

# AN INVESTIGATION INTO COMPUTER FORENSIC TOOLS<sup>†</sup>

K.K. Arthur<sup>1</sup> H.S. Venter<sup>2</sup>

Information and Computer Security Architectures (ICSA) Research Group  
University of Pretoria  
Pretoria

Department of Computer Science  
University of Pretoria  
Pretoria, 0002

Telephone: +27 (0) 12 420 250

<sup>1</sup>E-mail: [kweku@tuks.co.za](mailto:kweku@tuks.co.za)

<sup>2</sup>E-mail: [hventer@cs.up.ac.za](mailto:hventer@cs.up.ac.za)

## ABSTRACT

Cyber-crime has reached unprecedented proportions in this day and age. In addition, the internet has created a world with seemingly no barriers while making a countless number of tools available to the cyber-criminal. In light of this, Computer Forensic Specialists employ state-of-the-art tools and methodologies in the extraction and analysis of data from storage devices used at the digital crime scene. The focus of this paper is to conduct an investigation into some of these Forensic tools eg. Encase®. This investigation will address commonalities across the Forensic tools, their essential differences and ultimately point out what features need to be improved in these tools to allow for effective autopsies of storage devices.

## KEY WORDS

Forensic Tools, Computer Forensic Specialist, Storage Device, Data Recovery, Data Integrity, Information Security.

---

<sup>†</sup> This material is based upon work supported Telkom, IST and the NRF through THRIIP. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the Telkom, IST and the NRF do not accept any liability thereto.

# AN INVESTIGATION INTO COMPUTER FORENSIC TOOLS

## 1 INTRODUCTION

The internet is a network of networks, connecting millions of computing devices [1, p1], and has applications in business, communications and information interchange throughout the world. Undoubtedly, the advent of these connections has impacted all aspects of our lives. The decentralized nature of the internet forms its very foundation, yet ironically, this nature has opened networks and individual machines to a host of threats and attacks from cyber-criminals.

Cyber-crime includes, but is not limited to, the theft of trade secrets, theft of or destruction of intellectual property and fraud. Trade secrets and intellectual property is typically the foundation upon which many companies are built. This information gives each company a competitive advantage and to have such information compromised in any way could easily cost the company millions. In addition, since money is no longer exclusively paper based due to online trading, financial fraud such as credit card misuse is propagated once a criminal gains access to enterprise information systems. Cyber obscenity is one of the more popular forms of cyber crime. Essentially, pornographic material, such as child pornography, is hidden on storage media since perpetrators acknowledge the illegality of being in possession of these images.

Cyber-criminals associate themselves with one of or all of these crimes by making it their jobs to find vulnerabilities in operating systems, applications or services that run on a computer connected to the internet [2]. Once a vulnerability is discovered and exploited, the criminal is able to view or store sensitive information on some form of storage media. The storage medium can either be local, i.e. hard-drives or removable, i.e. floppy disks, zip drives, memory sticks or CDs. Once the crime is committed, prosecution becomes extremely difficult since the crime venue could easily be in different cities and countries and involve unsuspecting third parties. At this point, a computer forensic specialist (CFS) is tasked to investigate the digital crime scene by impartially scrutinizing a number of digital sources that are either involved or thought to be involved in the crime, and ultimately produce a single document reflecting a summary of the contents of the digital source.

Like any other forensic science, CFSs make use of a number of specialized software tools and hardware devices to carry out investigations. These investigations follow a strict methodology to maintain the credibility and integrity of all storage devices involved. The general methodology is to [3]:

- **PROTECT** the subject computer system during the forensic examination from any possible alteration, damage, data corruption or virus infection.
- **DISCOVER** all files on the subject system which includes existing normal files, deleted yet remaining files, hidden files, password-protected files and encrypted files.
- **RECOVER**, as much as possible, files that are discovered to be deleted.
- **REVEAL**, to the extent possible, the contents of hidden files as well as temporary files used by both the application programs and the operating system.
- **ACCESS** the contents of protected or hidden files if possible and legally appropriate.
- **ANALYZE** all relevant data found in special areas of the disk. The concept of special areas of a disk is explained later in section 3.
- **PRINT** out an overall analysis of the subject computer system. This analysis includes a listing of all relevant files and discovered file data. The print-out also provides an overview

of the system layout, file structures and data authorship information. Any attempts to hide, delete, protect or encrypt information will also be revealed through the print-out.

- **PROVIDE EXPERT CONSULTATION** and/or testimony as required. This testimony would typically be required to prove the points of a case in a court of law.

The subject of this paper will be those tools involved in each step of the, above mentioned, forensic methodology. The functionalities offered by the tools will also be discussed to offer a better understanding into the forensic process.

These tools generally differ in functionality, complexity and cost. In terms of functionality, some tools are designed to serve a single purpose [4] while others offer a suite of functions. Therefore, the functionalities offered by a tool are exactly what lead to its complexities. These complexities can either be related to design and algorithmic complexity or ease-of-use; in some instances, a tool can offer great functionality but fall short because of a complex interface. Cost is the final distinguishing factor. Some of the market-leading commercial products cost thousands of dollars while other tools are completely free [4]. With these limiting factors (functionality, complexity, and cost) in mind, the computer forensic expert now needs to evaluate the criticality of the crime and choose an appropriate tool(s) to help with his/her investigation.

In the remainder of this paper, a brief background to computer forensics is given. An explanation of some terms and concepts is given thereafter. The paper then offers an overview of forensic tools by identifying some functionalities and how they are achieved. The paper then also identifies differences between the evaluated tools. Finally, some findings are presented with suggested future add-ons for these tools and a brief conclusion is given.

## **2 BACKGROUND**

The term “Computer Forensics” was coined back in 1991 in the first training session held by the International Association of Computer Investigation Specialists (IACIS) in Portland, Oregon [5]. This science deals with the preservation, identification, extraction and documentation of computer evidence, and like any other forensic science, relates law and science.

In this day and age, the majority of correspondence is not paper based. Even when hard-copies of information are distributed, the probability that a soft copy still exists on the author’s computer is very high. As previously suggested in the introduction, if the author is found or suspected of distributing sensitive information, then forensic tools will be used to examine the author’s machine. As described by Sommer [14], acquiring a copy (image) of a disk would be the first essential step in evidence preservation. However, with standard hard disk capacities of 80GB and increasing storage media sizes, the imaging and examination processes will inevitably take longer. This is the basis for CFSs worries concerning increasing storage capacities.

With some crimes occurring between countries, dates and times become relevant to an investigation. As a result, the ability to associate a suspect to a crime through date and time evidence is a current field of study. Boyd and Forster [16], tell of an investigation that began when an e-mail trace identified an individual suspected of involvement in the communication of child abuse images. The investigation proceeds where the police obtain a warrant to seize the suspect’s computer equipment. The police and prosecution service then planned their case study while the defence made use of a CFS to comment on the digital evidence. When the defence presented their report to the prosecution, it had a number of allegations of malpractice by the police. Apparently, the seized computer was used while in police custody. This would inevitably tamper with the digital evidence by compromising the integrity of the data. This example illustrates how important an investigation methodology is, and how a CFS should be involved whenever evidence is digital. It also shows how the improper handling of evidence could affect time and date stamps [16] and hence, cause forensic tools to report inaccurate details of evidence.

Conversations with seasoned practitioners suggest that digital forensic practice is in a period of redefinition [15]. It no longer has to be associated with the examination of “conventional” storage media. Forensic examination can now be conducted on devices such as routers, personal digital assistants (PDAs) and digital cameras [4, 15]. With these developments, current forensic tools need to adapt to the changing environment or new tools need to be developed. Ultimately, forensic techniques and tools need to be found to keep CFSs ahead of the criminals who are seeking to hide from the digital forensic community pursuing them [15].

In order to have a better understanding of computer forensic tools, some CF terms and concepts are discussed below.

### **3 TERMS AND CONCEPTS**

The following section deals with concepts within computer forensics to provide a better understanding into the functionalities discussed later in the paper.

#### **3.1 Ambient Data**

In the mind of a normal computer user, once data is deleted, it is accepted that it is no longer in existence. On the other hand, the CFS should understand that the data could still exist in some other form or area on the storage medium. For example, file allocation differs between Windows 95 and Windows 2000. Windows 95 and Windows 2000 operating systems use FAT and NTFS [17, p554] filing systems respectively. This knowledge helps an investigator since he (she) would understand how data is “removed” and stored on storage media by the operating system (OS) in question. Thus, ambient data relates to the analyzing step of the computer forensic investigation methodology as given in the introduction. The analysis of all areas of storage media is important since they could be the site of relevant evidence.

In the Computer Forensics community, ambient data is used to describe data stored in unallocated space and file slack [11]. These two concepts are described in the following two sections.

##### **3.1.1 Unallocated Space**

When data is deleted, it is the reference to the data within the File Allocation Table that is actually deleted. That is, the data may still exist on the storage medium but the OS will not know how to access this data. From this point onwards, deleted data or files will be referred to as unreferenced data or files. Within Windows 95, when a file is deleted by a computer user, the clusters where the data is found can be reallocated to new data by the OS. However, until such a time, this “deleted” data remains in what is called unused or unallocated file space. This space has, more often than not, proven to contain data relevant to investigations hence, the need for its analysis.

##### **3.1.2 File Slack**

When files are created, their lengths vary depending on their contents. DOS, WINDOWS and WINDOWS NT-based computers store files in fixed length blocks of data called clusters. File sizes rarely match the size of one or multiple clusters perfectly. Hence, the unallocated storage space that exists from the end of the file to the end of the last cluster assigned to that file is what is referred to as the file slack [12]. This concept is also referred to as internal fragmentation [17, p308].

Such unallocated space should be investigated since they could contain previously created and relevant evidence [3, 12, 13]. This computer security weakness is exploited by forensic tools during investigations.

### **4 COMPUTER FORENSIC FUNCTIONALITY**

A number of functionalities are offered by the numerous forensic software tools. However, the focus of this paper is the disk imaging and hashing functionalities. Disk imaging is an important functionality since investigations should never be conducted on original storage media. Hence, disk

imaging is used to protect the integrity of any storage media to be investigated. If a storage medium's integrity is not maintained, results of an investigation could be rendered null and void in a court of law since defence attorneys are then able to bring the investigative process under question. Hashing and hash functions then become important since they offer a guarantee that an imaged device is actually the same as the original. These two functionalities are expanded within the following sections.

#### **4.1 Disk Imaging**

Typically, the first objective of a CFS would be to create an image of the storage device to be investigated. This image is and should be an exact replica of the original storage medium. Before this is done, it is vitally important to separate the suspect/owner from the computer immediately. If this is not done, it may be possible for the suspect to initiate a process on the target machine that overwrites the contents of the storage device [6].

Disk imaging can formally be defined as a physical sector-by-sector copy of a storage medium and the compression of the image into a file for forensic purposes. In addition, imaging tools will contain some internal verification mechanism, as described in section 4.2, to prove that the copy is exact and has not been altered. The image does not necessarily need the same geometry as the original storage device. This is because it is possible to simulate the geometry if it becomes necessary to boot into the acquired image [7].

In computer forensics, priority and emphasis are on accuracy, evidential integrity and security [7]. As such, the National Institute of Standards and Technology (NIST) offer some guidelines as to how disk imaging should occur. They suggest that [8]:

- 1) The tool should make a bit-stream duplicate or an image of an original disk or partition.
- 2) The tool should not alter the original disk.
- 3) The tool should be able to verify the integrity of a disk image file.
- 4) The tool should log I/O errors.
- 5) The tool's documentation should be correct.

The above mentioned guidelines help the CFS distinguish between tools since many free tools do not meet all these guidelines.

#### **4.2 Hashing Functions**

A hash function  $H$  is a transformation that takes an input  $m$  and returns a fixed-size string, which is called the hash value  $h$  [1, p590]. That is,  $h$  is the result of the hashing function being applied onto the input  $m$ . Hash functions form the foundation of the internal verification mechanism used by forensic tools to guarantee the integrity of the original media and the resulting image file. Message Digest 5 and the Secure Hash Algorithm are the most widely used hashing algorithms to date and these will be explained in the following two sections.

##### **4.2.1 Message Digest 5 (MD5)**

MD5 was developed by Professor Ronald L. Rivest of MIT [9]. The algorithm guarantees the integrity of an image file through the creation of a 128-bit message digest (hash value). This message digest is claimed to be as unique to an image file as a fingerprint is to a person. According to the Internet Engineering Task Force (IETF) [9], it is "computational infeasible" for any two data inputs to have the same message digest. MD5's author also claims, "it is conjectured that the difficulty of coming up with two messages having the same message digest is in the order of  $2^{64}$  operations, and that the difficulty of coming up with any message having a given message digest is in the order of  $2^{128}$  operations" [1, p594]. These guarantees make MD5 a credible hashing function.

### 4.2.2 Secure Hash Algorithm (SHA) 1

This is the second major hashing algorithm in use today. This algorithm is based on principles similar to those used in the design of MD4, the predecessor to MD5 [1, p594]. It produces a 160-bit message digest when an image file of size less than  $2^{64}$  bits is given as input to the algorithm. The SHA1 is called secure because, like the MD5 algorithm, it is computationally infeasible to find data which corresponds to a given message digest, or to find two different data files which produce the same message digest.





The following sections will now take a look at some forensic tools and discuss their effectiveness in carrying out some functionalities.

## 5 INVESTIGATION INTO SOME FORENSIC SOFTWARE TOOLS

Protecting a specific computer system from data corruption or the alteration of data, is the first step within the forensic investigation methodology. This is initially achieved through the complete isolation of the computer system from the suspect; typically the computer owner. Data integrity is further guaranteed by employing trained CFSs in the handling and investigative process. Forensic software tools are largely not involved within the protection step of the investigation methodology, as a result, the subject of data protection will not be considered any further.

The ability to access and analyse the contents of files is paramount to the success of an investigation. From figure 1, it can be seen that data access and analysis are not effectively supported. This matter needs to be addressed since the core of a CFSs task within the investigation is to discover specific contents of files that would link a suspect to the crime.

Figure 1 names some forensic tools and highlights their effectiveness at achieving the requirements of the investigation methodology.

	 PC Inspector File Recovery	 Encase	 Forensic Toolkit	 FTK Imager
<b>File (Data) Discovery</b>	●	●	○	●
<b>File (Data) Recovery</b>	○	○	○	○
<b>Reveal File Contents</b>	○	●	○	○
<b>File (Data) Access &amp; Analysis</b>	○	●	○	○
<b>Imaging</b>	○	●	○	●
<b>MD 5</b>	○	●	●	●
<b>SHA1</b>	○	○	●	●
<b>Summary Print-Out</b>	○	●	○	○

Key: ○ Not Supported    ○ Supported but Not Reliable    ○ Supported and Effective

The forensic tools used during the investigation were chosen for illustrative purposes and because of availability. Hence, the tools essentially cover the spectrum of issues that are dealt with in the paper. In addition, the investigated tools are presented in no particular order. PC Inspector File recovery is a freely available tool while Encase, Forensic Toolkit and FTK Imager are commercial tools. As a result, figure 1 was able to illustrate how a freely available forensic tool compares, in terms of functionality, with some commercial tools. It is also noted that Encase, Forensic Toolkit and FTK Imager were all demo versions and it is the author's belief that the full versions do incorporate more effective functionalities. The functionalities against which the tools are measured were chosen because of their importance to any forensic investigation. In particular, these functionalities follow the forensic investigation methodology and highlight common functionalities within the different tools.

In the sections that follow, illustrations and brief discussions about these different forensic tools are given.

### 5.1 PC Inspector File Recovery

PC Inspector File Recovery is a freely available forensic tool. This tool serves two main purposes. Firstly, to reveal the contents of all storage media attached to the computer system and, secondly, to recover any deleted data from the media.

As suggested in figure 1, this tool is very effective at detecting all files resident on a storage device. That is, all the file categories mentioned within the discovery step of the investigation methodology. All unreferenced files are associated with a condition. This condition can either be "good" or "poor".

The "Find lost data" option of the tool performs a sector-by-sector scan, which includes unallocated space and file slack of the storage media and reveals any files that were either lost or seemed to be deleted. While experimenting with this tool, we found that the probability of viewing or recovering an unreferenced file was higher if the file's condition was good. Essentially, the tool makes no guarantees on the accessibility of any unreferenced files. Figure 2 illustrates how seemingly deleted files are viewed and potentially recovered within the tools interface.

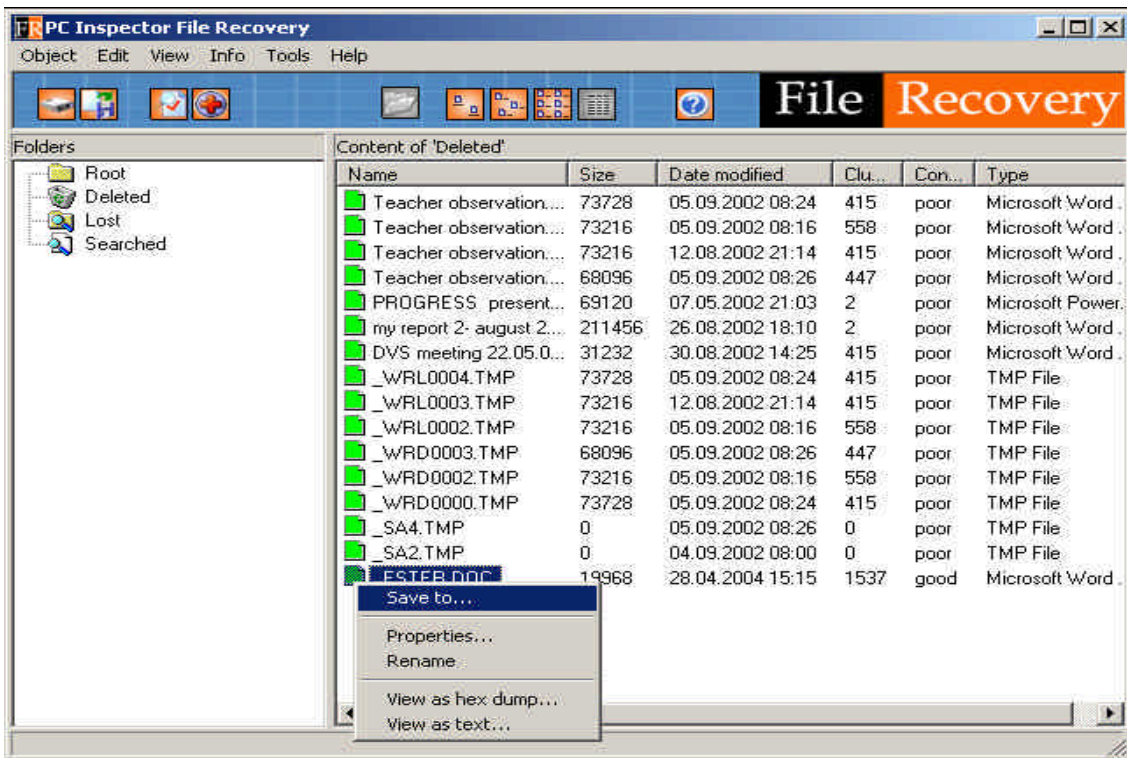


Figure 2: File recovery using PC Inspector File Recovery.



One of the appealing aspects about this tool is its simple interface which makes it useful to a CFS as well as the ordinary computer user. The tool is also very reliable in terms of discovering all the contents from storage media. However, PC Inspector File Recovery is generally not reliable in terms of data recovery.

## 5.2 Encase®

Encase [18] is a commercial forensic tool developed by Guidance Software. It was introduced to the forensics market in 1998. Encase’s functionalities include disk imaging, data verification and data analysis. An important feature is the recovery of data through the inspection of unallocated spaces. We must remember that these unallocated spaces could contain information relevant to an investigation.

A CFS using Encase would typically begin an investigation by seizing and imaging the storage device to be investigated. Encase refers to the resulting image file as an “Evidence File”. The Evidence File is a bit-stream image of the storage device. The software then verifies the integrity of the image file and the original storage media using the MD5 hash function. In order for the investigation to proceed, the imaged file is mounted by the tool to eliminate the need to restore the seized storage device [18].

The tool offers a cluster-by-cluster view of all files detected on the storage media. Vital information such as last access, time created, and last modifications of a file are all provided by this tool. Figure 3 illustrates how files are viewed with the Encase software tool. The first column, “File Name”, gives the names of some of the files being previewed within the tool. The “Description” column then gives the corresponding status of each file.

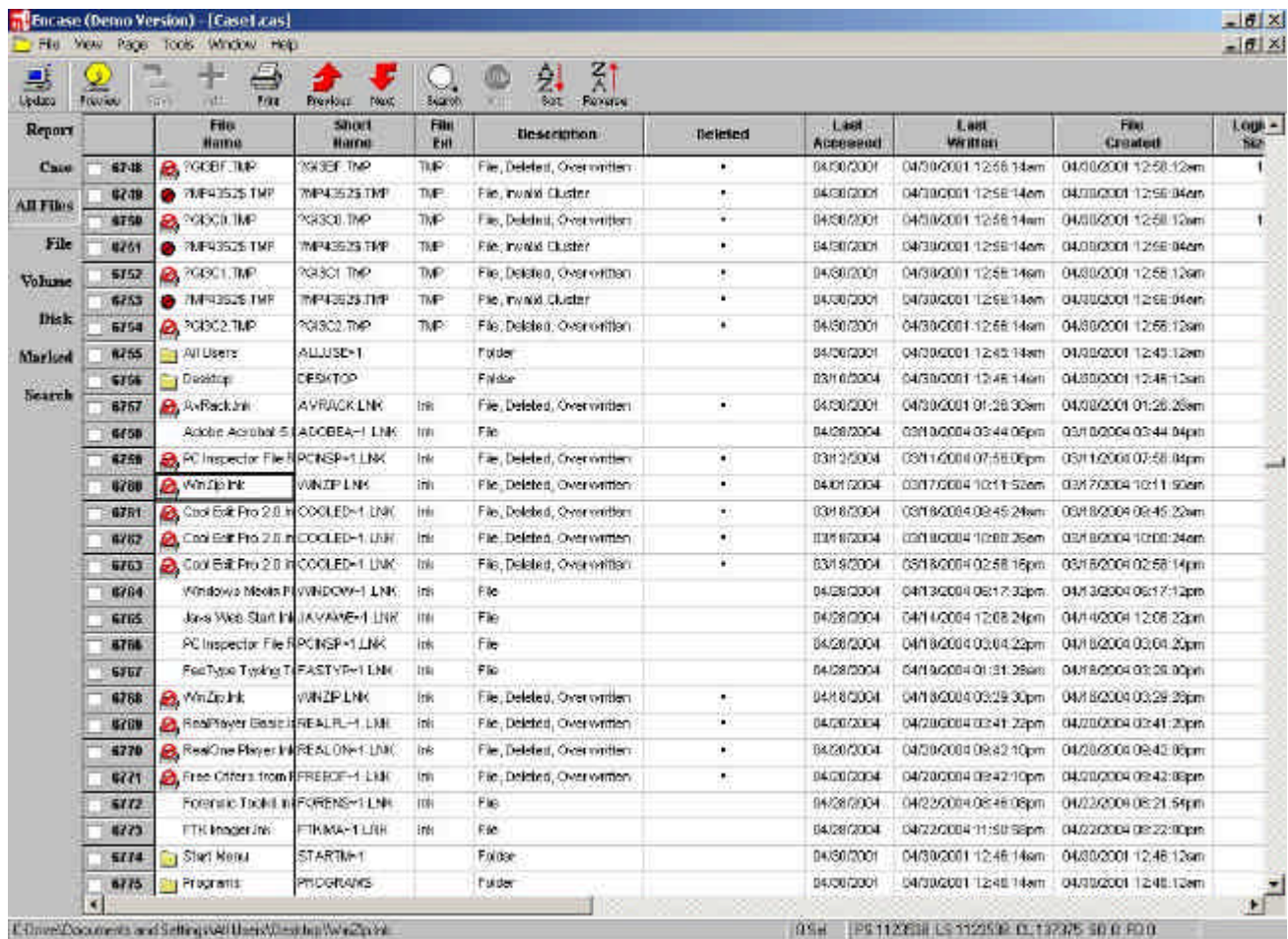


Figure 3: Data discovery view with Encase.



### 5.3 Forensic Tool Kit

Forensic Tool Kit is a commercial forensics tool developed by AccessData [20]. This tool allows the CFS to view all files on the chosen storage device. A function of this tool includes immediate generation of hash values for files that are viewed within an investigation.

From figure 1, it is clear that the most effective functionalities offered by this tool are the hashing functions. From the user-interface, it is apparent that the tool's developers intend the tool to be as simple and interactive as possible.

Unlike the above mentioned forensic tools, Forensic Tool Kit does not support data recovery. Since the data discovery functionality of the tool is not effective, data analysis and recovery are both affected. In light of all this, it is important to mention that all investigations were conducted on a trial version of Forensic Tool Kit. Therefore, it is our view that the full version does incorporate more effective and comprehensive functionality.

### 5.4 FTK Imager

FTK Imager is a commercial tool offered by AccessData [20]. Its main function is to view and to image storage devices. Data recovery can be attained in most instances as a result of the tool's ability to effectively preview these storage devices. It is worth noting that the tool's effectiveness at data recovery depends largely on the time when the file was actually deleted. The tool is also able to generate either MD5 or SHA hash values of all visible and accessible files. In particular, the MD5 hash value is generated and presented to the investigator as part of the completed process notification to guarantee the integrity of the original files.

Figure 4 illustrates a successful completion for the imaging of a floppy disk using FTK Imager. The integrity of the storage media is guaranteed through the generation of the MD5 hash value.

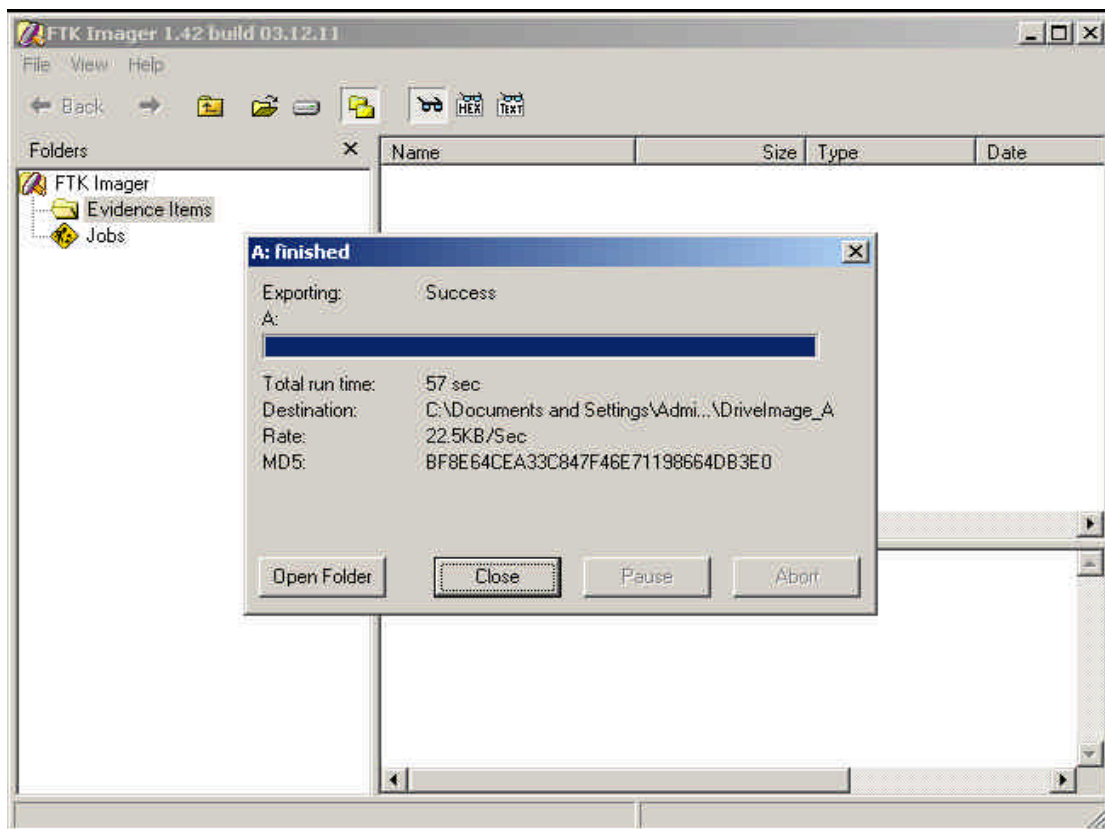


Figure 4: Storage device imaging with FTK Imager.

Now that a brief overview of the forensic software tools discussed in this paper has been given, the following section will provide findings and suggested improvements to these tools. Thereafter a conclusion will be given.

## **6 FINDINGS**

The ability to discover all original system files and unreferenced files from a storage device is of great importance to an investigation. From the forensic software tools that were investigated, it is clear, from figure 1, that this functionality is generally well supported. On the other hand, the data recovery functionality within these tools is in need of attention. From figure 1, it can be seen that none of the software tools, commercial or non-commercial, are able to guarantee the recovery of unreferenced files. A tool like PC Inspector File Recovery is able to recover files if these files were recently deleted. However, this would not necessarily assist an investigation if a storage device is only discovered several months, if not years, after the crime. This late discovery of a storage device could be the result of new leads within an investigation.

An aspect within computer forensics that is often overlooked is the reporting phase of an investigation. We find that the ability of the forensic tool to report, and assist the CFS in the investigation, is just as important as the tools' ability to view the contents off storage media. As mentioned in the background, storage device capacities are increasing at alarming rates. This implies that the search area that must be covered by the CFS is also increasing. Therefore, comprehensive reports of all activities undertaken by a CFS during an investigation would contribute positively to an investigation. It is our suggestion that reporting functionality be added to those tools lacking such reporting functionality and improved in the tools currently supporting this functionality.

Within the tools that supported imaging, we found that the process was consistently carried out successfully. Specifically, the NIST recently stated that the imaging functionality within the Encase tool was flawless [19]. This statement can be accepted without hesitation since Encase is considered an industry leader in the forensics software market [18, p53]. It should also be noted that the integrity of image files were always verified by the hashing functions supported within that imaging tool.

## **7 CONCLUSION**

This paper discussed some of the forensic software tools that CFSs use during their investigations. Four of these tools were evaluated with respect to their functionalities and effectiveness within the forensic investigation methodology. Finally, a discussion about these tools was given. The purpose of our approach was to highlight the shortcomings of current tools in order to provide suggestions for improvements. It is very important that CFSs are able to stay ahead of cyber-criminals through the use of forensic tools that allow them to reliably carry out their tasks within an investigation. We believe that if the suggested improvements to these tools are further researched, prosecutions of cyber-crimes will definitely increase.

## **8 REFERENCES**

- [1] Kurose, J.F, Ross, K.W. Computer Networking: A Top-Down Approach Featuring the Internet. Addison-Wesley, 2001. pp 1-6, pp 590-594.
- [2] Reinke, J, Saiedian, H. The availability of source code in relation to timely response to security vulnerabilities (ABSTRACT). Computers and Security, Vol 22, Issue 8, December 2003.
- [3] Judd, R, "An Explanation of computer Forensics".  
<http://www.computerforensics.net/forensics.htm>. [ACCESSED February 2004].

- [4] Rod, M. Options in Computer Forensic Tools. Computer Fraud and Security, Vol 2002, Issue 11, November 2002. pp 8-11.
- [5] "Computer Forensics Defined". New Technologies Inc.  
<http://www.forensics-intl.com/def4.html>. [ACCESSED February 2004].
- [6] Wolfe, H.B. The circumstances of seizure. Computers & Security, Vol 22, Issue 2, February 2003. pp 96-98.
- [7] Saudi, M.M. An Overview of Disk Imaging Tool in Computer Forensics. SANS Institute, 2001.
- [8] "Disk Imaging Tool Specification". Version 3.1.6. National Institute of Standards and Technology (NIST), 12 October 2001.  
<http://www.cftt.nist.gov/testdocs.html>. [ACCESSED February 2004].
- [9] "MD5" [http://searchsecurity.techtarget.com/gDefinition/0,294236,sid14\\_gci527453,00.html](http://searchsecurity.techtarget.com/gDefinition/0,294236,sid14_gci527453,00.html). [ACCESSED April 2004].
- [10] "SHA1 Encryption Algorithm". VOCAL Technologies, Ltd. 2003.  
<http://www.vocal.com/SHA1.pdf>. [ACCESSED April 2004].
- [11] "Ambient Data Defined". New Technologies Inc.  
<http://www.forensics-intl.com/def1.html>. [ACCESSED February 2004].
- [12] "File Slack Defined". New Technologies Inc.  
<http://www.forensics-intl.com/def6.html>. [ACCESSED February 2004].
- [13] "Unallocated File Space Defined". New Technologies Inc.  
<http://www.forensics-intl.com/def8.html>. [ACCESSED February 2004].
- [14] Sommer, P. The challenges of large computer evidence cases. Digital Investigation. The International Journal of Digital Forensics and Incidence Response, Vol 1, No 1, 2004. pp 16-17.
- [15] Stephenson, P. The right tools for the job. Digital Investigation. The International Journal of Digital Forensics and Incidence Response, Vol 1, No 1, 2004. pp 24-27.
- [16] Boyd, C, Forster, P. Time and date issues in forensic computing-a case study. Digital Investigation. The International Journal of Digital Forensics and Incidence Response, Vol 1, No 1, 2004. pp 18-23.
- [17] Stallings, W. Operating Systems: Internals and Design Principles, Fourth Edition. Prentice Hall International Inc, 2001. pp 305-563.
- [18] Casey, E et al. HANDBOOK OF Computer Crime Investigation: Forensic Tools and Technology. Academic press, 2002. pp 53-71.
- [19] <http://www.forensicdata.ca/products/software/forensic.htm> [ACCESSED April 2004].
- [20] <http://www.accessdata.com>. [ACCESSED March 2004].