Use of Computer Forensics in the Digital Curation of Removable Media

William Bradley Glisson
Humanities Advanced Technology and Information Instituted (HATII),
The University Glasgow, Scotland
George Service house
11 University Gardens, G12 8QQ
+44 (0)141 330 8591
b.glisson@hatii.arts.gla.ac.uk

ABSTRACT

The purpose of this paper is to encourage the discussion of the potential place and value of digital forensics techniques when dealing with acquisitions on removable media in the field of digital curation. It examines a basic computer forensics process, discusses a typical file system for removable media, and raises questions about necessary processes and incentives for addressing data capture in the field of digital curation.

General Terms

K.6.m Miscellaneous

Keywords

Computer, Computer Forensics, Digital, Digital Forensics, Digital Curation, Archival

1. INTRODUCTION

As has been noted in past papers, information is an extremely valuable asset in the world's increasingly globally networked environment [1]. The digital revolution has contributed to a radical paradigm transformation in today's information rich societies. Many of society's basic operational components such as financial markets, health care, and governmental agencies depend on information in digital formats.

An excellent example of the growth in digital information is provided by the White House. In 1978, Congress passed the Presidential Records Act, "which requires each president to maintain records of all activities, deliberations, decisions and policies that reflect on performance in office" [2]. There have been questions in the press as to whether the National Achieves can handle the volume of digital data produced during the Bush administration [3]. It has been estimated by archive officials that the "… electronic record(s) of the Bush years (are) about 50 times as large as that left by the Clinton White House in 2001" [3].

This paper discusses the potential place and value of digital forensics techniques for collecting institutions (e.g., libraries, museums, archives) that are dealing with acquisitions on removable media. The hope is that this paper will encourage future research in the adoption of computer forensics processes and tools for digital curation.

2. COMPUTER FORENSICS PROCESS

This section summarizes a typical digital forensics workflow.

2.1 Process for Retrieving Data

For this discussion, consider the forensic acquisition of a data key, a.k.a. thumb drive. One must first decide where to store the information. In order to counter data remembrance so that it does not contaminate the information stored on the target drive, the target drive needs to be forensically cleaned. This means that the target drive is wiped by writing all zeros or ones to the drive. Many companies promote a US Department of Defense (DOD) standard on this topic. However, the 2006 National Industry Security Program Operating Manual (that is also referenced as the DOD 5220.22-M) does not specify the number of passes required to achieve sanitation [4].

The Defense Security Service (DSS) Clearing and Sanitization Matrix which was updated on June 28, 2007, makes the statement that "DSS will no longer approve overwriting procedures for the sanitization or downgrading (e.g. release to lower level classified information controls) of IS storage devices (e.g., hard drives) used for classified processing" [5]. There is clearly controversy over the effectiveness of overwriting drives. A recent study that claims "…that correctly wiped data cannot reasonably be retrieved even if it is of a small size or found only over small parts of the hard drive" [6]. Even thought there appears to be a doubt as to the effectiveness of overwriting for sanitation purposes, it is still a good idea from a forensic practice perspective.

The next activity is to document the hardware. This includes any serial numbers and manufacturer information. The next activity in a forensic situation would be to start the chain of custody and to transport the device to a secure lab for processing.

At this point, a bit-stream copy of the removable media should be made by creating either a clone or a forensic image of the device. A bit-stream copy of the removable media copies every bit on the source drive [7]. When a bit-stream copy is saved to another drive, i.e., the target drive (generally, this drive needs to be identical to the source) so that the target drive is bootable, this is commonly referred to as a clone. When the bit-stream copy is saved to an image file, this is commonly referred to as a forensic image. It is possible to take a forensic image and restore the image to a drive making a clone of the source drive.

At this point, the forensic copy of the removable media then needs to be authenticated. This is typically done through the execution of a one-way hash on both devices to verify that they are identical.

The next step is the analysis of the drive to identify active files and inactive files. Active files are readily identifiable and can be accessed with the appropriate software and, in some cases, the required security information. Inactive files can be located by

carving the unallocated space and slack space off of the drive. Unallocated space is space that has not been used by the file system. It can contain deleted documents in Windows and DOS operating systems. Information can also be found in two types of unallocated slack space: file slack and RAM slack (sometimes both are referred to as drive slack) [7]. Any anomalies that are identified such as encrypted information, proprietary software formats and missing partitions are noted and examined individually. All of the information that is found would be documented appropriately.

The documentation is very detailed so that it includes all of the issues that were encountered and the evidence that was discovered in the process. It will also include the methods utilized in the investigation along with citations supporting the analysts' stated opinions.

2.2 Removable Media File Systems

The next issue to address is the file system. It can be argued that the file system is part of the application layer, the presentation layer and the session layer as defined in the Open Systems Interconnection (OSI) seven-layer model [8]. The file system is responsible for the organization of the files, i.e., it is responsible for the logical placement of the files on the storage drive. Hence, the file system is manipulating the sectors on a drive so that they are treated as clusters. These clusters are then linked together, as needed, so that they can be treated as a file with associated metadata. The size of the clusters will vary depending on the size of the hard disk and the file system [7]. Understanding this interaction is critical to the retrieval of data that has been accidentally or intentionally deleted on various types of files systems like the File Allocation Table (FAT) system, New Technology File System (NTFS), High Performance File System (HPFS) or the Hierarchical File System (HFS).

It is common, although not mandatory, for data keys to use a version of the File Allocation Table (FAT) system. When a file is deleted in a FAT system, the first character of the file is replaced with a non-readable character and the FAT entries linking the sector clusters are zeroed out. The data still exist on the system. The zeroing out of the entries linking the clusters simply tells the file system that the space is available for use. Hence, to restore a file, the name of the file would need to be amended and the links between the sector clusters would need to be re-established. If additional data has been saved to the system after a file has been deleted, the old data may have been over written.

3. INFORMATION APPLICATION

Now that we have an understanding of the basics of computer forensics, how can we apply this to the field of digital curation? In the case of the White house and the Presidential Records Act, does the administration have an obligation to not delete information? Does the archivist have an obligation to recover as much information as is possible from the digital media provided by the administration? If so, how does the archivist achieve this goal?

4. CONCLUSION & FUTURE WORK

This paper is intended to raise awareness of computer forensics concepts and to prompt discussion about the potential use and the need for computer forensics processes and tools in the field of digital curation. The long-term implications, obstacles, and hurtles for the integration of digital forensics processes and techniques into the field of digital curation are not fully understood. It is clear that the digital revolution will continue to penetrate all aspects of our globally networked information rich society. This continued integration raises the need to address the amount of information that is archived along with the examination of the processes that are implemented.

There are many associated questions to consider:

- When collecting institutions receive removable media, what procedures do they follow?
- Do they only capture "live files," or do they also capture deleted files?
- Under what conditions would it be beneficial for collecting institutions to copy entire drives, i.e. all of the bits, rather than only copying the live files from the drives?
- Are current practices of collecting institutions practical from a business perspective? To what extent do they conform to established digital forensics principles and practices?
- How often, and under what conditions, will the processes and storage arrangements of collecting institutions need to be upheld in a court of law?

Future research could focus on a closer examination of the process used in the field of digital curation. This can include conducting survey inquiries with collecting institutions in several different countries. It could also include the investigation and development of a digital recovery methodology specifically for use in digital curation. It could reasonably include a targeted educational effort toward librarians and archivists on relevant tools and the operation of specific file systems. The educational effort could be followed by empirical studies of the practicality and effectiveness of the developed methodologies to meet the needs of collecting institutions and their target user communities

REFERENCES

1. Glisson, W.B. and R. Welland. *Web Development Evolution: The Assimilation of Web Engineering Security*. in *3rd Latin American Web Congress*. 2005. Buenos Aires - Argentina: IEEE CS Press.
2. Aitoro, J.R. *Administration faces big challenge in records preservation*. 2008 [cited 2009 Januaruy 17]; Available from: http://www.govexec.com.
3. Pear, R. and S. Shane. *Keepers of Bush data face system overload as electronic records snowball*. International Herald Tribune 2008 [cited 2009 December 28, 2008]; Available from: http://www.iht.com.
4. Department of Defense, *National Industry Security Program Operating Manual*, D.o. Defense, Editor. 2006, Defense Technical Information Center: Washington, DC.
5. Defense Security Service. *Industry Security Letter*. 2007 [cited 2009 Jan. 17]; Available from: https://www.dss.mil/.
6. Wright, C., D. Kleiman, and S. Sundhar, *Overwriting Hard Drive Data: The Great Wiping Controversy*, in *Information Systems Security*. 2008, Springer Berlin / Heidelberg. p. 243-257.
7. Nelson, B., et al., *Guide to Computer Forensics and Investigations*. Third ed. 693. 2008: Thomson Learning, Inc.
8. SearchNetworking.com. *OSI*. [cited 2009 Feb 1]; Available from: http://searchnetworking.techtarget.com.