A Comparison of Forensic Toolkits and Mass Market Data Recovery Applications

Joe Buchanan-Wollaston, Tim Storer and William Bradley Glisson

October 22, 2012

Abstract

Digital forensic application suites are large, expensive, complex software products, offering a range of functions to assist in the investigation of digital artifacts. Several authors have raised concerns as to the reliability of evidence derived from these products. This is of particular concern, given that many forensic suites are closed source and therefore can only be subject to black box evaluation. In addition, many of the individual functions integrated into forensic suites are available as commercial stand-alone products, typically at a much lower cost, or even free. This paper reports research which compared (rather than individually evaluated) the data recovery function of two forensic suites and three stand alone 'non-forensic' commercial applications. The research demonstrates that, for this function at least, the commercial data recovery tools provide comparable performance to that of the forensic software suites. In addition, the research demonstrates that there is some variation in results presented by all of the data recovery tools.

1 Introduction

Forensic software suites are used by thousands of digital forensics professionals throughout the world. The functionality of forensic software suites varies, although several features appear to be consistently provided, including: hard disk image preparation and storage; data hashing of entire images or individual artifacts; disk image mounting and file system re-construction; data presentation and visualization; and data carving of damaged images or of deleted file contents.

Statistics on market share of forensic software tools appear to be a closely guarded secret. However, a review of online forums, corporate websites and the academic literature gives the *impression* that major vendors of forensic software for personal computers (PCs) are Guidance Software Inc. and AccessData Group LLC. These vendors market the EnCase (Guidance Software, 2011) and Forensic Tool Kit (FTK) (Access Data, 2011) software suites, respectively. These software suites are widely used, perhaps due to the integration of different 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 of these vendors provide compelling arguments for employing their software in digital forensic investigations. The Guidance Software website, for example, describes the EnCase suite as "the industry-standard computer forensics investigation solution" (Guidance Software, 2011), while AccessData claim that FTK is "the most advanced computer forensics software available" (Access Data, 2011). Similarly, both vendors have argued that their software products are designed and validated to meet standards of forensic evidence. FTK is described as a "court-validated digital investigations platform" (Access Data, 2011). EnCase is described similarly as having "an unsurpassed record of court acceptance" (Guidance Software, 2011). These claims are difficult to assess. It is unclear what level of scrutiny has been applied to the evidence produced by forensic toolkits, or what standards they have been assessed against.

Few independent reviews comparing the functionality and performance of forensic software suites exist. SC Magazine has conducted group tests of forensic software, typically considering around eight to ten products each year (SC Magazine, 2006). The National Institute of Standards and Technology (NIST) in the United States has developed the Computer Forensics Tools Testing (CFTT) program (NIST, 2011). In particular, the program has developed a

draft testing framework for data recovery tools for forensic purposes (NIST, 2009). To date, this framework has only been applied to a small selection of forensic software suites (Hildebrandt et al., 2011).

Anecdotal evidence suggests that forensic software suites may not be defect free. The release of FTK Version 2, in February 2008, received a considerable amount of bad press (Where is Your Data? Weblog, 2009; Forensic Focus Blog, 2008; Forensic Focus, 2008; Ball, 2008; Nelson, 2008). One review described the software as "an unmitigated disaster" (Where is Your Data? Weblog, 2009). The blog noted that users reported problems with installation and running the software. The documented minimum computer specification was also reported to be inadequate for the software. Mercuri (2010) has noted that CFTT program testing of both the EnCase and AccessData suites revealed defects in the hard disk image preparation processes. Both tools were unable to recover some data from NTFS formatted logical disk partitions.

These software applications incur a significant cost for forensic investigators. As of mid-2011, inquiries with the respective vendors established that a single user license for either EnCase or FTK is approximately $2,995 and the cost of software maintenance and support is charged at an additional $599 for EnCase and $840 for FTK (both fees charged annually). In addition, the computer hardware requirements for these tools impose significant further costs. For example, the recommended system specification for FTK (Version 3) includes an Intel i9 Dual Quad Core Xeon Processor, 12GB of RAM and a 160GB solid state hard drive dedicated entirely to an Oracle database for case management. The online shopping service provided by a popular vendor of IT equipment in the UK was used in mid-November 2011 to prepare an estimate for a machine with the minimum specification required. The estimate suggests that such a machine would cost approximately £2600 (approximately $4100), not including peripheral equipment or sales taxes.

Despite the popularity of commercial software forensic suites, numerous disparate software applications are available which provide equivalent functions for much lower, or even no cost. Assembling a toolkit of mass market applications which has equivalent functionality to a forensic suite could be an attractive proposition, not just in order to reduce costs. In addition, 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.

The novel contribution of the research presented here is a direct comparison of the results of the data recovery function from two commercial forensic products with that provided by several mass market applications. Data recovery is the extraction and presentation of file contents from a disk image formatted in a known file system. This definition excludes the recording of the disk image itself and also recovery of file contents stored in areas of the disk image that are not managed by a file system. Previous research has investigated methods for evaluating software for replicating disk images (NIST (2009) for example). The research presented here investigates the extent of agreement between the different forensic and mass market data recovery applications.

The paper is structured as follows. Section 2 describes the experimental setup for comparing data recovery functions of several different forensic and commercial tools. The experimental setup generates several disk images representing the evolution of the data stored as a result of user actions. These disk images are then processed using a selection of recovery tools. Section 3 describes the results of the comparison of the recovery tools. This analysis comprises a whole data-set comparison between tools, and an analysis of the recovery of a number of 'known' marker files deliberately added to the disk image. Section 4 discusses the findings from the experiment, and relates this to the previous literature on evaluation of forensic software. Finally, Section 5 draws some conclusions from the work relevant to forensic investigation and tool validation, and identifies some areas of future work.

2 Method

The purpose of this research was to compare the data recovery capabilities of a selection of software tools on a 'typical' desktop personal computer setup. We *assume* the typical applications of a personal computer to include office applications, web browsing, email communication and media playback. We note that some progress has been made in identifying *realistic* data sets for forensic tool analysis, based on data found on re-sold hardware (Garfinkel et al., 2009; Jones et al., 2009; Glisson et al., 2011). However, we are not aware of any research that establishes a *characteristic* test set. Consequently, we have chosen to develop and report our own data set to maintain control of the experiment.

In summary, the 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 on a user's 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 over-writing user-deleted files.

An image of the disk was taken at each stage of the experiment, referred to as Image1 through Image6, using FTK Imager. To gain assurance as to 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 dcfldd tool (Harbour, 2006). MD5 hashes of images from both dcfldd and FTK were computed using both the dcfldd and FTK tools. All four image hashes for Image6 matched. Files were then recovered from the images using a selection of data recovery tools. All files were hashed and file hashes and reported file system paths were analyzed for variations.

The following subsections provide more detail concerning the different phases of the experiment.

2.1 Target System Setup

A hard disk with a 20GB capacity was chosen for the installation as this would be sufficiently large to hold the operating system plus a variety of files including photos and videos whilst being relatively quick to image and process. Typical hard disks available in a new computer (as of 2012) are in the region of 500GB (laptop) and 1 to 2TB (desktop) however using such large disks would considerably extend the amount of time required for imaging and processing the disk multiple times. The hard disk was acquired from an IT hardware provider and an image was taken to check that the provider had scrubbed the disk before selling it (Image1) and this was found to be the case. However, the previous contents of the disk were scrubbed again using the dcfldd tool, available with most popular distributions of Linux. Another image of the disk (Image2) was then made. This step ensured that the state of the disk at the start of the setup was not dependent on processes undertaken by the hardware provider.

Windows XP Professional SP3 was installed on the disk as the operating system with a single user account. The operating system installation process formatted the target hard disk using the New Technology File System (NTFS). Software for a Netgear Wireless USB Adapter,

Internet Explorer 8, Firefox 5, Microsoft Office 2007 and Skype 5.5 were all installed. An image of the disk was taken (Image3). Windows was then activated.

2.2 Image Preparation with User 'Marker' Files

Internet Explorer and Firefox were opened and a number of websites were visited in each browser. In each case the URL was typed directly into the browser's address bar rather than being accessed via a search engine or other link. The sites visited were selected as 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 easily identified.

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 back from the same contact. This would create 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 would result in user content being stored in the outlook.pst data file.

The next step was to open Windows Explorer and create a new folder within the 'My Documents' folder. Within this folder a new blank text document was created, some text was added and the file was saved. A number of files were prepared and saved on an external hard drive. These encompassed a variety of file types that might be found on a user's computer including word processing documents, spreadsheets, PDF documents, photographic images, plain text files, audio, video and executable program files. The external hard drive was connected to the computer via a USB cable and the files copied across to folders within the 'My Documents' folder. A new Microsoft Word document was then created, some text added and the file saved. A new Microsoft Excel document was also created. A total of 86 user files were added to the 'My Documents' folder. Table 1 summarises the different file types and locations that were added to the disk. An image of the disk (Image4) was made at this stage using FTK Imager.

A number of the files that had been copied across to the disk were deleted, with 81 being 'moved' to the user's Recycle Bin. Of these files, 42 were then removed from the Recycle Bin.

file IDs	file type	File Extension	Added before Image4	Left in Recycle Bin for Image5	Deleted and Removed from Recycle Bin	Permanently deleted in Image4	Altered before Image6	Added before Image6	EnCase	FTK	Recuva	R-Studio	Stella Phoenix	EnCase	FTK	Recuva	R-Studio	Stella Phoenix
1	A - Created Documents	docx	✓		✓				✓	✓	✓	✓	✓					
2	A - Created Documents	txt	✓		✓													
3	A - Created Documents	xlsx	✓		✓				✓	✓	✓	✓	✓					
4,5	Excel Docs	xls	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6	Excel Docs	xls	✓				✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	Excel Docs	xls	✓		✓													
8	Excel Docs	xls	✓		✓				✓	✓	✓	✓	✓					
9-12	Movies	AVI	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
13-17	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	PDFs	pdf	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
42-46	PDFs	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 Docs	doc	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
83	Word Docs	doc	✓				✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
84,85	Word Docs	doc	✓		✓				✓	✓	✓	✓	✓					
86	Word Docs	docx	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
87	A - Created Documents	docx						✓						✓	✓	✓	✓	✓
88	A - Created Documents	xlsx						✓						✓	✓	✓	✓	✓
89	Excel Docs	xls						✓						✓	✓	✓	✓	✓
90-93	Excel Docs	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	PDFs	pdf						✓						✓	✓	✓	✓	✓
130-159	Photos	jpg						✓						✓	✓	✓	✓	✓
160-164	Text files	txt						✓						✓	✓	✓	✓	✓
165	Word Docs	doc						✓						✓	✓	✓	✓	✓
166-169	Word Docs	docx						✓						✓	✓	✓	✓	✓

Table 1: Summary of file manipulations made to Image4 through Image6.

Two files were reported as being 'permanently deleted' by a Windows prompt, as these files were too large for the Recycle Bin. The browsing history was deleted from Internet Explorer and Firefox. Calendar items and all emails were deleted from Outlook, and the 'Deleted Items' folder was then emptied in Outlook. The history was cleared from Skype. Approximately half of the files in the Recycle Bin were deleted, with a record being kept of which files remained. The disk was then imaged again (Image5).

The external hard drive was reconnected to the computer and another selection of 81 pre-prepared files copied across to sub-folders in the 'My Documents' folder. The total size of the files copied had been calculated so as to almost fill the remaining space on the disk whilst leaving a small amount of space for files created by the operating system during use and by Internet browsing. This replicates behavior in which a user fills up a hard disk with data and then must remove some old data to free up space. Firefox and Internet Explorer were each opened and a number of websites were visited in each browser, again by typing the URL directly into the address bar. Two new documents were created in each of Microsoft Word and Excel, text was added and the files saved. Further use of Skype and Microsoft Outlook was made and the disk imaged again (Image6).

2.3 Data Recovery Procedures

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)

All five tools were installed on a single HP Z400 workstation with Intel Xeon Dual Core W3503 Processor (2.40 GHz), 8GB RAM, with a 750GB hard disk, running Windows 7 Enterprise 64-bit. The disk images were all stored on the same 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 a direct comparison of the tools challenging. The configuration of each of the tools is presented here as a narrative to support experimental repeatability. Not all option configurations can be described exhaustively. The narrative records where non-default options (as presented to the user) are selected for an application.

For both of the forensic suites, the first step in processing the evidence is to start a new case. A case can contain all the evidence, bookmarks, information and reports and allows searching through the evidence. The three mass market data recovery tools have no case management options. In the forensic suites, the number of options that can be chosen before the scanning and recovery processes can be run is far greater than for a tool that only recovers data.

Three of the tools, EnCase, FTK and R-Studio are designed to allow a raw disk image to be loaded directly into the software whilst the other two tools, Recuva and Stellar Phoenix do not offer this facility. For these tools, the image must first be mounted in Windows as a logical drive, which was achieved using Mount Image Pro Version 4.48(828) software (Get Data, 2011).

Configuration of each of the tools was 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 processing was started.

FTK A new case was created and the image was added as evidence. All options were disabled except the generation of MD5 hashes and the process 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 selected.

Stellar Phoenix The 'Search Drive' dialog was opened for the image, and the 'Physical Drive' method was used. The options selected were 'Deep Scan' and 'Advanced Scan'.

3 Results

This section presents the results of several different analyses of the data recovered during the processes described in Section 2. The results reported in the following sections concern data recovered from Image4 through Image6, since these images represent the state of the target hard disk after user activity had been simulated.

3.1 Analysis of all Recovered Files

All of the tools recovered between 13,500 and 15,200 files from each of the three images Image4, Image5 and Image6. The analyses below does not assume that any one tool provides an accurate base line 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 extent of differences between tools is investigated. Several reasons for the variation between tools were identified, as discussed below.

The forensic software suites, FTK and EnCase, recover space not allocated by the file system as multiple logical files. EnCase provides options for the user to specify the size of each file created from unallocated space, whereas FTK automatically decides how to divide up unallocated space and files are then named according to the cluster number.

File slack is the disk space between the end of a file's contents and the end of the last cluster where the file is saved. FTK recovers file slack as files named:

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

FTK exported 1244 slack space files from Image4, 1171 from Image5 and 1200 from Image6. EnCase exports slack space files for every item in the case, even when the resulting file contains no data. The data that can be recovered from unallocated space and file slack is of potential interest to a digital forensic investigator. These regions of a disk may contain remnants of content from files that have been deleted by a user. However, the research presented here excludes the recovery of data not managed by the file system, so these are not included in the analysis.

Many file systems support the addition of one or more supplementary data attributes, known as alternate data streams (ADS) in addition to the default stream (Carrier, 2005). An ADS

can be used to store supplementary information about a file aside from the file's core contents, such as recording the zone identifier of a file downloaded from a web server.

An ADS can also be used to store data in a manner which is not obvious to a casual browser of a file system. Not all file systems support ADSs, so different recovery tools present this data in different ways. EnCase and FTK recover ADS which are used to denote zone identifiers as separate files named:

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

R-Studio recover ADSs and incorporate them into the original file if the host file system supports them. Recuva and Stella Phoenix do not recover ADSs.

Every directory in an NTFS file system contains a directory index file, $I30, listing the directory's file contents and sub-directories (Carrier, 2005). FTK recovers some of these files and labels them `$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 it was determined that no files are duplicated and it is simply a matter of presentation. For the purposes of comparing the result across the tools, all these 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' on file names starting with a $ symbol. These have been removed before analyzing the results further.

The tools differ in their presentation 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 can be found in the `INFO2` file, a normally hidden file created the first time the Recycle Bin is used (Cross, 2008).

All of these sources of data are of potential interest to a forensic investigator, so it is desirable that it is included by a data recovery process. In addition, both sources of data are managed by the NTFS file system under analysis. Consequently, two analyses are presented below:

1. All recovered files excluding unallocated space and file slack.

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

Figure 1 illustrates the number of files recovered from Image4, Image5 and Image6 by each of the recovery tools. Figure 1(a) compares the number of files recovered by each tool from each of the disk images recorded. The figure also illustrates the total number of files recovered from each image using all of the tools, and the total number of files recovered in common 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. Figure 1(b) illustrates the same totals, except that files representing ADS content, and index files are excluded.

Figure 1 shows the number of files recovered by the different tools from the total data set for each of the three images examined. The analysis shows that out of the total files recovered from a single image by all tools, any single tool recovers between:

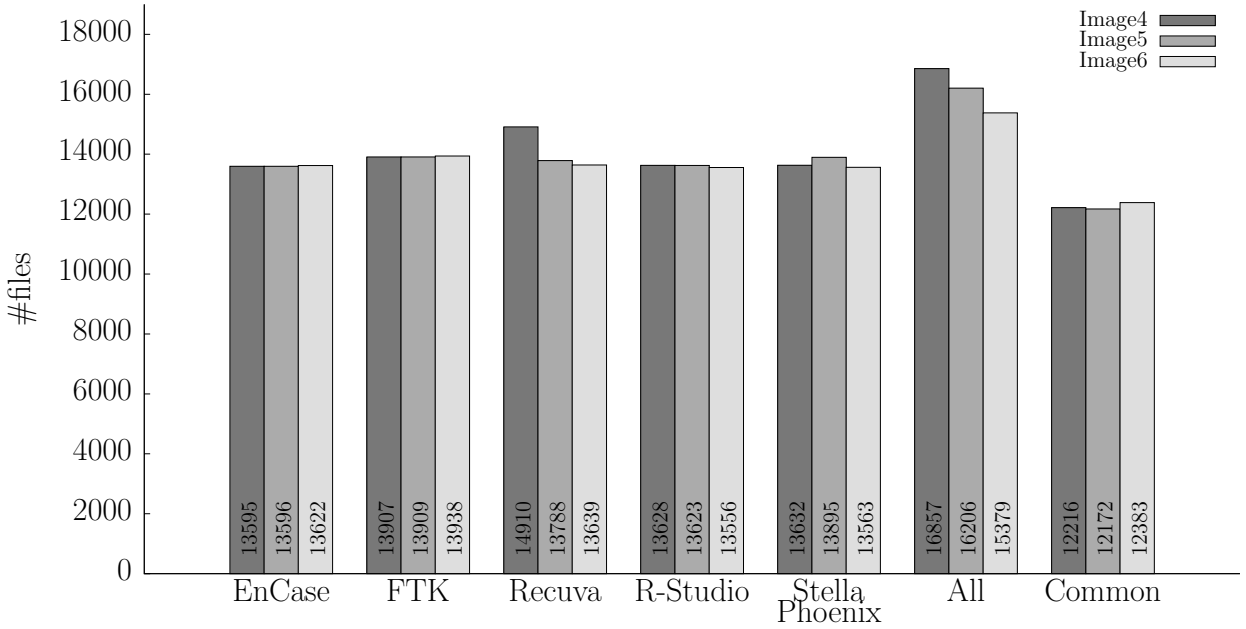
- 80.6% (EnCase from Image4) and 90.6% (FTK from Image6) when all files are considered; and
- 82.2% (FTK from Image4) and 91.3% (Recuva from Image6) when index and ADS files are excluded.

Although the figure illustrates that no one tool recovers all the files found by all other tools, it is unclear from this analysis whether certain tools are recovering similar types or locations of files.

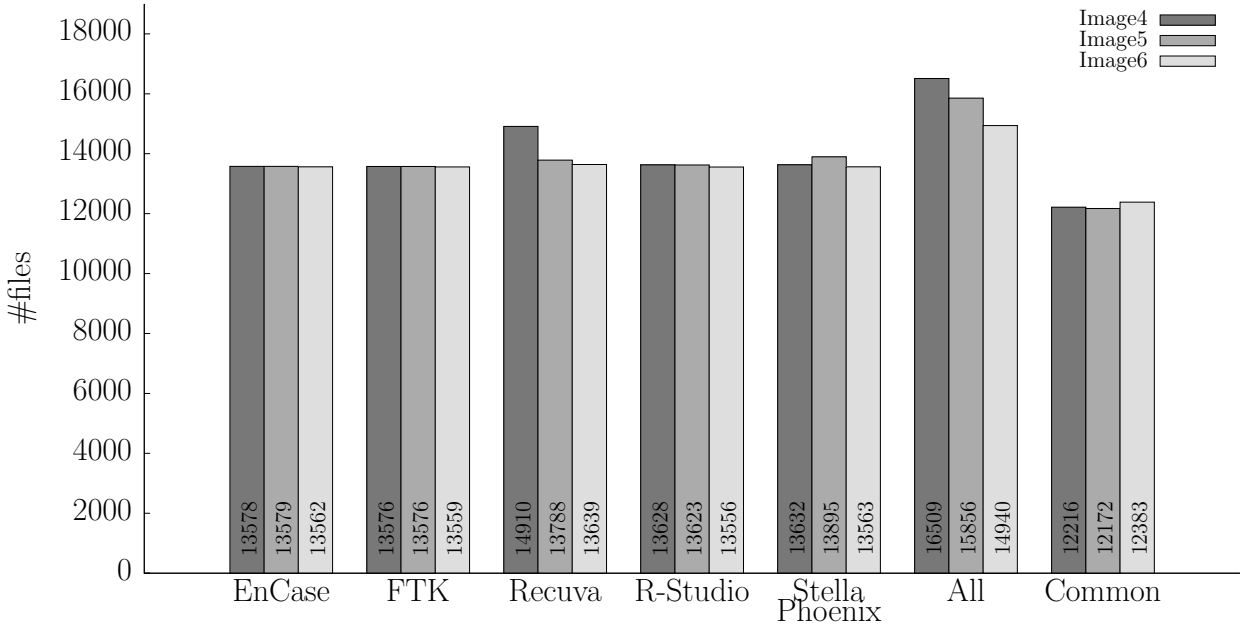
3.2 Analysis of User Created Files

As described in Section 2.2, a number of marker files were copied to the disk during the experiment so that their presence in the recovered files could be analyzed for each image. Some of these files were deleted and of those deleted to the Recycle Bin, some were removed from it. Further files were copied to the disk installation so that some previously deleted files might be over-written. The following sections present the findings of this analysis.

A total of 86 files were created within or copied to the user's 'My Documents' folder. EnCase, FTK, Recuva and R-Studio successfully recovered all 86 files with hashes matching



(a) total files



(b) excluding index and ADS content

Figure 1: Total files recovered from Image4, Image5 and Image6 by the different recovery tools.

the originals. Stellar Phoenix successfully recovered 84 of the 86 files. The remaining two 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 summarises the recovery of files from Image5 and Image6. A total of 83 files were deleted during the preparation of Image5, as described in Section 2.2. EnCase, FTK, Recuva and R-Studio successfully recovered 82 of the 86 files from Image5 with hashes matching those of the originals. Two files were not recovered at all by any of the tools (Files 17 and 18) and two were recovered but their MD5 hashes did not match the originals (Files 2 and 7). File 2, a text file was missing some content part way through the file. File 7 an Excel spreadsheet, was recovered, but was substantially corrupted (approximately 40% of bytes in the file had changed).

It is also noted that the tools display the names of files from the Recycle Bin in different ways, making comparison of results between tools more challenging. For example, File 2, a deleted text file is recovered as 'Dc79.txt' and File 7, a deleted Excel spreadsheet document is recovered as 'Dc2.xls' by FTK and R-Studio.

Two further files were not recovered by Stellar Phoenix, which were the same files as 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, as well as being different from the original.

EnCase, FTK, Recuva and R-Studio successfully recovered 125 of the 169 files with hashes matching the originals from Image6 (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 described for Image5, two of the additional unrecovered files were text files. The third file, an ISO formatted disk image was recovered but had some additional content (2048 bytes) at the end of the file compared to the original.

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

3.3 Differences in File Hashes due to Image Mounts

It was observed in that, under certain circumstances, different hashes could be produced by two tools for identical files, due to the way the image containing the file is mounted as a logical drive.

As described in Section 2.3, 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 (Pascoe, 2011). 14 files from Image4 had different hashes as computed by md5summer after mounting the image with the Mount Image Pro '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. 12 of the 14 files which had different hashes calculated using the 'Physical and Logical' option now 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' mount option. However, hashes produced for numerous other files under these conditions 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, having been mounted using Mount Image Pro, with the 'Physical and Logical' option selected. In both cases FTK Imager calculated the same file hashes for all files recovered.

In summary, this investigation demonstrates that care must be exercised when using tools to mount a disk image as a file system. The way in which it is mounted can result in different hashes being obtained for files that are contained within a disk image. Further investigation as to the cause of these differences is left for future work.

4 Discussion and Related Work

Several authors have argued that software tools must be validated before they can be considered suitable for forensic purposes (Mercuri, 2010; Marsico, 2004; Carrier, 2002). The National Institute of Standards and Technology have developed the Computer Forensics Tools Testing Program (NIST, 2009). The program develops test sets and methods for evaluating different functions of forensic software such as image acquisition and hashing. Mercuri (2010) has com-

mented on apparent defects in image acquisition functions in the EnCase and FTK forensic tool suites, as identified by the program.

However, data recovery is perhaps an intrinsically harder function to verify than creating an image of a disk, because the design and implementation of software tools requires judgments to be made as to how recovered data files are to be presented to a user. However, several authors have conducted empirical comparisons of digital forensic software, or software used for digital forensic purposes.

Childs and Stephens (2009) assess three specific Linux forensic tools: Vinetto, Pasco and Mork.pl. Each of these tools has been developed to perform a specific task and none is intended to fulfill the needs of the forensic investigator wishing to recover and analyze all files from a device. This comparison is thus limited in scope by the specific abilities of each of the tools assessed.

The use of computer forensics tools in an academic environment is discussed by Manson et al. (2007). They compare the open-source tool, The Sleuth Kit (Carrier, 2011) with EnCase and FTK and measure the performance of each against the same prototype images designed by the authors. As with the work presented in this paper, the disk images employed were smaller than those found in 'typical' computing hardware. The images used were an SD card (size not disclosed) from a phone, a 4GB and a 15GB hard drive. In the latter two cases, a small number of files were added to a Windows XP SP2 installation. The research also used versions of FTK (version 1.61a) and EnCase (version 5.05C) which have been superseded several times over in the intervening years. The work concluded that open source tool provided the same results as the closed-source major vendor commercial tools examined, although the usability of the open source software varies and is difficult to measure (Manson et al., 2007).

A brief survey of computer forensic tools is presented by Arthur and Venter (2004) who compare one freeware tool, PC Inspector File Recovery (PC Inspector, 2011) with FTK and EnCase. The authors admit that the versions of FTK and EnCase used were for demonstration purposes, with limited functionality, therefore the paper cannot form a conclusion based on using the full commercial versions of the software.

Hildebrandt et al. (2011) proposed a common scheme for the evaluation of forensic software. They note that evaluation of forensic software by NIST is restricted to testing the

quality of acquired images and that other features required for the forensic process are not tested. Hildebrandt et al. also note that there is no profile developed within the Common Criteria (a framework for evaluating the security properties of different software and information technology products) for forensic software (NIST, 1999).

Finally, Bariki et al. (2010) propose a standard for digital evidence to be used in reports generated using computer forensic software tools. They survey the reporting functionality of three tools, including EnCase and FTK, and note the variation in digital evidence items that are included in the reports. Their research concludes that a lack of standards leads to difficulties in producing reports that can be presented in a court of law.

5 Conclusions

This research compared the data recovery capabilities of five tools under identical conditions, to assess the speed with which the tools complete the process and the extent of variation between the toolkits in terms of files recovered. No two tools produced identical results, and no one tool recovered all files found on a disk image by all the tools combined. Further, it is possible that some files were resident on the disk image that were not recovered by any tool.

One conclusion that may be drawn is that forensic investigators need to use multiple, diverse tools to obtain a higher proportion of files from a disk image. However, as the results reported in Section 3 illustrate, the very variability amongst recovery tools should raise concerns as to the correctness of results. These results show that different subsets of files are recovered by different toolkits, and that the contents of recovered files can also sometimes differ. The results also demonstrated that the manner in which a recorded image is accessed by a recovery application can influence the results gathered. Care should be taken when relying on files that were only recovered by a single tool. Data recovery of user-deleted files further complicates this problem.

More widely, we note from this that, for the purposes of data recovery, comparing the results produced by different forensic software applications presents considerable challenges. The configuration options and user interface features of the different applications have been independently developed, meaning that establishing an equivalent configuration of two or more toolkits is difficult. Data that is recovered is presented to the software user in different ways,

depending on the design choices made by the software vendors.

This diversity of configuration options and presentation is unsurprising due to the lack of a widely accepted standard for data recovery methods amongst either digital forensic practitioners or academic researchers.

References

Access Data. Forensic Toolkit product website. `http://accessdata.com/products/computer-forensics/ftk`. Accessed 9th November, 2011, 2011.

K.K. Arthur and H.S. Venter. An investigation into computer forensic tools. In H.S. Venter, M. Coetzee, and L. Labuschagne, editors, *4th Annual Information Security South Africa Conference*, Johannesburg, South Africa, July 2004.

Craig Ball. FTK 2.0: Product review. Electronic Data Discovery Update Weblog. `http://commonscold.typepad.com/eddupdate/2008/05/ftk-20-product.html#more`. Accessed 10th November, 2011, May 2008.

Hamda Bariki, Mariam Hashmi, and Ibrahim Baggili. Defining a standard for reporting digital evidence items in computer forensic tools. In Ibrahim Baggili, editor, *Digital Forensics and Cyber Crime. Second International ICST Conference, ICDF2C 2010. Revised Selected Papers*, volume 53 of *Lecture Notes of the Institute for Computer Science, Social Informatics and Telecommunications Engineering*, pages 78–95, Abu Dhabi, United Arab Emirates, October 2010. Springer Verlag.

Brian Carrier. Open source digital forensic tools: The legal argument. White paper, @Stake, October 2002.

Brian Carrier. *File System Forensic Analysis*. Addison-Wesley Professional, Pearson Education Inc, One Lake Street, Upper Saddle River, NJ 07458, USA, first edition, 2005.

Brian Carrier. Sleuthkit product website. `http://www.sleuthkit.org`. Accessed 10th November, 2011, 2011.

Dave Childs and Paul Stephens. An analysis of the accuracy and usefulness of Vinetto, Pasco and Mork.pl. *International Journal of Electronic Security and Digital Forensics*, 2(2):182–198, 2009.

Michael Cross. *Scene of the Cybercrime*. Syngress Media, Elsevier Inc. 30 Corporate Drive, Burlington, MA 01803, second edition, 2008.

Forensic Focus. FTK 2 - any verdicts yet? Discussion forum. Forensic Focus Community Website. `http://www.forensicfocus.com/index.php?name=Forums&file=viewtopic&t=2475`. Accessed 10th November, 2011, May 2008.

Forensic Focus Blog. What happened to FTK2? Forensic Focus Blog `http://forensicfocus.blogspot.com/2008/05/what-happened-to-ftk-2.html`. Accessed 10th November, 2011, May 2008.

Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. Bringing science to digital forensics through standardized forensic corpora. *Digital Investigation*, 6(S1):S2–S11, – 2009. Supplement 1, The Proceedings of the Ninth Annual DFRWS Conference.

Get Data. Mount Image Pro v4 application website. `http://mountimage.com/`. Accessed 23rd February, 2011., 2011.

William Bradley Glisson, Tim Storer, Gavin Mayall, Iain Moug, and George Grispos. Electronic retention: What does your mobile phone reveal about you? *International Journal of Information Security*, 10(6):337–349, 2011.

Guidance Software. EnCase Forensic product website. `http://www.guidancesoftware.com/forensic.htm`. Accessed 9th November, 2011, 2011.

Nick Harbour. DoD Computer Forensics Laboratory DD (ddcfldd) application website. `http://dcfldd.sourceforge.net/`. Accessed 31st July 2011., December 2006.

Mario Hildebrandt, Stefan Kiltz, and Jana Dittmann. A common scheme for evaluation of forensic softwarwe. In *Sixth International Conference on IT Security Incident Management and IT Forensics*, pages 92–106, Stuttgart, Germany, May 2011. IEEE Computer Society.

Andy Jones, Glenn Dardick, Gareth Davies, Iain Sutherland, and Craig Valli. The 2008 analysis of information remaining on disks offered for sale on the second hand market. *Journal of International Commercial Law and Technology*, 4(3):162–175, 2009.

Dan Manson, Anna Carlin, Steve Ramos, Alain Gyger, Matthew Kaufman, and Jeremy Treichelt. Is the open way a better way? Digital forensics using open source tools. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, Big Island, Hawaii, USA, January 2007. IEEE Computer Society.

Christopher V. Marsico. Computer evidence v. Daubert: The coming conflict. Technical Report 2005–17, Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, Indiana, 47907-2086, USA., March 2004.

Rebecca Mercuri. Criminal defense challenges in computer forensics. In Sanjay Goel, editor, *Digital Forensics and Cyber Crime. Proceedings of the First International ICST Conference, ICDF2C 2009*, volume 31 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 132–138, Albany, NY, USA, September/October 2010. Springer.

Sharon D. Nelson. Craig Ball beats me to it: The disaster that is FTK 2.0. Published online. `http://ridethelightning.senseient.com/2008/05/index.html`. Accessed 17th November, 2011, May 2008.

NIST. *Common Criteria for Information Technology Security Evaluation Part 1.* National Institute of Standards and Technology, 2.1 edition, August 1999.

NIST. *Active File Identification and Deleted File Recovery Tool Specification.* National Institute of Standards and Technology, US Department of Commerce, draft for comment 1 of version 1.1 edition, March 2009.

NIST. Computer forensics tool testing programme. `http://www.cftt.nist.gov/`. National Institute of Standards and Technology. Accessed 10th November, 2011, April 2011.

Luke Pascoe. Md5 summer application website. `http://md5summer.org`. Accessed 12th August, 2011., 2011.

PC Inspector. PC Inspector product website. http://www.pcinspector.de. Accessed 9th November, 2011, 2011.

SC Magazine. Forensic tools group test 2006. Available at: http://www.scmagazineus.com/forensic-tools-2006/grouptest/37/. Accessed 15th September, 2011, July 2006.

Where is Your Data? Weblog. Forensics: FTK 2. Where is Your Data? Weblog. Published online. whereismydata.wordpress.com/2009/03/01/forensics-ftk-2/. Accessed 10th November, 2011, March 2009.