Evidentiary Considerations for Collecting and Examining Hard-Drive Media

Anthony F. DeSante
November 28, 2001
I. Introduction

Specific procedures for gathering and examining evidence are used to ensure the admissibility of hard-drive media data. These procedures are designed to specifically address the technical nature of the computer forensic examination process. Court precedent clearly indicates that computer-generated and computer-stored evidence is admissible with adequate foundation (Kerr, 2001, pp. 87-94). But, the transformation of data from a series of encoded bits to a trial exhibit is an abstract process that could lead a juror or judge to question the authenticity of computer-generated evidence. Opponents may scrutinize the examination process used by a computer forensic examiner to retrieve the data, when it is not clear that the evidence was properly handled during an investigation. Some foresee a trend toward more intense scrutiny, as awareness of the procedures used to ensure the authenticity and integrity of stored data increases (Kovacich & Boni, 2000, pp. 255-256). Such developments signify the need to follow specific procedures when gathering and examining evidence contained on a hard-drive. Exploitation of identified weaknesses in the evidence gathering and examination process may have negative ramifications during the evidence admittance process. The forensic examiner must ensure that the procedures used to process and examine computer evidence will not adversely affect the ability to achieve admittance (Kovacich & Boni, 2000, pp. 255-256). Therefore, the forensic examiner must consider admissibility rules concerning authentication and “best evidence.”

This paper analyzes four interrelated aspects of the computer forensic process to identify the proper procedures required to fulfill the evidence admissibility requirements. The analysis section addresses the Federal Rules of Evidence (FRE) of authentication and “best evidence” in the context of the collection, imaging, examination and preservation of hard-drive media. The research concludes that specific procedures must be used to address evidentiary requirements of hard-drive media. The procedures used to meet these requirements should be designed to ensure that the data contained on the hard-drive has not been altered, manipulated or damaged during each step of the investigation.

II. Analysis

Admissibility Considerations

A party opposing evidence derived from a hard-drive could challenge its admissibility on the grounds of authentication (FRE 901) and “best evidence” (FRE 1002/1003). The party submitting the hard-drive data must demonstrate that the recovered files are authentic to be considered relevant (Winn & Wright, 2001, chap. 20, p. 9). According to FRE 901(a), a proponent must offer evidence “sufficient to support a finding that the [computer record or other evidence] in question is what its proponent claims” in order to authenticate a record (Federal criminal code and rules, 1999, 308).

The best evidence rule requires that the original be submitted as evidence. This applies to information that is affixed to things, such as a magnetic impulse, mechanical or electronic recording (Winn & Wright, 2001, chap. 20, p. 31). Therefore, electronic data storage media (e.g., hard-drives) are also included. There is an exception to the rule. FRE 1003 states the following: “a duplicate of the original is always admissible to the same extent as the original
unless there is a genuine question as the original’s authenticity…”(Winn & Wright, 2001, chap. 20, p. 33). A copy of the hard-drive is used for data recovery and analysis (discussed in the *Imaging* section). This practice signifies the importance of successfully authenticating the original hard-drive.

The evidence proponent must meet two objectives when determining the procedures necessary to meet the authentication and best evidence requirements. First, the evidence proponent must demonstrate that the data recovered from a computer’s hard-drive provides a true and accurate reflection of the original data at the time of collection. This requires the proponent to address technical issues concerning *the process* used to examine the hard-drive. The second objective concerns the examiner’s ability to identify information derived from the hard-drive that specifically links the suspect with the recovered file. This addresses the *relevancy* aspect of authentication. The examiner’s articulation of the procedures and process used to collect, image, examine and preserve evidence must be persuasive enough to demonstrate that the evidence is relevant, or else it will not be admitted (Winn & Wright, 2001, chap. 20, p. 9 – 10). The persuasiveness of the examiner may be diminished if the appropriate procedures are not established and followed. Therefore, four functional aspects of computer forensics have been identified. Each of the following subsections discusses the procedures required to ensure the authentication of evidence recovered from a computer’s hard-drive and to ensure compliance with the “best evidence” requirement.

**Collection**

The collection of a hard-drive from a desktop computer during an investigation could involve many different scenarios. A primary concern in all situations is to establish a chain of custody. The chain of custody tracks evidence from its original source to what is offered as evidence in court. It is used to demonstrate that the evidence collected is authentic (Winn & Wright, 2001, chap. 20, p. 12; Feldman, 2001, p. 25). A chain of custody lends credibility to an examiner who is called upon to testify that a specific piece of evidence (e.g., hard-drive) is the same as the piece collected during the investigation. The documentation of the chain of custody is one of the primary means of differentiating a copy of the evidence used for analysis from the original (additional means are discussed in the *Imaging* section). The inability to clearly identify the copy from the original could cast doubt upon the integrity of the evidence and prevent the item from being properly authenticated (Casey, 2000, pp. 57-58).

Specific procedures exist to ensure that the integrity of the evidence is not diminished when dismantling a computer during collection. The examiner could jeopardize the authentication of evidence collected from a desktop computer if the appropriate actions are not taken and the proper documentation is not made at the time the item is seized. There are certain actions that may affect the evidence contained on a hard-drive at the time of collection. The sequence of steps taken to disconnect and disassemble a desktop computer is paramount to preserving data. The examiner should first observe the monitor screen before shutting down the computer. If it is evident that a program is running and is not destructive, then the program is allowed to complete prior to disconnecting the computer. The examiner should photograph the screen in order to preserve a visual image of the program running at the time of the collection. Then, the network, modem, power cord and battery are disconnected in this order.
(Lang, 2001, slide 6). Depending upon the operating system, the examiner has the option of either removing the power and communication lines to the computer or shutting down the computer using network commands (New Technologies, Inc., 1999, step 1). Evidence contained in Random Access Memory (RAM) may be lost if the power and communication feeds are cut off. In some circumstances it may be prudent to save programs running at the time of the collection to a clean floppy disk. This allows the examiner to obtain RAM memory without writing over existing evidence on the hard-drive (Casey, 2000, p. 54). The effect of this operation on the hard-drive must be evaluated before it is performed. Depending upon the operating system, this step may be unnecessary if the program was automatically saved to a swap file. Any operation performed prior to shutting down the computer should not overwrite or destroy evidence files at the time of collection (Anderson, 2000, second paragraph).

An inventory of the entire system should also be performed after the computer has been shut down. This entails documenting the hardware configuration by taking pictures of the computer from all angles. Also, the system’s hardware components and connections should be documented thoroughly. This includes tagging the ports and both ends of all cables to identify their connection location. This information may be used to reconstruct the system for trial, if necessary (Kovacich & Boni, 2000, pp. 259 – 260).

The examiner should use paper or antistatic plastic bags to package magnetic media such as a hard-drive (Technical Working Group, 2001, pp. 24, 35). Evidence bags should be identified and labeled with the investigator’s name, the method of contact, the date the item was taken into custody and a brief description of the case and the contents in the bag. Pertinent information, such as the make, manufacturer, data capacity (if applicable), model number, serial number, are documented for the collected hard-drive. If any damage or “wear and tear” is observable, then this should be noted, as well. A label with the investigator’s initial and date should also be put on each floppy disk seized. If there is more than one floppy disk seized, then each one is numbered in the order found (Stephenson, 2000, p. 249).

Additional procedures may also be employed when the forensic examiner is unable to remove the computer from the owner’s possession. If the computer must continue to be used by its owner, then two bitstream copies of the hard-drive (preferably in the presence of the owner or opposing counsel) should be made. The owner chooses one of them to be sealed in his presence. The label on the hard-drive is signed by the owner or counsel and the evidence is kept in a secure location. This protects the examiner against a challenge to the authenticity of the working copy, since the owner’s copy could be unsealed and compared to the one examined (Bates, 1997, Considerations for the courts).

Electromagnetic fields created by magnets, static electricity and radio transmitters encountered during the transportation of the computer may damage or alter data (Technical Working Group, 2001, p. 29). Therefore, equipment such as a radio transmitter should not be operated when transporting the computer. Moisture, high humidity, and excessive heat or cold could also adversely affect the computer’s components. Conditions like these should be avoided in order to ensure the integrity of the data stored on hard-drive media. (Technical Working Group, 2001, p. 36; Kovacich & Boni, 2000, p. 260).
The imaging process is a combination of software tools and procedures employed to produce a copy of the data contained on a hard-drive. The process of turning on a computer and/or utilizing a process to copy the original hard-drive may potentially contaminate the evidence (Sommer, 1998, p. 67). Therefore, it is recommended that a bitstream copy or other image of the original media be used for examination purposes (International Association of Computer Investigative Specialists, 2001, Hard Disk Examination). A bitstream or “mirror” image duplicates every sector contained on the surface of the hard-drive disk. Alternatively, a file-by-file copy does not capture the residual data necessary to perform a complete forensic analysis. Residual data includes deleted files, fragments of deleted files, slack space and other data not designated to a specific file name (Feldman & Kohn, 1998, Collect backup tapes). Swap files and cache files found on a disk may also contain evidence. Safeback is a common disk-imaging tool used by law enforcement to create a mirror copy. This product creates a single non-compressed file that is bit-for-bit copy of the original hard-drive (Stephenson, 2000, pp. 247, 254).1

There are several manufacturers of disk imaging tools, but not all disk imaging tools perform alike.2 The National Institute of Standards and Technology (NIST) publishes requirements for disk imaging tools. This publication provides a “measure of confidence” when selecting a disk imaging tool and specifies the following mandatory requirements:

5.1.1 The tool shall not alter the original.
5.1.2 If there are no errors accessing the source media, then the tool shall create a bit-stream duplicate of the original.
5.1.3 If there are I/O errors accessing the source media, then the tool shall create a qualified bit-stream duplicate. (A qualified bit-stream duplicate is defined to be a duplicate except in identified areas of the bit-stream.) The identified areas are replaced by values specified by the tool’s documentation.
5.1.4 The tool shall log I/O errors in an accessible and readable form, including the type of error and location of the error (NIST, 2001, pp. 4-5.)

One recent court case indicated that an imaging tool could be scrutinized if it does not meet the requirements similar to those specified by NIST numbered 5.1.1 and 5.1.2 (above). In Gates Rubber Co. v. Bando Chemical Industries, Ltd. 167 F.R.D. 90 (D. Colo. 1996), the plaintiff used data recovery software that overwrote seven to eight percent of the defendant’s hard-drive. The software also failed to record creation dates of files necessary for the identification of evidence pertinent to the litigation. In Gates, the judge specified that the plaintiff examining the hard-drive was obligated “to use technology which would produce the most complete and accurate picture of evidence.”

---
1 The fact that the file is non-compressed prevents a legal challenge concerning the potential alteration of data through compression or translation.
2 This was demonstrated by James Holley in his September 2000 survey of disk imaging tools. He discovered significant performance differences among four different disk-imaging tools (Byte Back, Drive Image Pro 3.0, Encase 2.08 and Linux “dd” 6.1). Byte Back and Encase have claimed to produce an exact copy or physical bitstream image, but Holley found that this was not the case with respect to some SCSI drives (Holley, 2000, pp. 1-8). This emphasizes the principal that a disk-imaging tool’s capability of creating a bitstream copy should be tested and confirmed prior to use.
NIST requirements 5.1.3 and 5.1.4 are also essential to the authentication process. FRE 901(b)(9) indicates that computer-generated evidence may be authenticated by describing the process or system used to produce the output (Winn & Wright, 2001, chap. 20, p. 12). The accuracy of the process used to image the disk may affect the admissibility of the evidence (Sommer, 1998, pp. 76-78). The input/output (I/O) log of errors and the identification of a qualified bit-stream duplicate provide the type of documentation necessary to confirm the results of the imaging process. Verifying the results of the imaging process and identifying the completeness of data is one method of authenticating computer evidence (Joseph, 1997, p. 36).

Differences between the original disk drive and the imaged copy could draw “best evidence” objections from an opponent. The forensic examiner can use mathematical techniques such as a message digest (MD) to verify that the copy is a true and accurate representation of the original (Noblett, Pollitt & Presley, 2000, Examining Computer Evidence). An MD is a signature that uniquely identifies the content of a file or a disk sector. The MD is created using a one-way hashing algorithm, such as MD5 (128-bit hash) or SHA (160-bit hash). A one-way hash function produces a unique “fingerprint” of a file, message, or other block of data. This is done by systematically processing bits of data grouped into blocks and producing a corresponding hash function (see Figure 1)(Stallings, 2000, p. 53). A change of one bit in the file would change the “fingerprint” produced by the hashing algorithm. A one-way hash value cannot be reversed, thus a file cannot be produced with knowledge of the hash value (Casey, 2000, p. 59).

The message digest is often complemented with an additional mathematical calculation called a Cyclical Redundancy Checksum (CRC). CRC detects errors caused by interference in the data stream between the original and the copy. Electromagnetic waves and excessive heat are two potential sources of interference during digital transmissions. A 32-bit cyclical redundancy check detects an error that is 32-bits or smaller. CRC performs this function by using an algorithm to calculate a number, which is transmitted as a checksum. The media device receiving the data stream performs the same algorithm on the data. This calculated checksum is compared to the one received. If the two values match, then no error is recorded. If there is no match then a digital alteration of the data stream has occurred during the transmission and an error is recorded (Morgan, 1996).

*EnCase* imaging software is one tool that utilizes both the CRC and MD5 algorithms to verify data during the examination process. CRC verifies the data stream during the copying process. The CRC value is written to the copy and stored as a file header with the MD5 hash value. The *EnCase* tool uses both of these values to verify the integrity of the data during an

---

3 One consideration concerning the admissibility of imaged hard-drives is that of expert scientific testimony provided by computer forensic examiners. The Supreme Court ruled in *Daubert v. Merrel Dow Pharmaceuticals* 509 U.S. 579 (1993) that a judge must make a “preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology can be applied to the facts at issue.” *Daubert* also states, “in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.” Therefore, a log of input/output errors assists in substantiating the practice of using an imaged hard-drive as evidence when considering the *Daubert* requirements for expert scientific testimony.
examination of the copy. The program accomplishes this by recalculating the CRC and MD5 hash values and comparing them to the first values created during the copying process. A report is generated to document this verification process.

### Simple Hash Function

*(Stallings, 2000, p.54)*

<table>
<thead>
<tr>
<th>Description</th>
<th>Bit 1</th>
<th>Bit 2</th>
<th>⋯</th>
<th>Bit 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>$b_{11}$</td>
<td>$b_{21}$</td>
<td>⋯</td>
<td>$b_{n1}$</td>
</tr>
<tr>
<td>Block 2</td>
<td>$b_{12}$</td>
<td>$b_{22}$</td>
<td>⋯</td>
<td>$b_{n2}$</td>
</tr>
<tr>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
</tr>
<tr>
<td>Block $m$</td>
<td>$b_{1m}$</td>
<td>$B_{2m}$</td>
<td>⋯</td>
<td>$b_{nm}$</td>
</tr>
<tr>
<td>Hash code</td>
<td>$C_1$</td>
<td>$C_2$</td>
<td>⋯</td>
<td>$C_n$</td>
</tr>
</tbody>
</table>

**Figure 1**

*Examination*

The forensic examination process is subject to many variables that must be controlled and/or documented in order to ensure authentication. Attention should be paid to every step performed regarding the examination. Notes of observations must be in ink, so as to have a permanent quality. Corrections to these handwritten notations should be made with a single strikeout and initialed by the person who made them. All handwritten documentation should be retained. No erasures should be made (Scientific Working Group, 1999, Standards and Criteria 1.6). A forensic examiner who utilizes these practices is better capable of convincing all concerned parties that decisions made during the examination process did not bias the results. The inability to demonstrate the “completeness of the evidence” reviewed may pose a challenge to its authentication (Joseph, 1997, p. 36).

The examination process entails following certain standard procedures to avoid contamination of the evidence. “Sterile” conditions should exist during an examination (IACIS, 2001, Forensic Examination Procedures). The hard-drive media used as the examination copy should be completely wiped, scanned for viruses and verified before being used. The computer used during the examination should be booted from floppy disk contained in the “A” drive (Sommer, 1998, p. 67). The boot sequence and/or running an application (like Microsoft Windows) on the evidence media may alter the file slack, erased files and Microsoft swap files. When running Windows, new files could be created, existing ones opened or file directories updated (such as file dates) (Anderson, 2000, second paragraph). As a precaution, both the original evidence media and the copy should be “write-protected” to ensure that no data is added or changed during the examination process (Feldman, 2001, 26). Evidence extracted from the copy of the media examined should be transferred to a CD-ROM. A CD ROM is a “Write Once, Read Many” medium that cannot be written over (Sommer, 1998, p. 67). These practices ensure that the data examined is the same data obtained from the original, a primary objective of authentication.

The examiner may also be called upon to assist in linking the evidence examined with the person who created it. Under Rule 901(b)(4) of the federal rules of evidence, the
authentication of evidence may come from its “[a]pearance, contents, substance, internal patterns, or other distinctive characteristics, taken in conjunction with the circumstances” (Winn & Wright, 2001, chap. 20, p. 11). For this reason, the forensic examiner may be required to play a key role in establishing that the recovered data is the work of the suspect. Data created with the file can be used to authenticate evidence. The “distinctive characteristics” mentioned in Rule 901(b)(4) can be obtained from the Master File Tables (MFT) or Root Directories. This data may provide the necessary link to the suspect by representing unique or unusual file information known only to the creator (Overly, 1996, Authentication of Electronic Documents). For instance, the New Technology File System (NTFS) creates attributes stored for each file in the Master File Table (MFT). Some of these attributes are necessary for the proper operation of the MFT and are always located with the file. These “resident” attributes specify information such as the date and time of when the file was created, modified and accessed. Additionally, the examiner is capable of identifying whether or not the file is read-only or hidden. These attributes may provide circumstantial evidence to link the suspect with a file. For instance, a file opened from a CD ROM disk will have a read-only attribute. Therefore, by identifying the fact that the file located on a hard-drive was designated as read-only, a link can be drawn between the file and a particular CD ROM. File directories located on a Disk Operating System (DOS) also records attributes for each file created. An “S” attribute is an example of one attribute that may be used for authentication purposes. This attribute indicates that the file is a system file. A system file usually refers to one of several hidden files that are used to boot the operating system (e.g., IO.SYS, MSDOS.SYS). If this attribute is found with a non-system file, then there is an indication that the creator may have been trying to disguise it. Forensic techniques such as these provide the examiner with a piece of knowledge that would have been known only by the creator. This authenticates the file with respect to linking the creator to the actual file produced.

Preservation

The logical and physical aspects of an imaged hard-drive must be preserved before, during and after an examination is performed. Physical protection is ensured by maintaining the chain of custody as discussed in Collection. This entails bagging, tagging and restricting access to the hard-drive to authorized individuals. A custodian maintains the evidence when it is not in use and ensures that this process is followed. If the evidence is passed on to another custodian, a change in official control must be documented (Stephenson, 2000, p. 248). Logical protection ensures the protection of the data contained on the hard-drive. This entails attaching a digital signature to the message digest. Public key encryption is used to create a digital signature. This signature is used to identify the individual who produced the message digest (Casey, 2000, p. 60). Public key encryption has two keys: a public key and a private key. The private key is used to encrypt the file and the public key is used to decrypt the file. The public key cannot re-encrypt the file after having been decrypted. This encryption methodology allows the examiner or other party to access the file for analysis, but gives the examiner control of the encryption process through the use of a private (or secret) key. The forensic examiner is capable of opening the original file and running a CRC and MD check against this file. The “seal” is shown to be intact when these authentication programs generate the same values for both the original and the copy. These programs allow the examiner to
testify with a degree of certainty that the original and imaged copy are one and the same (Stephenson, 2000, pp. 250-251).

An inventory of files contained on the hard-drive is generated and preserved soon after the disk has been imaged. FileList is a New Technologies, Inc. utility that identifies every file on a disk. The “inventory” list is presented logically in a standard directory tree format. FileList also provides the date of last change and file size and an MD5 message digest of the file. MD5 ensures that a listing from FileList is a nearly perfect and a practically unalterable representation of files present in logical disk space. The data files created by running the FileList program should be signed by the examiner and sealed in an evidence bag (Stephenson, 2000, p. 250, 258).

III. Conclusion

A forensic examiner has two primary evidentiary concerns when considering the admissibility requirements of a hard-drive. First, procedures to authenticate evidence must be employed to ensure the integrity of the hard-drive data when collecting, copying, analyzing and reconstructing evidence from the hard-drive. Second, the forensic tool used to image the hard drive should not alter, manipulate or damage the data on the original hard-drive. This would comply with the “best evidence” rule by making the original no different from the copy (Kerr, 2000, pp. 89, 94). These evidentiary issues pose challenges to the forensic examiner at both the physical level (handling of evidence) and the digital level (examination of evidence). The primary goal is to generate data during the examination process that accurately represents the evidence collected from the suspect.

---

4 One could argue that FileList should be run on the original media prior to imaging, but this may subject the original evidence to the potential for alteration or destruction.
References


