Recovering Deleted and Wiped Files: A Digital Forensic Comparison of FAT32 and NTFS File Systems using Evidence Eliminator

Phil Nabity
University of Dallas, 1845 East Northgate Drive, Irving, TX
Phone: 214-262-7617
phill.nabity@necam.com

Brett J. L. Landry
University of Dallas, 1845 East Northgate Drive, Irving, TX
Phone: 972-636-8633
blandry@gsm.udallas.edu

ABSTRACT

Digital Forensics is the art and science of recovering lost and deleted information. This study compares two popular files systems, FAT32 and NTFS to examine differences in recovering files after deletion and wiped using Evidence Eliminator (EE). After 12 lab trials, it was determined that the use of EE for FAT or NTFS deleted means the files can not be forensically recovered.

INTRODUCTION

It is nearly a given of today's World Wide Web based society that Internet surfers have become accustomed to downloading any type of material from sites visited on the web. There is the plethora of applications, photos, documents, music, movies, games, and other potentially illegal material freely available via download so long as allowed by the Internet Service Provider and country of residence. It is the discovery of this illegal material and its intended purposes on the machine that causes serious ramifications for computer owners – whether aware or unaware of the material's existence.

Forensics has long been labeled as the field of science that pursues standardized processes in an attempt to collect, preserve, and finally present all available evidence for some interpretation of the law in criminal proceedings. These standardized scientific processes, when used in a structured and chronological methodology, will aid in the venture for definitive and undeniable answers to any potential breach of security investigation. Zatyko (2007) defines contemporary computer forensics as "the application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence after proper search authority, chain of custody, validation with mathematics, use of validated tools, repeatability, reporting, and possible expert presentation" (p. 1).

To assist in trailblazing the guidance of this experiment Zatyko's (2007) definition above offers the best explanation of digital or computer forensics. When looking at computer forensics (digital computer forensics or digital forensic investigations), there are three basic, generally accepted constructs used in the digital forensics investigation process. These constructs are the acquisition stage, the analysis stage, and the presentation stage (Carrier, 2002). In their book *Computer Forensics: Incident Response Essentials*, Heiser and Kruse (2001) introduce a forth construct called the "assessment stage" that is placed before the acquisition stage. Additionally, Brian Karney of Guidance Software, (www.guidancesoftware.com), believes that there is warrant enough to add a fifth phase to the investigation process. He suggests that an "authentication phase" should be added after the presentation phase and that this authentication phase is vital because investigators need to be able to ultimately prove that the evidence they have obtained is truly authentic (Miller, 2007). Each of these stages is to be discussed in this paper as to its relevance to the data collected, observed chain of custody, and eventual analysis of results.

In this experiment, the digital forensic investigation process was followed according to the constructs identified by Heiser and Kruse. The purpose of the study was to identify any differentiating aspects of deleted files that have been forensically recovered. Specifically, those deleted files (known variables) were retrieved from a computer running Windows XP Professional Service Pack 3 (SP3) and using the 32 bit entry File Allocation Table (FAT32) file system and compared against the same "known" deleted files from a computer running Windows XP Professional SP3 and using the Windows New Technology File System (NTFS).

FORENSICS FRAMEWORK

The Computer Investigation Model (CIM) from the reference book Computer Forensics: Incident Response Essentials was the model followed throughout this study. The book's authors, Heiser and Kruse (2001), have designed a flowchart that describes the logical flow of an investigation; all the while stressing the importance of documentation and proper chain of custody of the evidence. This model was followed throughout the study to ensure that the achieved results are able to be extrapolated, verified, replicated, and interpreted. Diagram 1 is a visual representation of the CIM.

To verify that the image taken from each hard drive for each analysis trial is indeed a mirror image, or in other words an exact duplicate of the original physical hard drive, a hash is calculated and if the hashes from both images are identical, then the two images are indeed identical. When traversing each phase it is critical that the proper implementation, preservation, handling, and documentation of the data retrieved so that any possible litigation that the data may be used in can be proven to possess evidentiary status.

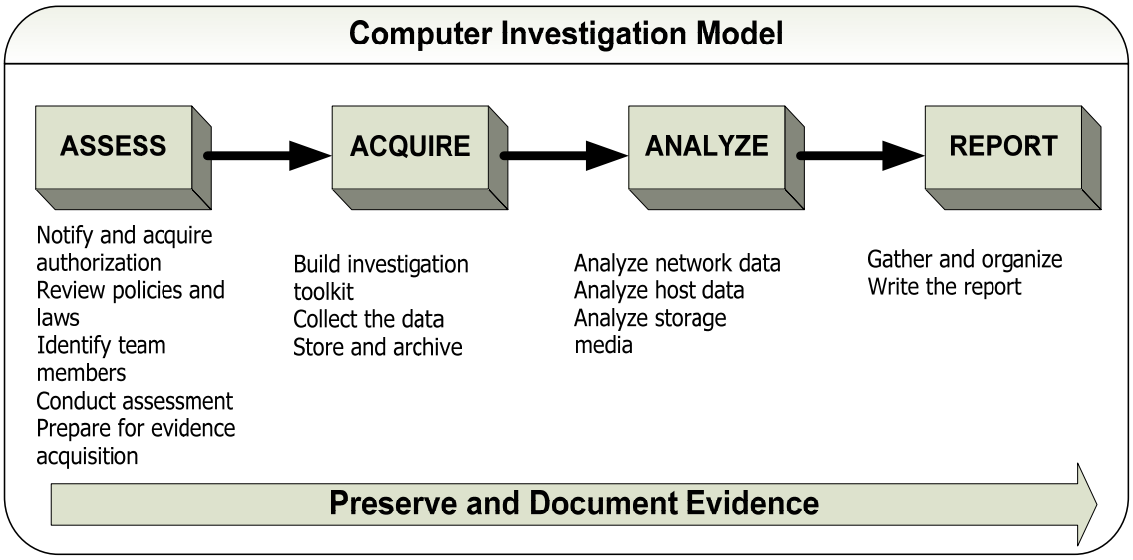

Diagram 1: Computer Investigation Model, Heiser and Kruse, 2001

Assessment Phase

The physical area constituting the investigation site needs to be documented both in writing and pictorially. Once all preparations are made for the site to be investigated by the internal company security team, then all team members should be briefed on the situation and possible outcomes. After all of this initial contact work has been completed, the team is ready to start the second phase: the acquisition phase.

Acquisition Phase

This phase is considered one of the more important phases in that during this phase all of the prep work done on the site during the assessment phase is physically implemented. As part of this phase the chosen method of imaging the hard drive is completed and an authentication hash (commonly MD5) is done on both the original hard drive as well as the image taken. It is during this phase that structured chain of custody requirements for the handling of all potential evidence (the original hard drive, the image of the hard drive, other storage devices) need to be implemented.

Analysis Phase

The analysis phase is just that: all data recovered during the acquisition phase is gathered in a central location and analyzed via various forensic tools for retrieval and interpretation purposes. The information retrieved (data files, email, music files, application files, Internet history files, web activity files, and so forth) is compiled, verified, and then tabulated for report preparation.

Reporting Phase

In the reporting phase, all analyzed and stored data from previous phases (following verified strict chain of custody procedures) is compiled into a standard reporting structure. The presentation of the report compiles and concludes the four phase investigation and categorically states the possible evidence retrieved.

TECHNICAL CONSIDERATIONS

There are a number of factors to consider when performing a forensic recovery including file and operating systems, hard drive sections, file slack space, windows file deletion, and wiping tools. Each of these must be considered or key information could be missed during the recovery. Each of these is discussed below.

File Systems and Operating Systems

In attempting to better understand the inner-workings of a computer system at its core level, it is necessary to understand the difference between a file system and an operating system and the functionality of each. Examples of different file systems include the following: FAT (FAT12, FAT16, FAT32) and NTFS which are implemented on Windows operating systems; HFS and HFS+ which are implemented on Macintosh operating systems; and finally, ext2 and ext3 which are implemented on Linux operating systems. When trying to understand the mechanics of how computers actually manage data, it is necessary to delve into the physics of how a computer's OS uses the file system architecture.

Sections of a Hard Drive

Regardless of the file system used to format the hard drive, all formatted hard drives can be categorized into three basic sections. The sections are labeled hard drive information, file storage information, and basic data. The hard drive information section contains specifics for that particular drive such as the sector size, the cluster size, name of drive, and the locations of other general information. The file storage information section stores all the names of every file and directory created on that particular drive. It also contains specifics on the clusters used by the named files. On a hard drive that has a single primary partition these three sections contain every piece of information needed by the computer to locate and access any stored data at the time when the user requests it.

File Slack Space

In addition to these areas, another important area on the hard drive is termed file slack space. File slack space is generally the amount of physical space that is left over or unused starting from the end of the written data file to the actual end of the cluster being filled on any file system (www.forensics-intl.com, 2008). To grasp the slack space concept consider this example: a computer is running XP with either FAT32 or NTFS as the file system. The default cluster size for either instance is 4096 (4K) bytes. A saved

data file with a size of 5750 bytes would actually need to use 2 clusters. The first 4K cluster would be completely filled and the second cluster would be partially filled with the remaining 1654 bytes. The 2442 bytes (4096 bytes – 1654 bytes) leftover in the cluster can not be used by the computer for data file storage so this space is in essence wasted.

All filled or partially filled clusters contain file path information needed for the computer to be able to retrieve the entire data file when needed. Consequently, the computer recognizes the left over space in the cluster and is able to store bits and pieces of other types of data in this space. Depending on the size of the hard drive and the file system used, there could potentially be much wasted space on the hard drive.

Windows File Deletion

When it is determined that a data file needs to be deleted in a Windows environment (regardless of file system used) it is most often placed into the recycle bin. All that has happened at this point to the file is that its logical path (for possible retrieval) has changed. Now for example, instead of the file's location being on the desktop or on the C:\ drive, the new location pointing to the file is the recycle bin. Once the recycle bin gets emptied, the file's data has still not been physically destroyed or deleted but rather the file information (path on cluster, sector information, creation date, modification date) and only this information for that particular file has been "erased". What this means is that the operating system would not be able to retrieve that file without the assistance of a freeware or commercial third party application (Undelete, Uneraser, WinUndelete, SoftPerfect) so to the operating system that file is no longer in existence.

The same functionality is involved if the file were to be deleted without putting it in the recycle bin (either by right-clicking on the file and selecting delete or if the file were simply too large to be placed in the recycle bin). When the file has been deleted (the file information is gone), the OS is notified by the file system that the space previously taken up by that data file has been freed up and is available for use. The next time the user saves a file, the older data is simply written over by the new data. Hence, in our last example of the 5750 byte file, a new file of size 2000 bytes is saved. Since this file does not even fill up the first 4K cluster, only 2000 bytes are used in the first cluster are used while the second cluster remains untouched. The data from the previously deleted file is still present in the second cluster and can be retrieved by forensic recovery tools. To add to the complexity to this example, the remaining 2096 bytes from the first cluster are now considered slack space and this slack space will contain bits and pieces of the first file that had initially been deleted.

Evidence Eliminator (EE)

In both experiment scenarios, the easy to use, commercially available, wiping tool Evidence Eliminator ^{TM} (EE) was chosen to clean the deleted material from the hard drive. After making a forensic image of each hard drive, a forensic recovery analysis tool was used to try to completely recover any deleted material.

Experiment Design

The purpose of the experiment was to prove or disprove that variations in file systems affect the recoverability of files that were deleted and then wiped. The difference between the trials is the procedural steps undertaken when the 32 bit entry File Allocation Table (FAT32) or Windows New Technology File System (NTFS) file system architecture is used locally to store and manipulate data on the computer. Microsoft OS Vista (and all of its flavors) was not considered for this experiment because Vista cannot be installed on a FAT32 partition (http://support.microsoft.com/kb/927520/en-us, 2008) and all new copies of the OS are defaulted to run on NTFS.

To maintain the validity of the procedures set forth in the CIM, the following steps were taken to ensure that the process is the same for each trial. In this experimental study the chosen method of investigation was cold and each machine was powered down before the start of the imaging process. The volatile memory was not taken into account during the acquisition phase due to the fact that the study is focused on finding differences in how the two file systems store data. The operating system XPSP3 and all parameters related to it (service packs, OS security updates, and so forth) were identical for each case. Additionally, all applications and software updates, all drivers, and all hardware (hard drive, mouse, keyboard, display,) were identical for each trial.

The study set-up consisted of a single stand alone Dell desktop computer with a Pentium III, 2.6 GHz Central Processing Unit (CPU) using two unformatted 20.5 gigabyte (GB) IBM Deskstar hard drives. To ensure that there was no previous data on any of the hard drives each drive was fully formatted six times using the Windows XP CD using either of the following patterns:

A) NTFS, NTFS, FAT32, FAT32, NTFS, NTFS or
B) NTFS, NTFS, FAT32, FAT32, NTFS, FAT32

Pattern A was used to format the NTFS drive and Pattern B was used to format the FAT32 drive. The following steps were employed in each of the trials.

1. The operating system used on the first hard drive was XPSP3 and the formatted file system was FAT32. The operating system used on the second hard drive was also XPSP3 and but the formatted file system was NTFS. Each hard drive had all current Microsoft OS and Microsoft Office 2003 security patches, application updates, and service packs installed as of the first image date. Each hard drive was able to support the same file extensions and also had Internet access.

2. MS Office 2003 Word, Evidence Eliminator (EE) version 6.0, and miscellaneous DVD copying and viewing software: QuickTime, Real Player, Flash Player, and WinDVD were installed on each hard drive.

3. At the start of each trial, there were three files of known size stored on different locations on the computer (known variables). The files included a 966 kilobyte (KB) digital picture (.jpeg extension); a 20.1 megabyte (MB) movie trailer (.mov

extension); and a 78 KB generic Microsoft Office 2003 word document (.doc extension). In addition to the files, there were also two other items controlled: a 2.3 MB freeware application called "WinRAR" (downloaded from www.downloads.com), and the IE web browsing history stored in cache. The WinRAR application was chosen because it represents a widely available, easily downloadable, freeware application that can be on any computer. The IE browsing cache was chosen because it contains the history of all sites visited on the web and its data may potentially be used by administrators and law enforcement as evidence to substantiate a possible crime.

4. The stored locations for the files were as follows: the word document was saved to the desktop in a folder called "docs", the photo was saved on the C:\ drive in a folder titled "photos", the WinRAR application in Program Files on the C:\ drive, the movie trailer in a folder titled "Movies" in the Windows Temp folder, and the browsing history is stored in the Windows Temp folder.

5. Step 5 involves deleting files. Each file was deleted and the hard drive "cleaned" using EE. All default parameters for EE were used except for the "clean swap space" parameters. Once applied and saved, the application completes a safe shutdown process where the saved configuration is processed. Upon a restart of the computer, the computer was checked to verify that the files were deleted. Since the safe shutdown process would be the normal activity of a user, this option will again be chosen and the computer will power down normally. At this point the Acquisition Stage of the CIM is undertaken by unplugging the computer from AC power, physically removing it, and then labeling it as the source hard drive. In any forensic investigation proper chain of custody measures would be meticulously followed: completely documenting the scene (writing in a notebook, taking pictures, interviewing suspects, etc), labeling all material, and finally packaging all material for transport back to the lab (if required). Since the hard drives and images never left the physical location, it can be assumed that a proper chain of custody was followed.

6. A second desktop computer (Pentium IV, 2.66GHz CPU, two 120 GB Maxtor hard drives) running XPSP3 served as the "target" disk. The target drive was defined as the drive that stores the image after it was been created. The second hard drive was formatted as FAT32 and consisted of four ~30 GB partitions. The Forensic Toolkit Imager application was executed on the computer and used to image the source hard drive by the "bit-wise" or "RAW" process; meaning that every bit on the drive is to be cloned. SAFE Block, a software write-blocker application, was used in all of the imaging trials to prevent the target OS from writing any data to the source disk upon connection.

7. All aspects of the imaging process using the software write-blocker and Forensic Toolkit Imager were successful. The one aspect of the application vital to the proof of a successful forensic image was the calculation of the encryption hashes (this application uses SHA-1 and MD5). The pre and post MD5 and SHA-1 hash calculations for all experiment trials matched identically. This indicated that exact

duplicates of the source hard drive had been imaged and saved accordingly to the partitions on the target drive for both NTFS and FAT32 trials. In this study, the imaging process was performed twice on each hard drive so that the hashes could be verified and there would be no mistake as to the validity of each forensic image. In all, the imaging procedure using Forensic Toolkit Imager was performed four times and the corresponding hash calculations for each trial were verified.

8. Once the image was successfully saved onto one of the four partitions of the target drive, the analysis phase began. Usually the image file is loaded onto multiple DVDs or CDs (depending on size) for transport and storage, but in this study the whole hard drive was moved and then mounted to a third setup for analysis. Each image file was copied from the 120 GB Maxtor hard drive to a Dell Latitude D630 laptop running the forensic retrieval software. The forensic recovery application Forensic Toolkit (FTK) version 1.80 (www.AccessData.com) was used to obtain a file by file analysis of both the FAT32 and NTFS images.

Once the imaging process was complete, the FTK recovery application was used to analyze each of the stored images on the target drive. This application performed a file by file recovery of any and all data on the image and stored this in a newly created case on the target drive. This resulting FTK case displayed a complete reconstruction of the file structure of the drive and attempted to show in as much detail what data was able to be retrieved. In some cases a file's metadata was retrievable but not the actual data. There was a noticeable difference in how the application displayed the reconstructed FAT32 drive in comparison to how it displayed the reconstructed NTFS drive. This is due to the structure of each file system's format; not how the forensic application chooses to rebuild the file structure. Nonetheless, the differences in file system structure did not detract from the available data that was retrieved; meaning a file recovered in the FAT32 trials was also recovered in the NTFS trials and data that was not recoverable in one file system was not recovered in the other.

According to FTK, none of the deleted files (doc file, mov file, jpeg file) were able to be recovered. This is also true for the application and IE browsing history that had been deleted in each trial. In addition to this, a function of the Windows OS is to store all actions performed on or to the computer by the user in various files on the hard drive. A few of these files include the pagefile.sys file, multiple index.dat files, and the different event log files. Each of these files had all data that was initially stored in them wiped clean as a function of the wiping tool. One conclusion that can be made from analyzing the FTK cases is this: when the EE application was configured and run on the source computer it truly did wipe the files clean. The use of the EE wiping tool resulted in there being no forensically recovered files from either the FAT32 or the NTFS file systems so that no comparison of these file systems were able to take place.

To validate the results and determine if EE was the sole reason for the files not being able to be recovered, the experiment trials were replicated in all aspects using two other sizes of hard drives. Two 13.6 GB IBM Deskstar hard drives and two 6.4 GB Maxtor hard drives were used to duplicate the experiment in its entirety. Again one disk

was formatted with FAT32 and the other formatted with NTFS using the patterns described previously. The assessment, acquisition, and analysis phases were followed exactly as before and the results of all trials were the same. Specifically, the results were duplicated on each file system loaded on all three sizes of hard drives for a total of 12 duplicated experimental trials.

To validate the functionality of the experiment using the process without the wiping tool, the entire process was replicated using a 500 MB partition (for both NTFS and FAT32) and 20 MB of miscellaneous files (pdf, doc, winRAR) were copied to that partition. The files were deleted by putting them into the Recycle Bin and then the Recycle Bin was emptied. The EE application was not used to clean this partition.

Furthermore, the forensic tool did a thorough job of recovering and reconstructing any deleted material that has not been subject to a powerful wiping tool such as EE. The end result of this additional experiment proves that the wiping tool itself is responsible for the lack of recoverable data from any of the drives.

To summarize the analysis of the experiment, the use of the Evidence Eliminator (EE) wiping tool in all of the twelve trial experiments allowed for no recoverable data from any of the forensic images regardless of the size of hard drive used. Even though the Computer Investigation Model was diligently followed and due care taken for the proper chain of custody, the wiping tool simply made the data unrecoverable. Therefore, files deleted from a FAT32 and files deleted from an NTFS drive when using EE will result in no recoverable data.

The initial scope of the experiment was to discover, identify, and report whether or not there were any differentiating aspects of files forensically recovered from images where the only discernable difference was the file system format when using a wiping tool. The (forensically recovered) deleted files from a computer using the FAT32 file system were to be compared to those same (forensically recovered) files deleted from a computer using the NTFS file system. After successfully duplicating the experiment twelve times, one asserted claim resulting from the analysis of the data is certain: EE does not discern between file systems. In fact, to the wiping application, any data located in a file or on any drive (partition) that has been identified (in the configuration) to be wiped, was thoroughly wiped, and was not recoverable regardless to file system; at least with FTK 1.8.

Future Work

Other experimental studies using EE may include the realm of virtual machine applications (VMWare) as well as virtual drives. How EE resides on a virtual server and carries out wiping functionality on these drives could mean possible disaster. The prominence that EE will have in upcoming years leads credence to the possibility of digital evidence being completely destroyed and never recoverable. This brings to mind the legal ramifications of the application. This is not to say that EE is a only a malicious tool and it can be used for good intentions as well such as preventing identity theft or the

complete destruction of old bank records? One thing for sure is that the wiping tools are here to stay and it is up to the digital forensics industry to be able to get the evidence needed when it is needed.

REFERENCES

Carrier, B. (2002). "Open Source Digital Forensics Tools - The Legal Argument" First published as @Stake Research Report.

Kruse, W. G. & Heiser, J. G. (2001) *Computer Forensics: Incident Response Essentials.*

Miller, R. (2007). "The Truth Is… Sleuthing for Data with Digital Forensics" Whitepaper March, 2007. Retrieved April 2, 2008 from www.econtentmag.com.

"You Cannot Select or Format a Hard Disk Partition When You Try To Install Windows Vista" Article ID # 927520. Microsoft, Inc. Retrieved June 20, 2008 from www.support.microsoft.com/kb/927520/en-us

Zatyko, K. (2007). "Commentary: Defining Digital Forensics" Forensic Magazine Issue Feb/ Mar 2007. Retrieved April 11, 2008 from www.forensicmag.com/article.

NTFS stands for New Technology File System and this took over from FAT as the primary file system being used in the Windows system. This NTFS file system is not only secure but also supports larger file sizes and hard drives. Indeed, before making a choice you need to know about their capabilities. The table below will give you a rough idea of the features and respective support. Cool Info: If we go back a few years, this New Technology File was first introduced alongside the corporate-oriented Windows NT 3.1. around July 1993 and then was used as a default file system for Windows XP, Windows