NTFS 被快速格式化成 NTFS 后数据恢复的研究

陈培德[1,2],王丽清[1,2],吴建平[1,2]
(1. 云南大学 信息学院,云南 昆明 650223;
2. 云南省高校数字媒体技术重点实验室,云南 昆明 650223)

摘 要:快速格式化是高级格式化中的一种特殊形式。逻辑盘被快速格式化后数据能否恢复取决于快速格式化操作对原来文件系统中所存储数据的破坏程度。以 Windows 7 为平台,WinHex 15.08 为分析工具,对 NTFS 文件系统结构进行分析,将逻辑盘由 NTFS 文件系统快速格式化成 NTFS 文件系统,通过快速格式化后与快速格式化前对元文件 $MFT 变化的对比,提出了恢复快速格式化前 NTFS 文件系统数据的基本思路、方法与步骤。实验结果表明,将逻辑盘由 NTFS 文件系统快速格式化成 NTFS 文件系统后,只要恢复格式化前的元文件 $MFT 的 80H 属性值,通过 CHKDSK 命令,便可以恢复被快速格式化破坏的 NTFS 文件系统结构,除部分数据被覆盖无法恢复外,其他未覆盖的数据均可全部恢复。

关键词:格式化; FAT32 文件系统;NTFS 文件系统;数据恢复
中图分类号:TP311.12 **文献标识码**:A **文章编号**:1673-629X(2018)08-0191-05
doi:10.3969/j.issn.1673-629X.2018.08.040

Research on Data Recovery of Quick NTFS Formatted

CHEN Pei-de[1,2], WANG Li-qing[1,2], WU Jian-ping[1,2]
(1. School of Information Science and Engineering, Yunnan University, Kunming 650223, China;
2. Key Laboratory of Digital Media Technology of Universities and Colleges in Yunnan Province, Kunming 650223, China)

Abstract: The quick disk formatting is advanced and special. After the logic disk has been quickly formatted, the data recovery depends on the extent to the data stored in the original file system. Based on Windows 7 and WinHex 15.08, we analyze the systematic structures of NTFS and the logical disk quick formatted into NTFS. Contrasting the change of the $MFT to the quick formatting before and after, we present the basic idea, method and steps to restore the data of NTFS file system. Experiment shows that as long as the 80H value of the $MFT is restored before quick formatting, by running CHKDSK command, most of the data on the logical disk will recover after the logical disk is quickly formatted into NTFS file system. But the partial data be overwritten cannot recover.

Key words: format; FAT32 file system; NTFS file system; data recovery

0 引 言

格式化是指对磁盘或磁盘中的分区进行初始化的一种操作,这种操作通常会导致现有的磁盘或分区中所有的文件被清除。格式化通常分为低级格式化和高级格式化[1]。如果没有特别指明,对硬盘的格式化通常是指高级格式化。

外存储器在生产出来后,一般要经过低级格式化、分区和高级格式操作后,才能用来存储数据[1]。低级格式化针对的是整个硬盘,一般由外存储器生产厂商来完成;而对外存储器的分区和高级格式化一般由销售商或者用户来完成,高级格式化针对的是某个分区。

对逻辑盘进行高级格式化后数据能否恢复,取决于格式化操作对原来文件系统中数据的破坏程度[2]。而快速格式化则是高级格式化的一种特殊形式。文中对 NTFS 文件系统的逻辑盘被快速格式化成 NTFS 文件系统后的数据恢复进行了大量实验。结果表明,在 Windows 7 平台下,NTFS 的逻辑盘被快速格式化成 NTFS 后,除被元文件 $MFT 覆盖后的记录无法恢复外,其他未被覆盖的记录均可全部恢复。

1 NTFS 文件系统

从整体结构上讲,NTFS 文件系统由元文件、用户

收稿日期:2017-07-05 **修回日期**:2017-11-16 **网络出版时间**:2018-03-07
基金项目:云南省科技创新强省计划项目(2014AB021);云南省高校数字媒体重点实验室开放基金项目(2015KFKT002)
作者简介:陈培德(1966-),男,工程师,研究方向为文件系统与数据恢复技术。

文件以及数据等组成。NTFS 文件系统在创建时,会将一些重要的系统信息以文件的形式分散地存储在 NTFS 卷中[3],存储这些重要系统信息所对应的文件就是元文件[4],它是 NTFS 文件系统最重要的组成部分。除根目录外,元文件的名称均以“$”符号开头[5];元文件是隐藏的系统文件[6],用户不能直接对元文件进行访问,在资源管理器中也查看不到元文件。

2 元文件 $MFT

在 NTFS 文件系统中,最重要的元文件就是 $MFT[7-8],它是 NTFS 文件系统中所有文件和文件夹(目录)的集合[9]。它记录着 NTFS 文件系统中所有文件和文件夹的基本情况[10],包括卷的信息、引导记录、元文件 $MFT 本身等的重要信息,以及文件名(或文件夹名)、文件安全属性、文件大小、数据运行列表等等[11]。元文件 $MFT 由许许多多记录组成[12],每条记录的大小固定为 1 024 字节[13],一般情况下,每个文件或文件夹在元文件 $MFT 中只占用一条记录[14]。每条记录以“FILE”作为开始标记,一般以第 1 个“FF FF FF FF 00 00 00 00”或者“FF FF FF FF 82 79 47 11”(存储形式)为结束标志[15]。元文件 $MFT 的结构大致如图 1 所示[15]。

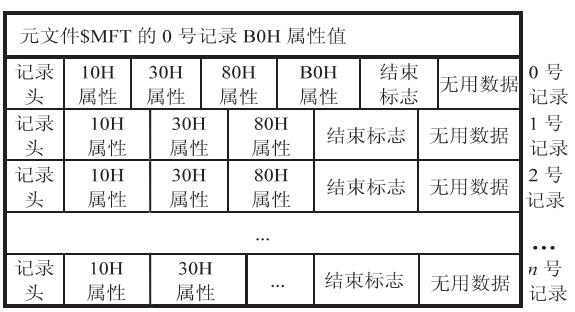

图 1 元文件 $MFT 结构示意图

(注:假设元文件 $MFT 的 0 号记录 B0H 属性值存放在元文件 $MFT 记录之前)

3 实验环境及制作实验素材

3.1 实验环境

(1)操作系统:Windows 7;

(2)逻辑盘容量:5.99 GB;

(3)分区形式:MBR;

(4)数据恢复工具:WinHex 15.08。

3.2 制作实验素材

制作实验素材步骤如下:

(1)在 Windows 7 操作系统下,使用计算机管理功能中的磁盘管理功能建立一个虚拟磁盘文件,文件名为 abcd.vhd,文件大小为 6.0 GB,将该虚拟磁盘文件 abcd.vhd 附加成虚拟盘;

(2)将虚拟盘初始化,分区形式选择 MBR,分区大小为 5.99 GB,假设对应的盘符为 G 盘;

(3)将 G 盘格式化,文件系统选择 NTFS,每个簇的扇区数选择“默认配置大小”;

(4)格式化完成后,复制 125 000 多个文件(夹)到 G 盘中,复制完成后;G 盘的基本情况如下:

文件系统:NTFS;

总容量:6 439 301 120 字节(5.99 GB);

已用空间:1 552 965 632 字节(1.44 GB);

可用空间:4 886 335 488 字节(4.54 GB);

每个簇的扇区数:8;

元文件 $MFT 记录范围:0 ~ 125 183。

(5)快速格式化前,NTFS 主要元文件在 G 盘分布情况如表 1 所示。

表 1 NTFS 主要元文件在 G 盘分布情况(快速格式化前)

元文件	扇区号	对应簇号
$Boot	0 ~ 15	0 ~ 1
$MFTMirr	16 ~ 23	2
$UpCase	24 ~ 279	3 ~ 34
$LogFile	1 962 736 ~ 2 029 743	245 342 ~ 253 717
$Attrdef	2 030 264 ~ 2 030 271	253 783
$Bitmap	2 096 752 ~ 2 097 135	262 094 ~ 262 141
$MFT	2 097 152 ~ 2 347 519	262 144 ~ 293 439

(6)将 G 盘进行快速格式化,文件系统选择 NTFS,每个簇的扇区数选择“默认配置大小”;快速格式化后,G 盘的基本情况如下:

文件系统:NTFS;

总容量:6 439 301 120 字节(5.99 GB);

已用空间:61 476 864 字节(58.60 MB);

可用空间:6 377 824 256 字节(5.93 GB);

每个簇的扇区数:8;

元文件 $MFT 记录范围:0 ~ 255。

(7)快速格式化后,NTFS 主要元文件在 G 盘分布情况如表 2 所示。

表 2 NTFS 主要元文件在 G 盘分布情况(快速格式化后)

元文件	扇区号	对应簇号
$Boot	0 ~ 15	0 ~ 1
$MFTMirr	16 ~ 23	2
$UpCase	24 ~ 279	3 ~ 34
$LogFile	1 962 736 ~ 2 029 743	245 342 ~ 253 717
$Attrdef	2 030 264 ~ 2 030 271	253 783
$Bitmap	2 096 752 ~ 2 097 135	262 094 ~ 262 141
$MFT	2 097 152 ~ 2 097 663	262 144 ~ 262 207

至此,在 Windows 7 平台下,G 盘由 NTFS 文件系统被快速格式化 NTFS 文件系统实验素材已制作完成。

4　快速格式化对 G 盘 NTFS 元文件的影响

从快速格式化前、后,G 盘 NTFS 主要元文件分布对比情况来看,G 盘 NTFS 主要元文件变化情况如下:

(1)元文件 $MFT 所占簇数由 31 296 个缩小为 64 个,元文件 $MFT 的记录号范围由 0 ~ 125 183 缩小为 0 ~ 255,而快速格式化前元文件 $MFT 的 256 ~ 125 183 号记录均完好保存;由于快速格式化后元文件 $MFT 的记录号范围为 0 ~ 255,所以,快速格式化前元文件 $MFT 的 256 ~ 125 183 号记录已不再起作用。

(2)快速格式化前,元文件 $MFT 的 0 号记录 B0H 属性值记录了 125 184 条记录(记录号范围为 0 ~ 125 183)的使用情况,而快速格式化后,只记录了 256 条记录的使用情况,其中 35 ~ 255 号记录为空记录。

(3)元文件 $MFTMirr 所占簇数和位置未发生变化,但是,元文件 $MFTMirr 的 0 号记录已被快速格式化后元文件 $MFTMirr 的 0 号记录所取代。

(4)元文件 $Bitmap 所占簇数和位置未发生变化,但是,元文件 $Bitmap 的内容已被快速格式化后元文件 $Bitmap 的内容所取代,即快速格式化前用户文件和文件夹所占据的位图已被全部释放。

而 G 盘 MBR 分区表、元文件 $Boot、$LogFile、$Attrdef、$UpCase 和元文件 $MFT 的开始簇号等均未发生任何变化。

5　恢复数据的基本思路

从快速格式化操作对 NTFS 元文件 $MFT 的影响可知,恢复数据的关键在于重建快速格式化前 G 盘的 NTFS 元文件 $MFT,即将快速格式化后元文件 $MFT 的记录号范围由 0 ~ 255 恢复为快速格式化前的 0 ~ 125 183。

要实现这一目标,只需要修改元文件 $MFT 的 0 号记录 80H 属性中的如下值:

(1)元文件 $MFT 所占簇数;

(2)元文件 $MFT 结束 VCN;

(3)系统分配给元文件 $MFT 的空间;

(4)元文件 $MFT 实际占用空间;

(5)元文件 $MFT 初始化空间。

将上述属性值修改完成并存盘。

回到 DOS 提示符下,使用 CHKDSK 命令[13]以元文件 $MFT 中的记录作为依据对 G 盘受损的 NTFS 文件系统进行自动修复;修复完成后,便恢复出快速格式化前 G 盘中的大部分数据。恢复快速格式前的数据流程如图 2 所示。

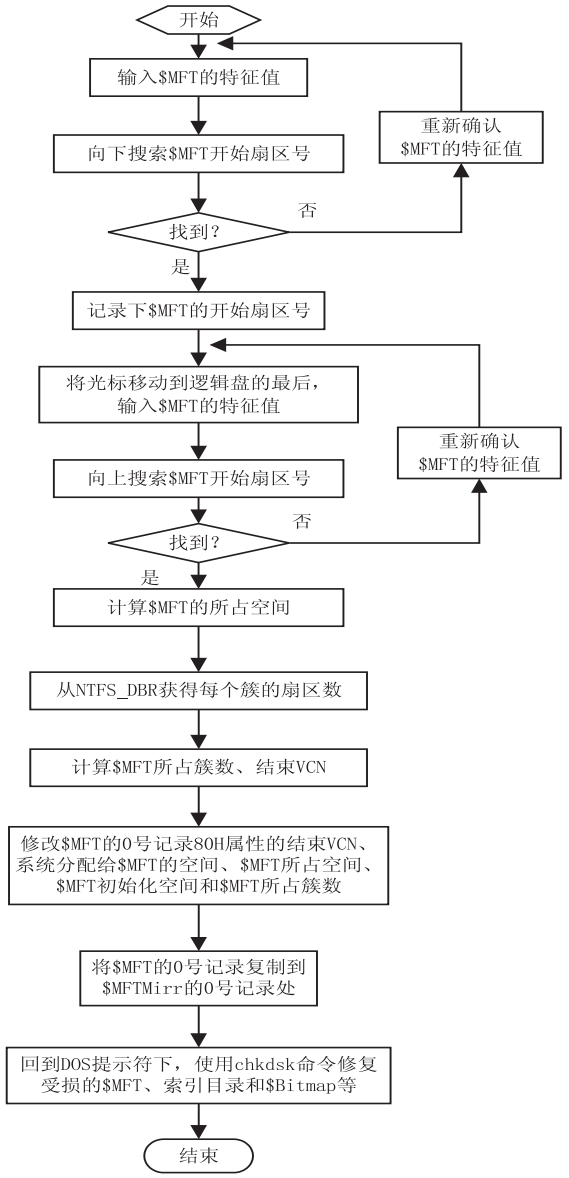

图 2　恢复快速格式化前数据流程

6　恢复数据的基本方法

计算元文件 $MFT 所占簇数、元文件 $MFT 结束 VCN 和系统分配给元文件 $MFT 的空间等的具体方法如下:

(1)元文件 $MFT 所占簇数 = (元文件 $MFT 结束扇区号 - 元文件 $MFT 开始扇区号 + 1)/每个簇的扇区数;

(2)元文件 $MFT 结束 VCN = 元文件 $MFT 所占簇数 - 1;

(3)系统分配给元文件 $MFT 的空间 = 元文件 $MFT 所占簇数 × 每个簇的扇区数 × 512 字节;

(4)元文件 $MFT 实际占用空间 = 元文件 $MFT 所占簇数 × 每个簇的扇区数 × 512 字节;

(5)元文件 $MFT 初始化空间 = 元文件 $MFT 记

录所占簇数×每个簇的扇区数×512 字节。

每个簇的扇区数从 NTFS_DBR 中获得。

将以上 5 个值转换成在元文件 $ MFT 的 0 号记录 80H 属性中的存储形式,并替换快速格式化后元文件 $ MFT 的 0 号记录 80H 属性中的这 5 个值,从而达到恢复快速格式化前元文件 $ MFT 的目的。

7 恢复数据基本步骤

7.1 计算元文件 $ MFT 的 0 号记录 80H 属性值

计算快速格式化前,元文件 $ MFT 的 0 号记录 80H 属性值步骤如下:

1. 计算快速格式化前,元文件 $ MFT 所占簇数。

操作步骤如下:

(1)启动 WinHex 软件。

(2)工具→逻辑硬盘→选择 G 盘,将光标移动到 24 号扇区(注:16 ~ 23 号扇区被元文件 $ MFTMirr 所占据,共计 8 个扇区)。

(3)搜索→查找文本→在查找文本窗口的文本框中输入元文件 $ MFT 记录的特征值“FILE”,扇区偏移地址为“0”,查找方向选择“向下”;在 2 097 152 号扇区找到,经确认为元文件 $ MFT 的 0 号记录所在扇区。

(4)将光标移动到逻辑盘的最后一个扇区;搜索→查找文本→在查找文本窗口的文本框中输入元文件 $ MFT 记录的特征值“FILE”,扇区偏移地址为“0”,查找方向选择“向上”;在 2 347 518 号扇区找到,经确认为元文件 $ MFT 最后一条记录的开始扇区号;所以,元文件 $ MFT 结束扇区号为 2 347 519。

(5)从 NTFS_DBR 中获得每个簇的扇区数为 8。

元文件 $ MFT 所占簇数 = (元文件 $ MFT 结束扇区号 - 元文件 $ MFT 开始扇区号 + 1)/每个簇的扇区数 = (2 347 519 - 2 097 152 + 1)/8 = 31 296 (即 0X7A40)

2. 计算元文件 $ MFT 结束 VCN。

元文件 $ MFT 结束 VCN = 元文件 $ MFT 所占簇数 - 1 = 31 296 - 1 = 31 295(即 0X7A3F)

3. 计算系统分配给元文件 $ MFT 的空间。

系统分配给元文件 $ MFT 记录空间 = 元文件 $ MFT 所占簇数×每个簇的扇区数×512 字节 = 31 296×8×512 字节 = 128 188 416(即 0X7A40000) 字节

4. 计算元文件 $ MFT 实际占用空间。

元文件 $ MFT 实际占用空间 = 元文件 $ MFT 所占簇数×每个簇的扇区数×512 字节 = 31 296×8×512 字节 = 128 188 416(即 0X7A40000) 字节

5. 计算元文件 $ MFT 初始化空间。

元文件 $ MFT 初始化空间 = 元文件 $ MFT 所占簇数×每个簇的扇区数×512 字节 = 31 296×8×512 字节 = 128 188 416(即 0X7A40000) 字节

将元文件 $ MFT 所占簇数、元文件 $ MFT 结束 VCN、系统分配给元文件 $ MFT 的空间、元文件 $ MFT 实际占用空间和元文件 $ MFT 初始化空间转换为在元文件 $ MFT 的 0 号记录 80H 属性中的存储形式,如下所示:

(1)元文件 $ MFT 所占簇数:40 7A;

(2)元文件 $ MFT 结束 VCN:3F7A000000000000;

(3)系统分配给元文件 $ MFT 空间:00 00 A4 07 00 00 00 00;

(4)元文件 $ MFT 实际占用空间:00 00 A4 07 00 00 00 00;

(5)元文件 $ MFT 初始化空间:00 00 A4 07 00 00 00 00。

7.2 恢复元文件 $ MFT 的 0 号记录 80H 属性值

恢复快速格式化前元文件 $ MFT 的 0 号记录 80H 属性中元文件 $ MFT 所占簇数、元文件 $ MFT 结束 VCN 等 5 个值,操作步骤如下:

1. 启动 WinHex 软件。

2. 工具→逻辑硬盘→选择 G 盘,将光标移动到 2 097 152号扇区,即快速格式化后,元文件 $ MFT 的 0 号记录 80H 属性处,如图 3 所示。

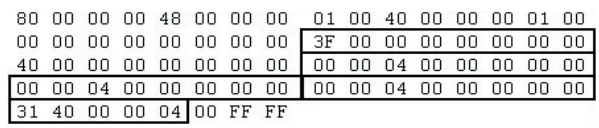

图 3 快速格式化后,元文件 $ MFT 的 0 号记录 80H 属性

3. 修改元文件的 0 号记录 80H 属性中的以下 5 个值。

(1)将元文件 $ MFT 的 0 号记录 80H 属性中的数据运行列表由“31 40 00 00 04 00 FF FF”修改为“32 40 7A 00 00 04 00 FF”,即元文件 $ MFT 所占簇数由 0X40 修改为 0X7A40,也就是将元文件 $ MFT 的记录号范围由 0 ~ 255 恢复为快速格式化前元文件 $ MFT 的记录号范围 0 ~ 125 183;

(2)将元文件 $ MFT 结束 VCN 由 0X3F 修改为“0X7A3F”,即将快速格式化后元文件 $ MFT 的虚拟簇号范围由 0 ~ 63 修改为快速格式化前的 0 ~ 31 295;

(3)将系统分配给元文件 $ MFT 的空间由“0X040000”修改为“0X07A40000”;即将快速格式化后系统分配给元文件 $ MFT 的空间由 262 144 字节修改为快速格式化前的 128 188 416 字节;

(4)将元文件 $ MFT 实际占用空间由“0X040000”修改为“0X07A40000”;即将快速格式化

后元文件 $MFT 实际占用空间由 262 144 字节修改为快速格式化前的 128 188 416 字节;

(5)将元文件 $MFT 初始化空间由“0X040000”修改为“0X07A40000”;即将快速格式化后元文件 $MFT 初始化空间由 262 144 字节修改为快速格式化前的 128 188 416 字节。

4. 修改完成后,元文件 $MFT 的 0 号记录 80H 属性的值如图 4 所示,即快速格式化前,元文件 $MFT 的 0 号记录 80H 属性值,然后存盘并退出 WinHex。

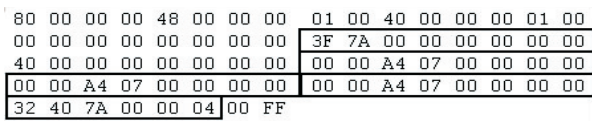

图 4　快速格式化前,元文件 $MFT 的 0 号记录 80H 属性

至此,元文件 $MFT 的 0 号记录已恢复到快速格式化前的状态。

5. 在 DOS 提示符下,使用“CHKDSK G:/F/I”[14] 命令完成如下 NTFS 文件系统的自动修改:

(1)通过元文件 $MFT 的 0 号记录自动修复受损的元文件 $MFT 和元文件 $MFT 的 B0H 属性值;

(2)通过元文件 $MFT 自动修复受损的元文件 $Bitmap;

(3)通过元文件 $MFT 自动修复受损的根目录、其他索引目录以及其他元文件等。

经过半个多小时,G 盘的元文件 $MFT、元文件 $Bitmap、根目录、索引目录以及其他元文件等已修复完成。G 盘的基本情况如下所示:

(1)文件系统:NTFS;

(2)总容量:6 439 301 120 字节(5.99 GB);

(3)已用空间:1 561 186 304 字节(1.44 GB);

(4)可用空间:4 878 114 816 字节(4.54 GB);

(5)每个簇的扇区数:8;

(6)元文件 $MFT 记录范围:0 ~ 125 183。

到资源管理器中,可以查看到 G 盘中的文件和文件夹。

8　结束语

在 Windows 7 平台下,将 NTFS 文件系统快速格式化成 NTFS 文件系统后,在每个簇的扇区数没有变化的情况下,只要正确修复元文件 $MFT 的 0 号记录 80H 属性相应值,并在 DOS 下使用 CHKDSK[16] 命令对受损的元文件以及索引目录进行自动修复,便可恢复快速格式化前 NTFS 文件系统中的大部分数据。但是由于元文件 $MFT 的 38 ~ 255 之间的记录已被填充为有效记录,这 218 条记录对应的文件内容可以通过 WinHex 软件的“按类型恢复文件”进行恢复。

综上所述,在 Windows 7 平台下,NTFS 文件系统被快速格式化成 NTFS 文件系统后,数据恢复的核心在于恢复快速格式化前元文件 $MFT 的 0 号记录 80H 属性中的相应值。通过实践证明,该方法方便快捷、简单实用,并且用户在资源管理器中可以看到所恢复出来的大部分数据。

参考文献:

[1] 陈培德,吴建平,王丽清,等. FAT32 被格式化成 NTFS 后数据恢复的研究[J]. 实验科学与技术,2018,16(1):9-12.

[2] 陈培德,吴建平,王丽清. NTFS 文件系统实例详解[M]. 北京:国防工业出版社,2015.

[3] 高志鹏,尤俊生. NTFS 空间再分配清零策略研究[J]. 计算机科学,2015,42(10A):92-94.

[4] 刘　伟. 数据恢复技术深度揭秘[M]. 北京:电子工业出版社,2010.

[5] 马　林. 数据重现:文件系统原理精解与数据恢复最佳实践[M]. 北京:清华大学出版社,2009.

[6] 赵双峰,费金龙,刘　楠,等. Windows NTFS 下数据恢复的研究与实现[J]. 计算机工程与设计,2008,29(2):306-308.

[7] 李步升. 基于 NTFS 的计算机反取证研究与实现[J]. 计算机工程,2010,36(19):274-276.

[8] 戴士剑,涂彦晖. 数据恢复技术[M]. 北京:电子工业出版社,2005.

[9] 白桂梅,张新颜,朱长青,等. NTFS 文件隐藏方式研究[J]. 计算机应用与软件,2014,31(8):298-299.

[10] IVENS K,GARDINILER K. Windows 200:the complete reference[M]. Beijing:China Machine Press,2000.

[11] CARRIER B. File system forensic analysis[M]. [s.l.]:Addison Wesley Professional,2005.

[12] RUSSINOVICH M E,SOLOMON D A,LONESCU A. Windows internals[M]. Beijing:Post & Telecom Press,2009.

[13] SOLOMAN D A. Inside windows NT[M]. 2nd ed. Washington,USA:Microsoft Corporation,1998.

[14] 梁金千,张　跃. NTFS 文件系统的主要数据结构[J]. 计算机工程与应用,2003,39(8):116-118.

[15] 陈培德,吴建平,王丽清. 重建 NTFS 的 DBR 及分区表的研究与实现[J]. 实验科学与技术,2016,14(6):56-59.

[16] 李佟鸿,王　宁,刘志军. 计算机系统信息隐藏反取证技术[J]. 计算机系统应用,2013,22(5):1-4.