

2.3 文件签名

每一种文件都有自己的扩展名和特有的文件格式。文件有时没有扩展名，有时候会被人特地修改扩展名。在计算机中，出现文件类型与文件扩展名不匹配的情况很多。通过文件签名可以识别、验证一个文件的真实类型。利用取证软件可以快速地找到这些扩展名被修改的文件。此外，如果数据被删除，我们一样可以利用文件签名来进行数据恢复。本章实验重点练习如何有效地利用签名来进行数据分析。

一、文件签名

大多数文件都具有一些独特的字节，这些字节仅在此种文件格式中出现，我们称其为文件签名 (Signature)，或者称为文件头特殊标识 (Header)。这个标识可以是几个特殊的字符，也可能是几个十六进制字节。通过这些特殊的字节，X-Ways Forensics 可以依据自带的文件签名库文件，准确地识别出文件格式。在 Windows 操作系统中，Windows 注册表仅仅将文件扩展名与应用程序相关联，并利用相关联的程序打开相应的扩展名文件。

文件签名标识如果是正常 ASCII 码字符，例如，RAR 压缩文件的签名就是 Rar!这四个字符，看到这四个字符，我们立刻就可以读懂这个文件属于 RAR 压缩文件。但在多数文件头信息中，文件签名不是正常的字符。例如 MS OFFICE 文件中，文件签名就是 D0CF11E0A1B11AE1，通过这些十六进制数值可以认定属于 Office 2003 格式文件。

实验 2-15：分析文件的真实类型

本实验配合案例文件：C:\CDF\2-取证基础\2.4-File_Signature.zip。2.5-File_Signatures.pdf

要求：检查压缩文件中的文件确定其签名，并判断真实的文件类型。

步骤 1：解压缩“2.4-File_Signature.zip”文件。

步骤 2：打开 WinHex，创建案件，选择“添加目录”，添加解压缩后的文件夹 2.4-File_Signature。

步骤 3：选择 file1，查看文件的前八个字节为 50 4B 03 04 14 00 06 00。

文件名称

= (根目录)

= File Signature Examples (10)

File1

未验证

File Signature Examples

卷

文件

预览

详细

缩略图

时间轴

图例说明

同步

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
00000000	50	4B	03	04	14	00	06	00	08	00	00	00	21	00	E9	51	10	B0	8D	01	00	00	C2	05	00	00	13	00	08	02	5B	43
00000020	6F	6E	74	65	6E	74	5F	54	79	70	65	73	5D	2E	78	6D	6C	20	A2	04	02	28	A0	00	02	00	00	00	00	00	00	00
00000040	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000000C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000000E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000100	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000120	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000140	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000160	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000001A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000001C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
000001E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000200	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	

图 2-3-1 查看文件签名

步骤 4: 在文档 2.5-File_Signatures.pdf 中搜索 “50 4B 03 04 ”, 可知此签名与多种文件匹配。(PDF 中文件签名可参考网站 “http://www.garykessler.net/library/file_sigs.html”)

2.5-File_Signatures.pdf 中搜索 “50 4B 03 04 14 00 06 00”, 可知此签名仅与 Microsoft Office 的 Open XML 文件格式匹配。即与 Word 文档、Excel 文档 、PPT 演示文档包含相同的签名。

50 4B 03 04	PK...	ZIP	PKZIP archive file (Ref. 1 Ref. 2)
	Trailer: filename 50 4B 17 characters 00 00 00		
	Trailer: (filename PK 17 characters ...)		
	Note: PK are the initials of Phil Katz, co-creator of the ZIP file format and author of PKZIP.		
	ZIP	Apple Mac OS X Dashboard Widget, Aston Shell theme, Oolite eXpansion Pack, Opera Widget, Pivot Style Template, Rockbox Theme package, Simple Machines Forums theme, SubEthaEdit Mode, Trillian zipped skin, Virtual Skipper skin	
	APK	Android package	
	JAR	Java archive; compressed file package for classes and data	
	KMZ	Google Earth saved working session file	
	KWD	KWord document	
	ODT, ODP, OTT	OpenDocument text document, presentation, and text document template, respectively.	
	OXPS	Microsoft Open XML paper specification file	
	SXC, SXD, SXI, SXW	OpenOffice spreadsheet (Calc), drawing (Draw), presentation (Impress), and word processing (Writer) files, respectively.	
	SXC	StarOffice spreadsheet	
	WMZ	Windows Media compressed skin file	
	XPI	Mozilla Browser Archive	
	XPS	XML paper specification file	
	XPT	eXact Packager Models	
50 4B 03 04 0A 00 02 00	PK.....	EPUB	Open Publication Structure eBook file. (Should also include the string: mimetype=application/epub+zip)
50 4B 03 04 14 00 01 00	PK.....		
63 00 00 00 00 00	C.....	ZIP	ZLock Pro encrypted ZIP
50 4B 03 04 14 00 06 00	PK.....	DOCX, PPTX, XLSX	Microsoft Office Open XML Format (OOXML) Document
	NOTE: There is no subheader for MS OOXML files as there is with DOC, PPT, and XLS files. To better understand the format of these files, rename any OOXML file to have a .ZIP extension and then unzip the file; look at the resultant file named [Content_Types].xml to see the content types. In particular, look for the <Override PartName= tag, where you will find word, ppt, or xl, respectively.		
	Trailer: Look for 50 4B 05 06 (PK. .) followed by 18 additional bytes at the end of the file.		
50 4B 03 04 14 00 08 00	PK.....		
08 00	..	JAR	Java archive

图 2-3-2 搜索文件签名

实验 2-16: 查看峰会简版.BAK 文件的真实类型

本实验配合案例文件: C:\CDF\2-取证基础\2.1-CCFC.e01。

过滤分区 1 中的“峰会简版.BAK”文件，并以文件模式查看，可看到文件头部开始的几个 ASCII 字符是 Rar!。在文档 2.5-File_Signatures.pdf 中搜索“52 61 72 21 1A 07 00 ”可知该文件为 RAR v4.x 格式。

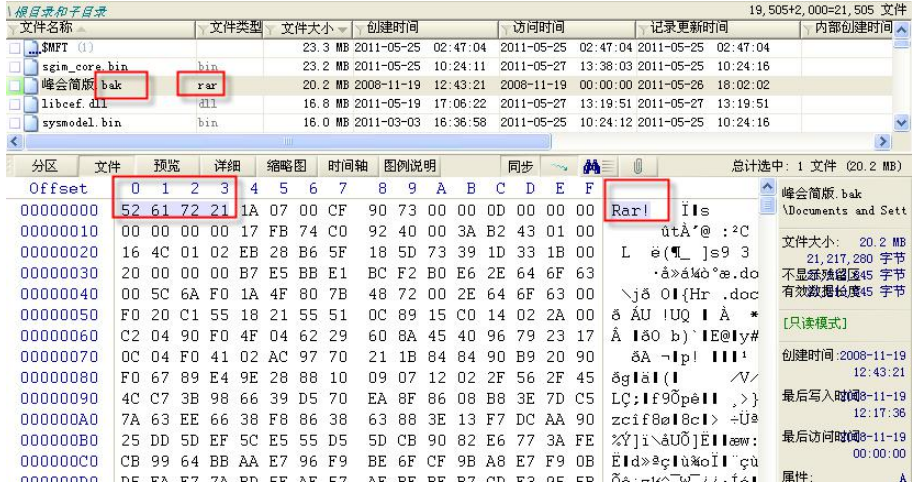


图 2-3-3 查看文件签名

二、签名状态

通常来说，文件的扩展名和文件签名应该是一致、匹配的。如 RAR 压缩文件的扩展名为 RAR，文件签名为文件开始前四个字节 Rar!，对应十六进制就是 51 61 72 21。当 X-Ways Forensics 读取到一个 RAR 压缩文件，会自动去和文件签名库中的特征字记录进行比对。如果两个记录匹配，则说明这是一个正常的 RAR 压缩文件，文件签名匹配。

但在一些特殊情况下，文件扩展名和文件签名会不匹配。例如没有扩展名的文件，bak 备份文件，加密文件，或者被人故意更改了扩展名，不愿意让别人轻易发现的文件。2.1-CCFC.e01 案例中的“峰会简版.RAR”文件就是被人为修改为“BAK”文件，即签名不匹配。利用取证分析软件可以快速筛选出来。

Winhex 中对于文件的签名状态，初始为“签名未校验”，经过比对文件签名库后：

- 1. 如果文件签名、扩展名和文件签名库中匹配，状态为“签名匹配”；
- 2. 如果文件类型在文件签名库中不存在，则签名状态是“不在列表中”；
- 2. 如果文件小于 8 字节，例如图 4 中的几个 0 字节文件，签名状态为“无关的”；
- 4. 如果扩展名在数据库中被引用，但文件签名未知，状态为“签名未校验”；
- 5. 如果文件签名在数据库中和某种文件类型匹配，但是文件扩展名是另一种文件类型或者根本没有扩展名，状态为“新确定”。

根目录和子目录 22 文件: 21,483 个被过滤掉

文件名称	文件类型	签名状态	文件大小	创建时间	修改时间	访问时间
已删除邮件. dbx	dbx	签名匹配	74.7 KB	2011-05-27	11...	2011-05-27 11...
草稿. dbx	dbx	签名匹配	74.7 KB	2011-05-27	11...	2011-05-27 11...
corpus. dbx	dbx	签名匹配	265 KB	2009-05-14	17...	2011-05-25 10...
ToDoFile. BOX	box	签名未校验	96 B	2009-05-25	12...	2011-05-25 09...
out. BOX	box	无关的	0 B	2011-05-25	09...	2011-05-25 09...
in. BOX	box	签名未校验	42.5 KB	2011-05-25	09...	2011-05-25 09...
Address. BOX	box	无关的	0 B	2011-05-25	09...	2011-05-25 09...
Send. BOX	box	签名未校验	0.6 KB	2011-05-25	09...	2011-05-25 10...
sent. BOX	box	无关的	0 B	2011-05-25	09...	2011-05-25 09...
spam. BOX	box	无关的	0 B	2011-05-25	09...	2011-05-25 09...
trash. BOX	box	无关的	0 B	2011-05-25	09...	2011-05-25 09...
out. BOX	box	签名未确认	134 KB	2011-05-25	10...	2011-05-25 10...
in. BOX	box	签名未确认	35.1 MB	2011-05-25	10...	2011-05-27 08...
sent. BOX	box	签名未确认	134 KB	2011-05-25	10...	2011-05-25 10...

图 2-3-4 签名状态

可以利用不同的签名状态进行文件过滤。



图 2-3-5 签名状态过滤

实验 2-17: 验证文件的签名状态

本实验配合案例文件: C:\CDF\2-取证基础\2.1-CCFC.e01。

要求: 判断所有文件签名状态, 发现分区 1 中签名不匹配的文件的数量。

X-Ways 能够根据文件签名库来校验一个文件的真实类型, 即检测那些文件扩展名与固定签名值不匹配的所有文件。例如: 如果某人故意将一张属于犯罪证据 JPEG 图片改名为 “invoice.xls”, 通过文件签名校验功能, X-Ways Forensics 能够自动将图片文件的真实类型 “jpg” 显示在目录浏览器的 “文件类型” 栏中。这属于磁盘快照的一个基本功能。有关磁盘快照的实验, 将在 2.6 节中详细练习。

1. 进行磁盘快照。选择快照选项。勾选 “在选定证据中搜索”、“基于文件签名和算法验证文件真实类型”, “选定的证据项中搜索” 后, 仅针对 “分区 1” 进行分析。

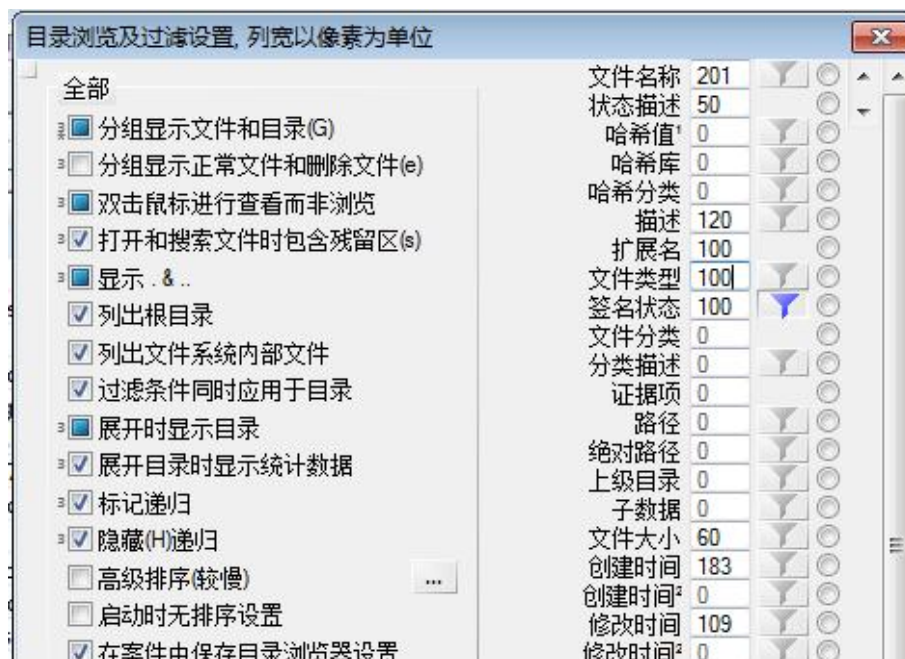


图 2-3-8 修改目录设置

3.通过“签名状态”过滤所有“不匹配”的文件

点击“签名状态”右侧的“漏斗”图标。选择“不匹配”，“激活”。



图 2-3-9 通过签名状态过滤

4.查看统计结果

点击“分区 1”，右键选择“浏览递归”。在窗口右侧，可以看到“1347+146=1493”个文件。

1 根目录和子目录		1,347+146=1,493 文件; 20,008 个被过滤掉				
文件名称	状态描述	描述	扩展名	文件类型	签名状态	文件大小 创建时间
7zNew.data	✓	文件, 存在的	data	7z	不匹配 (3)	32 B 2010-10-28, 12:37
rwebqq16.qq[1].66377934...	✓	文件, 存在的	663779343737383	ascii	不匹配 (2)	37 B 2011-05-27, 10:00
listen[1].2474558816114586	✓	文件, 存在的	2474558816114586	ascii	不匹配 (2)	43 B 2011-05-27, 12:44
rwebqq16.qq[1].05940078...	✓	文件, 存在的	05940078410346178	ascii	不匹配 (2)	38 B 2011-05-27, 10:42
rwebqq16.qq[1].77518076...	✓	文件, 存在的	775180769185406	ascii	不匹配 (2)	37 B 2011-05-27, 10:05
rwebqq16.qq[1].35500904...	✓	文件, 存在的	35500904552712053	ascii	不匹配 (2)	38 B 2011-05-27, 10:29
state[1].htm	✓	文件, 存在的	htm	ascii	不匹配 (2)	156 B 2011-05-27, 13:21
rwebqq16.qq[1].28476631...	✓	文件, 存在的	28476631093974386	ascii	不匹配 (2)	37 B 2011-05-27, 10:10
rwebqq16.qq[1].32186308...	✓	文件, 存在的	32186308301477895	ascii	不匹配 (2)	37 B 2011-05-27, 10:16
rwebqq13.qq[1].14489302...	✓	文件, 存在的	1448930299283101	ascii	不匹配 (2)	37 B 2011-05-27, 13:38
state[2].htm	✓	文件, 存在的	htm	ascii	不匹配 (2)	156 B 2011-05-27, 13:20
state[1].htm	✓	文件, 存在的	htm	ascii	不匹配 (2)	155 B 2011-05-27, 12:38
msgb_output_page[2].1890...	✓	文件, 存在的	18906458249241553&j...	ascii	不匹配 (2)	169 B 2011-05-27, 13:31
rwebqq16.qq[1].66196832...	✓	文件, 存在的	6619683296469445	ascii	不匹配 (2)	37 B 2011-05-27, 10:17
rwebqq16.qq[1].51193664...	✓	文件, 存在的	5119366403364639	ascii	不匹配 (2)	38 B 2011-05-27, 10:49
rwebqq16.qq[1].12733745...	✓	文件, 存在的	12733745423554393	ascii	不匹配 (2)	38 B 2011-05-27, 10:33
inetcorp.adm	✓	文件, 存在的	adm	ascii	不匹配 (2)	5.9 KB 2008-07-31, 00:00

图 2-3-10 过滤结果