

## Chapter 22

# COMPARISON OF THE DATA RECOVERY FUNCTION OF FORENSIC TOOLS

Joe Buchanan-Wollaston, Tim Storer and William Glisson

**Abstract** Commercially-available digital forensic tools are often large, expensive, complex software products, offering a range of functions to assist in the investigation of digital artifacts. Several authors have raised concerns about the reliability of evidence derived from these tools. This is of particular importance because many forensic tools are closed source and, therefore, are only subject to black box evaluation. In addition, many of the individual functions integrated into forensic tools are available as standalone products, typically at a much lower cost or even free. This paper compares – rather than individually evaluates – the data recovery function of two forensic suites and three standalone non-forensic commercial applications. Experimental results demonstrate that all the tools have comparable performance with respect to the data recovery function. However, some variation exists in the data recovered by the tools.

**Keywords:** Digital forensic tools, data recovery, testing

### 1. Introduction

Forensic tools are used by thousands of digital forensic professionals around the world. The functionality of forensic tools varies, although several features appear to be provided consistently, including hard disk image preparation and storage, data hashing of entire disk images or individual artifacts, disk image mounting and filesystem reconstruction, data presentation and visualization, and data carving of damaged images and deleted file contents.

Data related to the market shares of forensic tools appears to be a closely guarded secret. However, a review of online forums, corporate

websites and the research literature gives the impression that major vendors of forensic tools for personal computers are Guidance Software and AccessData, who market the EnCase [13] and Forensic Toolkit (FTK) [1] software suites, respectively. These software suites are widely used, perhaps due to the integration of several forensic applications in a common product. In addition, the provision of “push button” graphical user interfaces reduces the level of training and computing expertise required to conduct a forensic investigation.

Both Guidance Software and AccessData provide compelling arguments for employing their software in digital forensic investigations. For example, Guidance Software describes the EnCase suite as “the industry-standard computer forensics investigation solution” [13] while AccessData claims that FTK is “the most advanced computer forensics software available” [1]. Moreover, both vendors maintain that their software products are designed and validated to meet the standards of forensic evidence. FTK is described as a “court-validated digital investigations platform” [1]. EnCase is described similarly as having “an unsurpassed record of court acceptance” [13]. These claims are difficult to assess. In particular, it is unclear what level of scrutiny has been applied to the evidence produced by the forensic tools and what standards they have been assessed against.

A few independent reviews have compared the functionality and performance of forensic software suites. *SC Magazine* conducts annual group tests of forensic software, typically considering around eight to ten products [22]. Separately, the National Institute of Standards and Technology (NIST) has implemented the Computer Forensics Tools Testing (CFTT) Program [20]. This program has developed a draft validation framework for forensic data recovery tools [19]. To date, the framework has only been applied to a small selection of forensic software suites [15].

Anecdotal evidence suggests that forensic software suites may not be defect-free and that the results from different toolkits are likely to differ to some degree. The February 2008 release of FTK version 2 received bad press [2, 9, 23]. One review described the software as “an unmitigated disaster” [23]. It noted that users reported problems while installing and running the software. The documented minimum computer specification was also reported to be inadequate for the software. Mercuri [18] has noted that CFTT Program tests of EnCase and FTK revealed defects in the hard disk image preparation processes. Both tools were unable to recover some data from NTFS-formatted logical disk partitions.

Forensic tools can be expensive to purchase and operate. As of the middle of 2011, a single user license for EnCase or FTK was approximately \$2,995 and the annual cost of software maintenance and support

was an additional \$599 for EnCase and \$840 for FTK. The hardware requirements for these tools also impose significant costs. For example, the recommended system specification for FTK (version 3) includes an Intel i9 Dual Quad Core Xeon Processor, 12 GB RAM and a 160 GB solid state hard drive dedicated entirely to an Oracle database for case management. The online shopping service provided by a popular UK vendor was used in November 2011 to prepare an estimate for a machine with the minimum required specification. The estimate suggests that the recommended platform would cost approximately £2,600 (about \$4,100), not including peripheral equipment and sales taxes.

Despite the popularity of commercial forensic suites, numerous applications are available that provide equivalent functions at a much lower cost or even at no cost. Assembling a toolkit of mass market applications that has equivalent functionality to a forensic suite is an attractive proposition, not just to reduce costs. The provision of a supplementary, low-cost toolkit can ease the process of validating results generated by forensic suites and increase confidence in the reliability of evidence.

This paper compares the results of data recovery by digital forensic toolkits and mass market applications. Data recovery is the extraction and presentation of file contents from a disk image formatted using a known filesystem. This definition excludes the recording of the disk image itself and the recovery of file contents stored in areas of the disk image that are not managed by the filesystem.

The comparative approach mimics the situation faced by a digital forensic practitioner when confronted with a previously unseen data storage device. If the device is processed using different data recovery applications, there is the potential for the results to differ. This contrasts with the validation method adopted by NIST [19], in which a ground truth known data set of files or other data items is prepared and validated. In the work presented here, no assessment is made of the correctness of the forensic tools. Rather, the intention is to provide a means for quantifying the extent to which the data recovery results for different tools differ.

The comparative experiments described in this paper focused on several disk images representing the evolution of data stored on a computer as a result of user actions. The disk images were processed using a selection of recovery tools and the comparison results are presented. A holistic data comparison of the tools is conducted and the recovery of known marker files that were deliberately added to the disk image are assessed. Several experimental findings and their relation to previous literature are discussed, along with conclusions related to forensic investigations and tool validation.

## 2. Experimental Method

The purpose of this research is to compare the data recovery capabilities of a selection of software tools on a typical desktop personal computer setup. The personal computer is assumed to have office applications, web browsing, email communications and media playback. Some progress has been made in identifying realistic data sets for forensic tool analysis based on data found in re-sold hardware [10, 12, 16]. However, we are not aware of any research that establishes a characteristic data set for tool testing. We chose to develop our own data set to maintain control of the experiments.

A Windows XP operating system and a selection of desktop applications were installed on a pristine hard disk. A number of marker files representing what might be found in a typical user system were copied to the disk or created directly using the installed applications. Selected files were then deleted from the hard disk via the operating system user interface. Finally, a number of additional files were added to the disk, potentially overwriting files deleted by the user.

An image of the disk was taken at each stage of the experiment using FTK Imager; these images are referred to as Image1 through Image6. To gain assurance about the correctness of the images produced by FTK Imager (version 2.9.0.1385), the imaging process was repeated for the final image using the `dclfd` tool [14]. MD5 hashes of the `dclfd` and FTK images were computed using the `dclfd` and FTK tools. All four image hashes for Image6 matched. The files were then recovered from the images using several data recovery tools. All the files were hashed and the file hashes and reported filesystem paths were analyzed for variations.

### 2.1 Target System Setup

A 20 GB hard disk was chosen for the installation as it was deemed large enough to hold the operating system along with a variety of files (including photos and videos) while being relatively quick to image and process. Typical hard disks available in a new computer (as of 2012) range from 500 GB for a laptop to 2 TB for a desktop, but using such large disks would have considerably increased the time required for imaging and processing the disk multiple times.

Windows XP Professional SP3 was installed on the disk with a single user account. The operating system installation process formatted the target hard disk using NTFS. Software for a Netgear Wireless USB Adapter, Internet Explorer 8, Firefox 5, Microsoft Office 2007 and Skype 5.5 were also installed. At this point, the disk was imaged to create Image3. Windows was then activated. Earlier images (Image1 and Image2)

were recorded, but these images are not relevant to this work and are, therefore, not discussed.

## 2.2 Image Preparation with User Marker Files

Internet Explorer and Firefox were opened and a number of websites were visited using each browser. In each case, the URL was typed directly into the browser address bar instead of using a search engine or link to access the website. The sites visited were selected because they were known to be static sites without changing content such as advertising banners and graphics. The advantage of this approach is that if further analysis of the browser caches were to be performed, the origin of each web page and image file could be identified easily.

Skype was then opened and a voice call was initiated to one contact, followed by a short instant message session where a message was sent to the contact and a response received from the same contact. This created a history file for Skype that could be analyzed further, if required. Outlook was then opened, an email account was configured and a test email was sent to a contact. A reply was received back from the same contact. Four appointments were then added to the calendar. This resulted in user content being stored in the `outlook.pst` data file.

The next step was to open Windows Explorer and create a new folder in the My Documents folder. A new blank text document was created in the new folder, some text was added and the file was saved. A number of files were prepared and saved on an external hard drive. The files corresponded to those found on a typical computer, including word processing documents, spreadsheets, PDF documents, photographic images, plain text files, audio, video and executable files. The external hard drive was connected to the computer via a USB cable and the files were copied across to folders in the My Documents folder. A new Microsoft Word document was then created, some text was added and the file was saved. A new Microsoft Excel spreadsheet was also created. A total of 86 user files were added to the My Documents folder. At this stage, the disk was imaged to create Image4.

Table 1 summarizes the marker files, including the file IDs, file types and file extensions. The table also shows:

- **Manipulations to Images4 through Image6:** a = Added before Image4; b = Left in Recycle Bin for Image5; c = Deleted and removed from Recycle Bin; d = Permanently deleted in Image4; e = Altered before Image6; f = Added before Image6.
- **Recoveries from Image5:** g = EnCase; h = FTK; i = Recuva; j = R-Studio; k = Stella Phoenix.

Table 1. Summary of marker files.

File IDs	File Type	Ext	Manipulations/Recoveries													
			a	b	c	d	e	f	g	h	i	j	k	l	m	n
1	A - Created documents	docx	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
2	A - Created documents	txt	✓	✓												
3	A - Created documents	xlsx	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
4,5	Excel spreadsheets	xls	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
6	Excel spreadsheets	xls	✓			✓			✓	✓	✓	✓	✓	✓	✓	✓
7	Excel spreadsheets	xls	✓	✓												
8	Excel spreadsheets	xls	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
9-12	Movies	avi	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
13-16	Movies	avi	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
17,18	Movies	iso	✓	✓												
19	Music	mp3	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
20-23	Music	mp3	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
24,25	Music	mp3	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
26-28	Music	mp3	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
29,30	Other files	exe	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
31,32	Other files	exe	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
33,34	Other files	psd	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
35,36	Other files	psd	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
37-41	PDF files	pdf	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
42-46	PDF files	pdf	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
47-61	Photos	jpg	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
62-76	Photos	jpg	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
77,78	Text files	txt	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
79	Text files	txt	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓
80-81	Text files	txt	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
82	Word documents	doc	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
83	Word documents	doc	✓			✓			✓	✓	✓	✓	✓	✓	✓	✓
84,85	Word documents	doc	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓
86	Word documents	docx	✓✓						✓	✓	✓	✓	✓	✓	✓	✓
87	A - Created documents	docx						✓			✓	✓	✓	✓	✓	✓
88	A - Created documents	xlsx						✓			✓	✓	✓	✓	✓	✓
89	Excel spreadsheets	xls						✓			✓	✓	✓	✓	✓	✓
90-93	Excel spreadsheets	xlsx						✓			✓	✓	✓	✓	✓	✓
94-96	Movies	iso						✓			✓	✓	✓	✓	✓	✓
97	Movies	iso						✓			✓	✓	✓	✓	✓	✓
98,99	Movies	mts						✓			✓	✓	✓	✓	✓	✓
100	Music	mp3						✓								
101-109	Music	mp3						✓			✓	✓	✓	✓	✓	✓
110-118	Other files	exe						✓			✓	✓	✓	✓	✓	✓
119	Other files	msi						✓			✓	✓	✓	✓	✓	✓
120-129	PDF files	pdf						✓			✓	✓	✓	✓	✓	✓
130-159	Photos	jpg						✓			✓	✓	✓	✓	✓	✓
160-164	Text files	txt						✓			✓	✓	✓	✓	✓	✓
165	Word documents	doc						✓			✓	✓	✓	✓	✓	✓
166-169	Word documents	docx						✓			✓	✓	✓	✓	✓	✓

- **Recoveries from Image6:** l = EnCase; m = FTK; n = Recuva;  
o = R-Studio; p=Stella Phoenix.

A number of the files that had been copied across to the disk were deleted. Specifically, 81 files were moved to the Recycle Bin. Of these files, 42 were then removed from the Recycle Bin. Two files were reported as being permanently deleted by a Windows prompt because the files were too large for the Recycle Bin. The browsing histories were deleted from Internet Explorer and Firefox. Calendar items and all emails were deleted from Outlook, and the Deleted Items folder was then emptied. The Skype history was cleared. Approximately half of the files in the Recycle Bin were deleted, with a record being kept of the files that remained. The disk was then imaged again to create Image5.

The external hard drive was reconnected to the computer and another selection of 81 pre-prepared files were copied across to sub-folders in the My Documents folder. The copied files almost completely filled the remaining space on the disk, leaving a small amount of space for files created by the operating system such as during Internet browsing. This replicates behavior in which a user fills up a hard disk with data and is required to remove some old data to free up space. Firefox and Internet Explorer were opened and a number of websites were visited using each browser, again by typing the URLs directly into the address bars. Two new documents were created in Microsoft Word and Excel, text was added and the files were saved. Skype and Microsoft Outlook were then used and the disk was imaged again to create Image6.

## 2.3 Data Recovery Procedures

The following five tools were selected for comparison of data recovery functionality:

- EnCase (Guidance Software EnCase version 7.01.02.01)
- FTK (AccessData Forensic Toolkit version 3.1.2.2359)
- Recuva (Piriform Recuva version 1.40.525)
- R-Studio (R-TT R-Studio version 5.4, build 134130)
- Stellar Phoenix (Stellar Phoenix Windows Data Recovery version 4.2 Home Edition)

The first two tools are parts of commercially-available digital forensic software suites. They were selected due to their popularity with forensic practitioners and their availability for our research. The remaining three

tools are mass market data recovery applications that are not advertised as suitable for digital forensic recovery. A plethora of options exist in this category; the three tools were selected as representative of the market and available feature sets.

All five tools were installed on an HP Z400 workstation running Windows 7 Enterprise 64-bit with a Intel Xeon Dual Core W3503 Processor (2.40 GHz) with 8 GB RAM and a 750 GB hard disk. The disk images were all stored on this workstation.

Each of the applications presents a different selection of options to the user for the purpose of configuring the recovery process. This variability in features and presentation makes the direct comparison of the tools rather challenging. The configuration of each tool is presented here to support experimental repeatability. All the options cannot be described exhaustively, and the narrative records where non-default options (as presented to the user) were selected for an application.

For EnCase and FTK, the first step in processing evidence is to start a new case. A case contains all the evidence, bookmarks, information and reports, and allows searches of the evidence. The three mass market data recovery tools have no case management options. In the case of the EnCase and FTK forensic suites, the number of options that may be selected before the scanning and recovery processes is much greater than for a tool that only recovers data.

EnCase, FTK and R-Studio are designed to allow a raw disk image to be loaded directly into the software. The other two tools, Recuva and Stellar Phoenix, do not offer this capability. For these tools, the image must first be mounted in Windows as a logical drive; this was achieved using Mount Image Pro version 4.48(828) [11].

The five tools were configured as follows:

- **EnCase:** A new case was created and the image was added as evidence to this case with default options. The “Recover Folders” task was selected (only) and the processing was started.
- **FTK:** A new case was created and the image was added as evidence. All the options were disabled except for the generation of MD5 hashes and the processing was started.
- **Recuva:** The mounted drive was selected. In the recovery dialog, the options selected were “Show files found in hidden system directories,” “Show zero byte files,” “Deep scan” and “Scan for non-deleted files.”
- **R-Studio:** The disk image was opened and the “Whole disk scan” and “Detailed view during scan” options were selected.

- **Stellar Phoenix:** The “Search Drive” dialog was opened for the image and the “Physical Drive” method was used. The options “Deep Scan” and “Advanced Scan” were selected.

### 3. Experimental Results

The results reported in this section pertain to the data recovered from Image4, Image5 and Image6. These images represent the state of the target hard disk after user activity was simulated.

#### 3.1 Analysis of Recovered Files

All the tools recovered between 13,500 and 15,200 files from each of the three images, Image4, Image5 and Image6. The analysis below does not assume that any one tool provides an accurate baseline for the number of files to be recovered. Consequently, it is not possible to report the absolute proportion of files recovered by any one tool. Instead, the differences between the tools are investigated. Several reasons for the variations between tools were identified; these are discussed below.

The forensic suites, EnCase and FTK, recover space that is not allocated by the filesystem as multiple logical files. EnCase provides options for the user to specify the size of each file created from unallocated space while FTK automatically decides how to divide the unallocated space into files and names the files according to the cluster number.

File slack is the disk space between the end of the file content and the end of the last cluster in which the file is saved. FTK recovers file slack as files named:

```
<path>\<filename>.<extension>.FileSlack
```

FTK exported 1,244 slack space files from Image4, 1,171 from Image5 and 1,200 from Image6. EnCase exports slack space files for every item in the case, even when a file contains no data.

The data recovered from unallocated space and file slack is of interest to a digital forensic practitioner. These regions of a disk may contain remnants of deleted files. However, the research presented here excludes the recovery of data not managed by the filesystem, so these results are not included in the analysis.

Many filesystems support supplementary data attributes called alternate data streams (ADSs) in addition to the default stream [5]. An ADS can be used to store supplementary information about a file such as the zone identifier of a file downloaded from a web server. An ADS can also be used to store data in a manner that is not obvious to a casual browser of a filesystem. Not all filesystems support ADSs, so different recovery

tools present this data in different ways. EnCase and FTK recover ADSs that are used to denote zone identifiers as separate files. These files are named as follows:

```
<path>\<filename>.<extension>.<Zone.Identifier>
```

R-Studio recovers ADSs and incorporates them into the original file if the host filesystem supports them. Recuva and Stellar Phoenix do not recover ADSs.

Every directory in an NTFS filesystem contains \$I30, a directory index file that lists the directory files and sub-directories [5]. FTK recovered some of these files and labeled them \$I30. The other four tools did not recover directory index files.

FTK recovered files from the root folder into two folders: [root] and [root][1008]. After investigating this with the recovered files from Image4, we determined that no files were duplicated and that it was simply a matter of presentation. For the purposes of comparing the result across the tools, all the files were regarded as having been recovered to one root folder. It was also apparent that FTK had appended file ID numbers (e.g., 2708) to file names starting with the \$ symbol. These files were removed before analyzing the results any further.

The tools differ in their presentations of filenames for files that are in the user's Recycle Bin. EnCase and Recuva show the original filename whereas FTK, R-Studio and Stellar Phoenix show renamed files such as Dc1.xls and Dc5.avi. The naming convention is as follows:

```
D<original drive letter of file><#>.<original extension>
```

The mapping of the original filename to the renamed file is found in the INFO2 file, a normally-hidden file that is created the first time that the Recycle Bin is used [8].

These sources of data are of potential interest to a digital forensic practitioner, so it is desirable that they are included in a data recovery process. In addition, both sources of data are managed by the NTFS filesystem. Consequently, two analyses are presented below:

- All recovered files excluding unallocated space and file slack.
- All recovered files excluding unallocated space, file slack, Zone Identifier alternate data streams and \$I30 index files.

Figure 1 shows the numbers of files recovered from Image4, Image5 and Image6 by each of the recovery tools. Figure 1(a) compares the numbers of files recovered by each tool from the disk images. The figure also shows the total number of files recovered from each image using

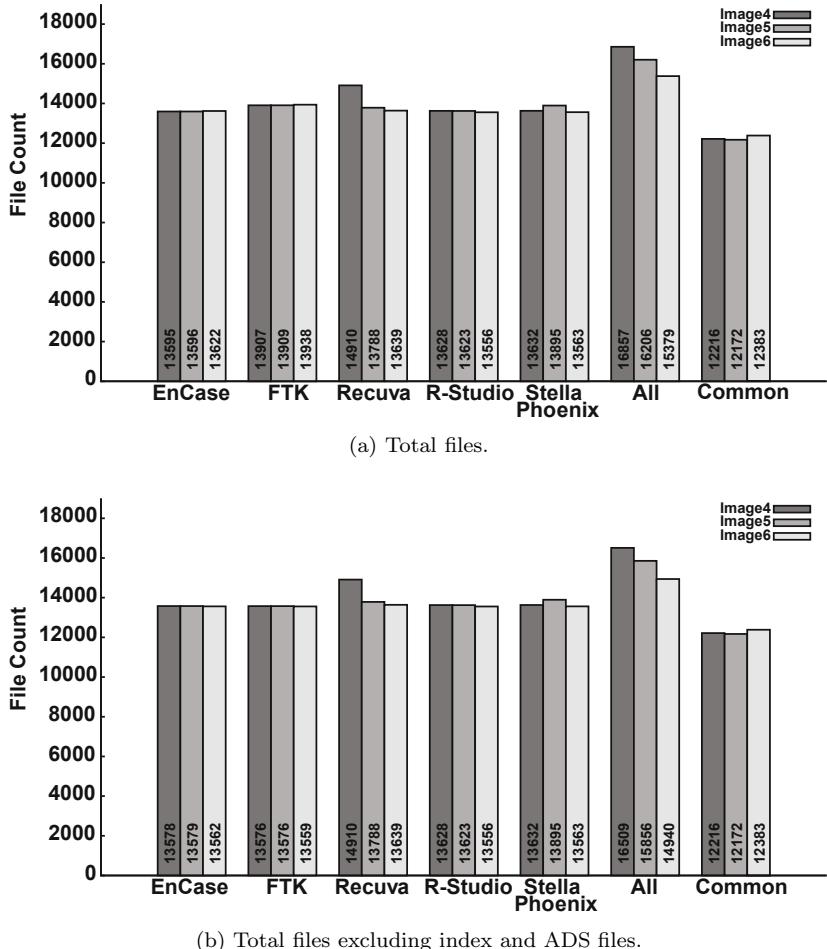


Figure 1. Files recovered from Image4, Image5 and Image6 using the tools.

all the tools, and the total number of common files recovered by all the tools from each image. Two files are considered identical if they have matching file paths (taking into account the adjustments described above) and matching MD5 hash values. Figure 1(b) shows the same totals, excluding the index and ADS files.

An analysis of the results reveals that, out of the total files recovered from a single image by all the tools, a single tool recovers between:

- 80.6% (EnCase from Image4) and 90.6% (FTK from Image6) when all the files are considered.
- 82.2% (FTK from Image4) and 91.3% (Recuva from Image6) when the index and ADS files are excluded.

Although the results show that no one tool recovers all the files found by all other tools, it is unclear from this analysis if the tools recover similar file types and file locations.

### 3.2 Analysis of User-Created Files

During the experiment, a number of marker files were copied to the disk so that their presence among the recovered files could be analyzed for each image. Some files were deleted by moving them to the Recycle Bin, and some of the files moved to the Recycle Bin were removed from it. Additional files were copied to the disk installation so that some previously-deleted files would be overwritten. This section analyzes the findings.

A total of 86 files were created within the user's My Documents folder or were copied to it. EnCase, FTK, Recuva and R-Studio successfully recovered all 86 files with hashes that matched the originals. Stellar Phoenix successfully recovered 84 of the 86 files. The remaining two files (Files 77 and 78) were substantially recovered, but two bytes in each file had been altered, so the MD5 hashes did not match. The recovery process from Image4 was repeated to confirm this result.

Table 1 summarizes the recovery of files from Image5 and Image6. A total of 83 files were deleted during the preparation of Image5. EnCase, FTK, Recuva and R-Studio successfully recovered 82 of the 86 files from Image5 with hashes matching the originals. Two files (Files 17 and 18) were not recovered at all by any of the tools and two (Files 2 and 7) were recovered, but their MD5 hashes did not match the originals. File 2, a text file, was missing some content part of the way through the file. File 7, an Excel spreadsheet, was recovered, but was substantially corrupted (approximately 40% of the bytes had been changed). Two files were not recovered by Stellar Phoenix from Image5, these were same files (Files 77 and 78) that were not successfully recovered from Image4. However, the file contents were changed at different locations compared with Image4, so the files recovered from Image4 and Image5 had different MD5 hashes and were different from the originals.

EnCase, FTK, Recuva and R-Studio successfully recovered 125 of the 169 files from Image6 with hashes matching the originals (the same files were recovered by each of these four tools). Stellar Phoenix successfully recovered 122 of the 169 files with hashes matching the originals. As in the case of Image5, two of the additional unrecovered files were text files. The third file, an ISO formatted disk image was recovered, but it had some additional content (2,048 bytes) at the end of the file compared with the original file.

In summary, EnCase, FTK, Recuva and R-Studio performed identically when recovering marker files from all three images (Image4, Image5 and Image6). Stellar Phoenix corrupted two bytes in each of two text files (even when these files had not been deleted) and added a padding of zeroes at the end of another file that had not been deleted.

### 3.3 Differences in Hashes due to Image Mounts

It was observed in that, under certain circumstances, different hash values were produced by two tools for identical files. This was due to the manner in which the image containing the file was mounted as a logical drive.

Mount Image Pro was used to mount images to enable data recovery for Recuva and Stella Phoenix. The MD5summer tool version 1.2.0.5 was used to compute hashes for files recovered using these tools [21]. Fourteen files from Image4 had different hashes computed by MD5summer after mounting the image with the Mount Image Pro's "Physical and Logical" mount option, compared with those computed by FTK Imager.

The "Mount File System" option for Mount Image Pro was also tested with MD5summer. When this option was used, twelve of the fourteen files that had different hashes for the "Physical and Logical" option agreed with those computed by FTK Imager. The hashes computed for the two other files matched those computed by MD5summer using the "Physical and Logical" option. However, the hashes of numerous other files did not match the hashes previously calculated using FTK Imager.

The Image4 file was also loaded into FTK Imager, both as a raw image and as a mounted drive. The mounting used Mount Image Pro with the "Physical and Logical" mount option. In both cases FTK Imager computed the same hash values for all the recovered files.

This analysis demonstrates that care must be exercised when using tools to mount a disk image as a filesystem. The way in which a disk image is mounted can result in different hashes being computed for the files in the disk image. Our future research will investigate the reason for these differences.

## 4. Related Work

Several authors have argued that software tools must be validated before they can be considered suitable for forensic purposes [4, 18]. NIST has created the CFTT Program [19], which develops test sets and methods for evaluating forensic software functions such as image acquisition and hashing. Mercuri [18] has commented on the apparent defects in

image acquisition functions in the EnCase and FTK forensic tool suites, as identified by the CFTT Program.

However, data recovery is perhaps an intrinsically harder function to validate than disk image creation. This is because software tools require judgments to be made about how the recovered data files are to be presented to a user. A comparative approach overcomes some of these problems by providing estimates of the differences between data recovery applications rather than setting a ground truth as an absolute standard.

Several authors have conducted empirical comparisons of digital forensic software and of software used for digital forensic purposes. Childs and Stephens [7] have assessed three Linux forensic tools, Vinetto, Pasco and `mork.pl`. Each of these tools is designed to perform a specific task and none is intended to fulfill the needs of a digital forensic practitioner who wishes to recover and analyze all the files from a device. The tool comparison conducted by Childs and Stephens is thus limited by the specific functions provided by each tool.

The use of digital forensic tools in an academic environment is discussed by Manson, *et al.* [17]. They compare the open-source Sleuth Kit [6] with EnCase and FTK, and measure the performance of each tool against prototype images designed by the authors. As in the case of our research, the disk images used were smaller than those found in typical computer systems. The images included a mobile phone SD card (size not disclosed), and 4 GB and 15 GB hard drives. A small number of files were added to the hard drives in a Windows XP SP2 installation. Manson, *et al.* used FTK (version 1.61a) and EnCase (version 5.05C), which have been superseded several times over in the intervening years. Their research concluded that open source tools provided the same results as commercial tools, although the usability of the open source tools varied and was difficult to measure [17].

Finally, Bariki, *et al.* [3] have proposed a standard for digital evidence to be used in reports generated using digital forensic tools. They surveyed the reporting functionality of three tools, including EnCase and FTK, and note the variations in the evidentiary items included in the reports. Their research concluded that a lack of standards leads to difficulty in producing quality reports for legal proceedings.

## 5. Conclusions

This research has compared the data recovery capabilities of five tools under identical conditions to assess the speed with which the tools complete the data recovery process and the extent of the variations between the tools in terms of the files recovered. No two tools produced identical

results, and no tool recovered all the files in a disk image (“all” is defined at the sum total of the distinct files collectively recovered by the tools). Of course, it is also possible that some files resident on the disk image were not recovered by any tool.

One conclusion is that digital forensic practitioners need to use multiple tools to obtain a higher proportion of files from a disk image. However, the variability of recovery tools raises concerns about the correctness of the results obtained. Specifically, different subsets of files are recovered by different toolkits, and the contents of the some recovered files differ. Also, the manner in which a recorded image is accessed by a recovery application can influence the results obtained. Data recovery of user-deleted files further complicates this problem. Therefore, digital forensic practitioners should take great care when relying on files recovered by a single tool.

Comparing the data recovery results of different forensic tools presents considerable challenges. Since the configuration options and user interface features were developed independently and recovered data is presented to the software and user in different ways, establishing equivalent configurations of the various forensic tools may not be possible.

The diversity of configuration options and presentation schemes is unsurprising due to the lack of an accepted standard for data recovery methods. Further research is required to understand the implications of these variations on the evidence produced in investigations. In particular, the extent to which the discrepancies between recovery methods can influence investigations must be better understood.

## References

- [1] AccessData, FTK, Linden, Utah ([accessdata.com/products/computer-forensics/ftk](http://accessdata.com/products/computer-forensics/ftk), 2011).
- [2] C. Ball, FTK 2.0: Product review, Electronic Data Discovery Update Weblog ([commonscold.typepad.com/eddupdate/2008/05/ftk-20-product.html#more](http://commonscold.typepad.com/eddupdate/2008/05/ftk-20-product.html#more)), May 8, 2008.
- [3] H. Bariki, M. Hashmi and I. Baggili, Defining a standard for reporting digital evidence items in computer forensic tools, *Proceedings of the Second International ICST Conference on Digital Forensics and Cyber Crime*, pp. 78–95, 2010.
- [4] B. Carrier, Open Source Digital Forensic Tools: The Legal Argument, White Paper, @Stake, Cambridge, Massachusetts, 2002.
- [5] B. Carrier, *File System Forensic Analysis*, Pearson Education, Upper Saddle River, New Jersey, 2005.

- [6] B. Carrier, The Sleuth Kit ([www.sleuthkit.org/sleuthkit](http://www.sleuthkit.org/sleuthkit)), 2011.
- [7] D. Childs and P. Stephens, An analysis of the accuracy and usefulness of Vinetto, Pasco and `mork.pl`, *International Journal of Electronic Security and Digital Forensics*, vol. 2(2), pp. 182–198, 2009.
- [8] M. Cross, *Scene of the Cybercrime*, Syngress, Burlington, Massachusetts, 2008.
- [9] Forensic Focus Blog, What happened to FTK2? ([forensicfocus.blogspot.com/2008/05/what-happened-to-ftk-2.html](http://forensicfocus.blogspot.com/2008/05/what-happened-to-ftk-2.html)), May 20, 2008.
- [10] S. Garfinkel, P. Farrell, V. Roussev and G. Dinolt, Bringing science to digital forensics through standardized forensic corpora, *Digital Investigation*, vol. 6(S), pp. S2–S7, 2009.
- [11] GetData, Mount Image Pro v4, Kogarah, Australia ([mountimage.com](http://mountimage.com)), 2011.
- [12] W. Glisson, T. Storer, G. Mayall, I. Moug and G. Grispos, Electronic retention: What does your mobile phone reveal about you? *International Journal of Information Security*, vol. 10(6), pp. 337–349, 2011.
- [13] Guidance Software, EnCase Forensic, Pasadena, California ([www.guidancesoftware.com/forensic.htm](http://www.guidancesoftware.com/forensic.htm)), 2011.
- [14] N. Harbour, `dcfldd` version 1.3.4-1 ([dcfldd.sourceforge.net](http://dcfldd.sourceforge.net)), 2006.
- [15] M. Hildebrandt, S. Kiltz and J. Dittmann, A common scheme for evaluation of forensic software, *Proceedings of the Sixth International Conference on IT Security Incident Management and IT Forensics*, pp. 92–106, 2011.
- [16] A. Jones, G. Dardick, G. Davies, I. Sutherland and C. Valli, The 2008 analysis of information remaining on disks offered for sale on the second hand market, *Journal of International Commercial Law and Technology*, vol. 4(3), pp. 162–175, 2009.
- [17] D. Manson, A. Carlin, S. Ramos, A. Gyger, M. Kaufman and J. Treichelt, Is the open way a better way? Digital forensics using open source tools, *Proceedings of the Fortieth Annual Hawaii International Conference on System Sciences*, pp. 266b, 2007.
- [18] R. Mercuri, Criminal defense challenges in computer forensics, *Proceedings of the First International ICST Conference on Digital Forensics and Cyber Crime*, pp. 132–138, 2009.

- [19] National Institute of Standards and Technology, Active File Identification and Deleted File Recovery Tool Specification, National Institute of Standards and Technology, Draft for Comment 1 of Version 1.1, Gaithersburg, Maryland, 2009.
- [20] National Institute of Standards and Technology, Computer Forensics Tool Testing Program, Gaithersburg, Maryland ([www.cftt.nist.gov](http://www.cftt.nist.gov)), 2011.
- [21] L. Pascoe, MD5summer ([md5summer.org](http://md5summer.org)), 2011.
- [22] SC Magazine, Forensic tools 2006, New York ([www.scmagazineus.com/forensic-tools-2006/group/test/37](http://www.scmagazineus.com/forensic-tools-2006/group/test/37)), July 11, 2006.
- [23] Where is Your Data? Weblog, Forensics: FTK 2 ([whereismydata.wordpress.com/2009/03/01/forensics-ftk-2](http://whereismydata.wordpress.com/2009/03/01/forensics-ftk-2)), March 1, 2009.