doi: 10.3969/j.issn.1671-1122.2012.01.020

Windows 缩略图缓存文件的分析和取证

刘浩阳

（大连市公安局，辽宁大连 116012）

摘　要：尽管缩略图文件在 Windows 系统中一直存在，但是它的数据结构一直少有人关注。缩略图文件作为保存图形文件的重要数据的数据库，在计算机取证中具有重要作用。文章通过对不同 Windows 系统下的缩略图文件的结构进行解析，阐述如何利用缩略图文件来恢复数据和获取关键证据。

关键词：Thumbs.db;Thumbcache；计算机取证；复合文件

中图分类号：TP393.08　**文献标识码：**A　**文章编号：**1671-1122（2012）01-0074-03

The Analysis and Forensic of Thumbs.db

LIU Hao-yang

(*Dalian Municipal Public Security Bureau, Dalian Liaoning* 116012, *China*)

Abstract: Windows operating system have a character.It is Thumbs.db(Windows7 change it to Thumbcache). Thumbs.db is a database.It have many information like timestamp,picture,etc.The structure of Thumbs.db is unkown. I decoded the Thumbs.db to show how to analyse it in computer forensic area.

Key words: Thumbs.db;Thumbcache;computer forensic;compound file

从 Windows98 开始，为了加速显示图片，Windows 定义了缩略图文件（Thumbs.db），这样不必为了查看图片内容而一个个的打开原始图片。当以缩略图查看时（图片或视频），将会生成一个 Thumbs.db 文件。Thumbs.db 文件是一个数据库，Thumbs.db 里面保存了这个目录下所有图像文件的缩略图，Thumbs.db 文件可缓存图像文件的格式包括：jpeg，bmp，gif，tif，pdf 和 htm。不管原始图片是哪种格式，缩略图的格式都是 jpeg。

Windows 98/me/2000/XP/2003 的缩略图文件 Thumbs.db 以隐藏文件形式保存在每一个包含图片或照片的目录中。在 Windows vista/7/2008 中，\Users\< 用户名 >\AppData\Local\Microsoft\Windows\Explorer 目录下的“thumbcache_ 像素值 .db”文件代替了 Thumbs.db 文件成为新的缩略图文件，它的大小可以达到 256×256 像素。需要注意的是，Windows2000 使用 NTFS 3.1 格式时不会使用 Thumbs.db 文件，而是使用 Alternate Data Stream（ADS）来代替它。而当 Windows2000 使用 FAT 时，使用 Thumbs.db。

表1 Windows系统的缩略图存放路径

操作系统	缩略图缓存文件保存位置
Windows 98/me/2000/XP/2003	图片目录内
Windows vista/7/2008	\Users\< 用 户 名 >\AppData\Local\Microsoft\Windows\Explorer

1 缩略图文件的数据结构

Windows98/me/2000 的缩略图缓存文件与 Windows XP/2003 的结构有所不同。Windows98/me/2000 的 Thumbs.db 文件不但保存了缩略图，还包含了文件名、驱动器、路径等信息；Windows XP/2003 除了缩略图之外，只保存文件名；Windows Vista/7/2008 连文件名也不保存，只保存文件名的哈希。因此在计算机取证中，Windows Vista/7/2008 的缩略图文件的调查获得的信息较少。

表2 Windows系统下的缩略图信息

	Windows98/me/2000	Windows XP/2003	Windows Vista/7/2008
驱动器	有	无	无
文件名	有	有	无
路径	有	无	无
文件名哈希	无	无	有

下面就分别以 Windows 98/me/2000/XP/2003 和 Windows Vista/7/2008 为例说明二者的缩略图缓存文件的格式。

收稿时间：2011-12-15

作者简介：刘浩阳（1978-），男，网警支队电子物证勘验大队大队长，电子物证检验鉴定实验室主任，硕士，主要研究方向：计算机取证、网络犯罪侦查。

1.1 Windows 98/me/2000/XP/2003的缩略图文件格式

Windows 98/me/2000/XP/2003 都为 Thumbs.db，保存在图片所在目录之中，以隐藏文件的形式存在。缩略图缓存文件 Thumbs.db 的结构属于复合文件形式，类似于 word 文件。虽然小，但是有着比较复杂的结构。数据通常以流（Streams）形式来保存，主要有目录流（Catalog）和缩略图（Thumbs.db）两种。目录流主要有索引号、图片的最后修改时间、文件名，缩略图包括变换的索引号（例如在目录流中的索引号为“123”，在此就是“321”）和缩略图。除此之外，在缩略图缓存文件的 Root Entry 中还包含有缩略图混存文件的最后修改时间（在 Windows Vista/72008 中这个最后修改时间被取消了）。

复合文件主要有两个特点：

1）复合文件是以物理结构描述目录和数据流的存储方式。

2）复合文件被划分为大小相等的存储空间。每个空间的大小为一个扇区（512 字节）。扇区按照在复合文件中顺序编号，编号从 0 开始。

每个缩略图缓存文件 Thumbs.db 有着标志性的文件头签名“D0 CF 11 E0 A1 B1 1A E1”。文件头占用 512 字节，描述了结构信息，是整个文档的关键部分。

以一个实例 Thumbs.db 文件来详细讲解缩略图缓存文件的结构。

表3 缩略图缓存文件结构

Thumbs.db 文件头格式		
00000000h	D0 CF 11 E0 A1 B1 1A E1 00 00 00 00 00 00 00 00	邢.唷??.....
00000010h	00 00 00 00 00 00 00 00 3E 00 03 00 FE FF 09 00	........>...?..
00000020h	06 00 00 00 00 00 00 00 00 00 00 00 01 00 00 00	.....
00000030h	01 00 00 00 00 00 00 00 00 10 00 00 02 00 00 00	.....
00000040h	02 00 00 00 FE FF FF FF 00 00 00 00 00 00 00 00	....? .....
00000050h	FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF	
00000060h	（以下均为 FF，略）	

D0 CF 11 E0 A1 B1 1A E1	Thumbs.db 文件头标志
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00	全为 0
FE FF	小端字节顺序
3E 00	文件格式修订号
03 00	文件格式版本号
09 00	扇区大小 =29=512 字节
06 00	短扇区大小 =26=64 字节
00 00 00 00 00 00 00 00 00 00 00 00	保留
01 00 00 00	扇区分配表占用的扇区总数 =0x01=1 个扇区
01 00 00 00	存放目录的目录流的第一个扇区编号 =0x01=1
00 00 00 00	保留
00 10 00 00	流的最小值 =4096 字节
02 00 00 00	短流扇区分配表的第一个扇区编号 =2
02 00 00 00	短流扇区分配表所占用的扇区总数 =2
FE FF FF FF	主扇区分配表的第一个扇区编号，如扇区分配表占用的扇区总数小于 109，则没有附加的主扇区，该值为有符号整数“FE FF FF FF” =-2
00 00 00 00	主扇区分配表占用的扇区总数 =0
00 00 00 00	主扇区分配表保存的前 109 个扇区编号。此例只有第一组有效值 =0（为第 0+1=1 个扇区，偏移量 0x01×0x200=0x200）

扇区表的组合方式与 FAT 扇区的组合方式一致，通过对扇区分配表的组合，可以计算数据块的长度，完整的恢复出图像缩略图。在某些软件无法恢复出完整的图像时，可以手动恢复出部分图像。

Thumbs.db 还有一个目录（catalog）数据，用于存储图片或者视频的文件名，如果图片或视频有元数据，则会存储元数据。由微软不公开缩略图文件的格式，这些元数据往往不为人所知。

表4 目录数据结构

名称	相对偏移量（十进制）	长度（字节）	说明
数据头	0	4	标记位“10 00 07 00”
	4	4	文件数量
	8	8	固定值“60 00 00 00 60 00 00 00”
元数据	16	4	元数据长度，含此 4 个字节。
	20	4	元数据数量
	24	8	64 位 Windows 时间表示的照片修改时间
	32	变长	以 Unicode 表示的文件名（以 0x0000 结尾）

以一个目录的数据为例：

表5 目录的短流数据

000047c0h	10 00 07 00 05 00 00 00 60 00 00 00 60 00 00 00	.....`...`...
000047d0h	2C 00 00 00 01 00 00 00 00 15 05 BE D3 77 CB 01	,..... 居 w?
000047e0h	49 00 4D 00 47 00 5F 00 35 00 30 00 39 00 33 00	I.M.G._.5.0.9.3.
000047f0h	2E 00 4A 00 50 00 47 00 00 00 00 00 2C 00 00 00	..J.P.G.....,...
00004800h	02 00 00 00 00 8B 88 9D D2 77 CB 01 49 00 4D 00	..... w?I.M.
00004810h	47 00 5F 00 35 00 30 00 38 00 37 00 2E 00 4A 00	G._.5.0.8.7...J.
00004820h	50 00 47 00 00 00 00 00 2C 00 00 00 03 00 00 00	P.G.....,.....
00004830h	00 CE 8F D5 D2 77 CB 01 49 00 4D 00 47 00 5F 00	. 找 w?I.
00004840h	35 00 30 00 38 00 39 00 2E 00 4A 00 50 00 47 00	M.G._.
00004850h	00 00 00 00 2C 00 00 00 04 00 00 00 00 8A 03 0A	5.0.8.9...J.P.G.
00004860h	D3 77 CB 01 49 00 4D 00 47 00 5F 00 35 00 30 00	....,.....?.
00004870h	39 00 30 00 2E 00 4A 00 50 00 47 00 00 00 00 00	?I.M.G._.5.0.
00004880h	2C 00 00 00 05 00 00 00 00 E9 58 50 D3 77 CB 01	9.0...J.P.G.....
00004890h	49 00 4D 00 47 00 5F 00 35 00 30 00 39 00 31 00	,..... P ?
000048a0h	2E 00 4A 00 50 00 47 00 00 00 00 00 00 00 00 00	I.M.G._.5.0.9.1.
.....	.....	..J.P.G.....
000049f0h	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00	.....

10 00 07 00	标志位
05 00 00 00	0x05=5 个元数据
60 00 00 00 60 00 00 00	固定值“60 00 00 00 60 00 00 00”
2C 00 00 00	元数据长度 =0x2C=44
01 00 00 00	元数据数量
00 15 05 BE D3 77 CB 01	照片修改时间 = 2010 年 10 月 30 日 09:42:42 GMT+8
49 00 4D 00 47 00 5F 00 35 00 30 00 39 00 33 00 2E 00 4A 00 50 00 47 00 00 00 00 00	名称 = I.M.G._.5.0.9.3...J.P.G....
其他 4 个元数据	

1.2 Window Vista/7/2008的缩略图文件格式

Window Vista/7/2008 使用了新的缩略图缓存文件，叫做 thumbcache。它们保存在 \Users\< 用户名 >\AppData\Local\Microsoft\Windows\Explorer 目录中。他们类似于以下文件名：

thumbcache_idx.db
thumbcache_ 像素值 .db 文件 ：thumbcache_32.db、thumbcache_96.db、thumbcache_256.db、thumbcache_1024.db
thumbcache_sr.db

当 Windows 使用 thumbcache 文件时，首先搜索索引文件 thumbcache_idx.db，在其中找到文件信息，再在 Thumbcache_32.db, Thumbcache_96.db, Thumbcache_256.db 和 Thumbcache_1024.db 获取缩略图。其中 Thumbcache_32.db, Thumbcache_96.db 储存 bmp 缩略图，Thumbcache_256.db 和

Thumbcache_1024.db 储存 jpg 缩略图。尽管缩略图缓存文件多了，但是相对于复合文件 Thumbs.db，Windows Vista/7/2008 中的 Thumbcache 文件的结构却相对简单。

Thumbcache_idx.db 主要分为两部分：文件头和节点

表6 Thumbcache_idx.db文件头结构

偏移量（十进制）	长度（字节）	说明
0	4	文件头标记 “49 4D 4D 4D” = “IMMM”
4	4	系统版本，例如 20=Windows Vista；21=Windows 7
8	4	未知
12	4	使用的节点数量
16	4	总节点数量
20	4	未知

表7 Thumbcache_idx.db节点结构

偏移量（十进制）	长度（字节）	说明
0	8	文件名哈希
8	8	最后修改时间，在 Win7 下，将不会有这项存在。
16	4	标记
20	4	在 Thumbcache_32.db 文件中的偏移量
24	4	在 Thumbcache_96.db 文件中的偏移量
28	4	在 Thumbcache_256.db 文件中的偏移量
32	4	在 Thumbcache_1024.db 文件中的偏移量
36	4	Thumbcache_sr.db 的偏移量

节点值如果是 -1 表明未使用。

thumbcache_ 像素值 .db 是实际保存缩略图的文件，其中的缩略图是连续存放的，以占位符结束。在新的缩略图加入后，将在文件的结尾增加。因此结构也比较简单，分为文件头和数据块。

表8 thumbcache_像素值.db文件头结构说明

偏移量（十进制）	长度（字节）	说明
0	4	文件头标记 “43 4D 4D 4D” = “CMMM”
4	4	系统版本，例如 20=Windows Vista；21=Windows 7
8	4	Thumbcache 的类型 0 =thumbcache_32.db, 32 x 32 1 = thumbcache_96.db, 96 x 96 2 = thumbcache_256.db, 256 x 256 3 = thumbcache_1024.db, 1024 x 768 4 = thumbcache_sr.db
12	4	第一个数据块的偏移量（也可以认为是文件头大小）
16	4	最后有效数据块的偏移量
20	4	数据块数量

紧着文件头就是数据块内容，结构如下

表9 Thumbcache_像素值.db数据块结构说明

偏移量（十进制）	长度（字节）	说明
0	4	标记 “43 4D 4D 4D” = “CMMM”
4	4	数据块大小
8	8	文件名哈希
16	8	文件扩展名（UTF-16 编码，有结束标记）。如果在 Win7下，此项不存在
24	4	识别符大小
28	4	附加大小
32	4	数据大小
36	4	未知
40	8	数据 CRC64 校验值
48	8	文件头 CRC64 校验值
56	识别符大小	识别符（UTF-16 编码，有结束标记）
……	附加大小	附加（一般为 0）
……	数据大小	数据

2 缩略图文件在取证中的应用

缩略图文件有一个特性，就是将原始图片删除，由于缩略图文件不会立即自动更新，而且是隐藏文件。因此还可以浏览到被删除的图片的缩略图，这一点特性在偏重于图形文件的调查的案件（例如色情案件）中极为有用。即使犯罪嫌疑人使用了加密来保存图片，例如 Microsoft Encrypting File System（EFS）来加密 jpeg、gif、bmp 和 ppt 文件，通过缩略图缓存文件仍然可以看到内容。

通过解析缩略图文件的结构，可以手动恢复其中的缩略图和获取元数据，但是使用工具更为快捷和简便。在 Windows 98/me/2000/XP/2003 的缩略图 Thumbs.db 文件中，对于数字取证有价值的信息主要是内嵌的 Thumbs.db 创建时间和图片的修改时间，再就是图片本身内容，这些都可以通过 Windows File Analyzer（WFA）来查看。

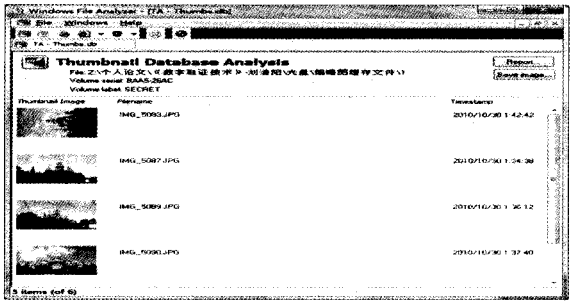

图1 Thumbs.db文件的解析结果

如图 1 所示，其中包含了6 幅图片的缩略图。可以看到保存的分区的卷标是“SECRET”，卷标是 BAA5-26AC。IMG_5093.JPG 的时间戳是 2010 年 10 月 30 日 1:42:42。值得注意的是，WFA 解析出来的是 GMT 时间，应该将其转换为中国地区时间（GTM+8），转换后的结果是 2010 年 10 月 30 日 9:42:42。

由于 Windows 7 中取消了缩略图缓存文件的时间戳和路径，因此，分析 Thumbcache 文件的元数据的意义已经不大，可以利用工具来解析图像。使用 ThumbDb Viewer 解析 Thumbcache 文件，因为 Thumbcache 的包含的信息较少，只能看到缩略图文件内容和大小，时间戳并不保存。

在很多案件中，调查人员都将工作重点放在了恢复图形文件上。但是由于图形文件本身的数据残缺，往往不能达到满意的效果。不妨变换一下思路，通过对缓存文件的分析，可以获取更多的线索和证据。（责编 岳逍远）

参考文献：

[1] 米佳，刘浩阳 计算机取证技术 [M]. 北京：群众出版社，2007.

[2] 麦永浩，孙国梓，许榕生，戴士剑 . 计算机取证与司法鉴定 [M]. 北京：清华大学出版社，2009，3.

[3] 丁丽萍 . 网络取证及计算机取证的理论研究 [J]. 信息网络安全，2010，(12)：38-41.