Chapter 2

APPLYING TRADITIONAL FORENSIC TAXONOMY TO DIGITAL FORENSICS

Mark Pollitt

Abstract Early digital forensic examinations were conducted *in toto* – every file on the storage media was examined along with the entire file system structure. However, this is no longer practical as operating systems have become extremely complex and storage capacities are growing geometrically. Examiners now perform targeted examinations using forensic tools and databases of known files, selecting specific files and data types for review while ignoring files of irrelevant type and content. Despite the application of sophisticated tools, the forensic process still relies on the examiner's knowledge of the technical aspects of the specimen and understanding of the case and the law. Indeed, the success of a forensic examination is strongly dependent on how it is designed. This paper discusses the application of traditional forensic taxonomy to digital forensics. The forensic processes of identification, classification/individualization, association and reconstruction are used to develop "forensic questions," which are applied to objectively design digital forensic examinations.

Keywords: Digital evidence process, forensic taxonomy, forensic examination

1. Introduction

Early forensic practitioners from a variety of jurisdictions and backgrounds recognized that evidence stored in electronic form is easily changed with improper handling. In the early 1990s, the International Association of Computer Investigative Specialists (IACIS) promulgated what was, perhaps, the first set of guidelines for digital forensics. The Association of Chief Police Officers (United Kingdom) followed with a good practice guide. Subsequently, the International Organization on Computer Evidence (IOCE) and the G-8 developed a set of principles for computer-based evidence. All these documents stipulate that digital

Please use the following format when citing this chapter:

Pollitt, M., 2008, in IFIP International Federation for Information Processing, Volume 285; *Advances in Digital Forensics IV*; Indrajit Ray, Sujeet Shenoi; (Boston: Springer), pp. 17–26.

evidence be acquired in its totality and that it not be altered during any subsequent examination [8].

These guidelines and principles are reinforced in virtually every digital forensic model. Despite their differences, most forensic models [1–4, 14, 16] follow evidence acquisition with evidence preservation, typically by creating a digital image of the media. Interested readers are referred to [13] for a review of the principal forensic models.

As a result, virtually all forensic examinations start with the totality of the evidence. The examiner is then required to locate, extract and present the material of forensic value. The two fundamental approaches are selection and reduction, and they are often used in combination. Selection involves searching the data (e.g., using string searches) to locate information of probative value. Reduction involves the removal of information that is not of forensic value. This process often uses "negative hashing," where the hash values of known "good" files are used to eliminate unknown files. Negative hashing is facilitated by repositories of file signatures such as those available at the National Software Reference Library [11].

The selection and reduction approaches are both less than optimal. When applying selection, forensic examiners must know, with some degree of specificity, what they are looking for and where it might be located. The irony of this approach is that the more deterministic the approach, the less complete the answer. In the case of reduction, the evidentiary material that remains is often so voluminous as to be unmanageable. To refine their approach to examinations, forensic examiners carefully consider the facts of the case, the elements of the violation and the behavior of computer users. Experiential knowledge is vital to conducting examinations that are efficient and effective, but efforts to objectively identify and articulate this knowledge have not been very successful.

2. Traditional Forensic Science

Science has provided a foundation for legal proceedings for more than 100 years. During this time, the science practiced in the legal system has differed from traditional scientific endeavors in its form and application, not in its content. Moreover, while traditional science engages the "scientific method" to drive methods of proof, the legal system has demanded additional approaches to ensure the reliability of the evidence, the scientific methods applied and the resulting testimony. These requirements are the result of judicial decisions rather than scientific research and discourse [17].

Edmond Locard, an early 20th century French criminologist, is considered to be the pioneer of modern forensic science. His celebrated "Exchange Principle" postulated that when objects contact one another, there is an exchange of material [5, 9]. A long list of distinguished forensic scientists have added a number of principles to the corpus of forensic scientific knowledge. Nevertheless, there have been surprisingly few attempts to develop ontologies for these principles. This paper draws on two important approaches, Inman and Rudin's Unifying Paradigm of Forensic Science [5] and Lee and Harris' General Concepts in Forensic Science, to further develop a model for digital forensics [7].

3. Need for Structure

Thomas Kuhn's seminal work, *The Structure of Scientific Revolutions* [6], discussed the importance of paradigms:

> "The study of paradigms is what mainly prepares the student for membership in the particular scientific community with which he will later practice. Because he here joins men who learned the bases of their field from the same concrete models, the subsequent practice will seldom evoke overt disagreement over fundamentals. Men whose research is based on the shared paradigms are committed to the same rules and standards for scientific practice."

The adoption of a paradigm certainly facilitates the instruction of students, but it also allows for the formulation of an accepted practice that adds to the efficiency, effectiveness and reliability of the practitioner's work. The question then becomes: What paradigm?

4. Application of Traditional Forensic Science

Inscribed on one of four large statues in front of the U.S. National Archives is the quotation: *"What is Past is Prologue."* Many credit Shakespeare for this quotation, but it was, in fact, modified from the original (Act II of *The Tempest*) by John Russell Pope, the architect of the building [10]. It is appropriate that an idea from several hundred years ago that was adapted to modern usage lights the way for the newest forensic science. Traditional forensic science has been developing its paradigm for decades and some of its concepts can be adapted to digital forensics.

Locard's Exchange Principle influenced a number of forensic scientists to develop new ways for looking at evidence. Inman and Rudin [5] have analyzed six of these approaches, categorizing two of them as "principles" and four as "processes."

The two principles are "transfer" and "the divisibility of matter." The first is recognized as Locard's observation; the second was proposed by

Inman and Rudin as a way of explaining the ability to impute characteristics to the whole from a separated piece. It is easy to see how these principles underlie many of the biological, physical and chemical examinations conducted by traditional forensic scientists. The two principles also apply to digital forensics – digital evidence exhibits transference in its interactions and electronic duplicates are representative of the original evidentiary items. But these principles do not have a great deal to offer in terms of developing examination strategies.

On the other hand, the four processes of "identification," "classification/individualization," "association" and "reconstruction" have the potential to be very useful from the perspective of planning digital forensic examinations. The following sections analyze these four processes and discuss how they might be adapted to digital forensics.

4.1 Identification

Inman and Rudin credit Saferstein [15] with defining the concept of identification as the physiochemical nature of the evidence. They note that being able to accurately describe an item or its composition may be sufficient for a given forensic purpose. For example, when the mere presence of illicit drugs is an important element of a crime being investigated, the identification of a white powder as containing cocaine, dextrose and talc may be all that is required.

In the discipline of digital forensics, identification helps describe digital evidence in terms of its context – physically (a particular brand of hard drive), structurally (the number of cylinders, heads and sectors), logically (a FAT32 partition), location (directory and file) or content (a memo, spreadsheet, email or photograph). The presence of metadata or the existence of a particular letter (not necessarily their content) may be probative in an investigation. In other situations, as in child pornography cases, the nature of the content is dispositive. On the other hand, the mere presence of connections between certain computers may demonstrate a key fact in an intrusion case.

Examiners are routinely asked to find evidence on computer storage media, but the tasking is usually done in an investigative context as opposed to a digital context. This places the burden on the examiner to translate the task into an examination plan or strategy. By focusing on the characteristics of the potential evidence, it is possible to search for it in the same way that one looks for cocaine in a drug investigation – by conducting specific examinations.

This process is done best by working backwards. First, we ask, What information is desired? The next logical question is: In what form might

this kind of information be stored? Finally, Where might this information be located? Selecting a tool and query that searches in specific locations for limited types of data that have particular characteristics significantly reduces the forensic burden. Simultaneously, it produces "rich" information that may be sufficient for the investigation.

4.2 Classification/Individualization

Inman and Rudin draw on the work of several forensic scientists to explain the concepts of classification and individualization. Classification is an attempt to determine a common origin; individualization uses a set of characteristics to uniquely identify a specimen. The notions are clarified using an example.

A video surveillance camera captures the shooting death of a victim. The perpetrator cannot be identified from the video, but the image is clear enough to identify the type of firearm. A bullet is recovered from the victim and submitted for examination. Based on the bullet's weight and composition, and the size and twist of the rifling marks, the examiner may be able to identify an ammunition manufacturer, the caliber of the weapon and, potentially, its manufacturer. These are all class characteristics, which, on their own, do not link the suspect to the weapon or the weapon to the bullet.

After a suspect is identified, a search reveals a box of unused ammunition and a weapon consistent with the one in the surveillance video. The characteristics of the seized ammunition are identical to the bullet obtained from the victim. As a result, it can be determined that the bullets have a common origin and are therefore "class evidence." The recovered weapon is test-fired and the resulting bullet and the bullet recovered from the victim are microscopically examined. Matching the micro-striations on the bullets allows the examiner to identify the two bullets as coming from the recovered weapon, to the exclusion of all others. This is the process of identification, which yields what is referred to as "individual evidence."

The application of these concepts to digital evidence is relatively straightforward. File systems, partitions and individual files have characteristics that allow for their classification. The location and structure of data on storage media can determine the partition type and the file system. Objects such as file allocation tables, master file tables and inodes define certain file systems. Individual files may have naming conventions as well as internal data structures (headers, footers, metadata, etc.) that determine their origin (common source). An example is a Microsoft Word file, which has a well-documented internal structure. It

would be accurate to describe the origin of such a file as being produced by Microsoft Word. All of these are class characteristics. Conversely, a file may be positively identified based on its mathematical signature (i.e., hash value), which corresponds to the process of identification.

4.3 Association

Inman and Rudin bemoan the lack of an accepted definition of the term "association" in the forensic context. They proceed to define it as "an inference of contact between the source of the evidence and a target."

Inman and Rudin use an example where reference fibers are compared with the fibers actually found on a body. When considered in the context of all the facts in the case and all other sources of the same fibers, the examiner may be able to justify a conclusion that the victim had been in contact with a particular source of the fibers.

The physical transfer of evidence is uncommon in digital evidence cases, but it does occur. An item of digital media may be linked to a computer by Windows Registry entries [12]. In malware and intrusion cases, it is often necessary to link the presence of specific files or code to the perpetrator and victim computers. The association of files is also important in intellectual property investigations.

In digital forensics, it is necessary to identify the items (files, data structures and code) that need to be associated and to determine where they might be located and the tools that could be used to locate the items. The required information is then extracted and the associations are presented.

Lee and Harris [7] observe that forensic evidence may demonstrate the commission of a crime (*corpus delecti*) or document the methodology of the crime (*modus operandi*). They identify other modalities, but most of them overlap with the Inman and Rudin taxonomy and are not addressed here. However, Lee and Harris describe one additional area that must be discussed in the context of digital forensics – that of providing investigative leads.

Computers and digital media are potentially valuable sources of lead material. The problem, from the time management and efficacy perspectives, is that it is difficult to define specific goals and objectives for many categories of lead material. Some will be discovered in the normal course of identifying material on known targets. Much will not and will only be linked based on a thorough knowledge of the case, the crime or both. This situation has often been used to justify the assignment of

sworn officers to forensic duties. However, the discussion of this issue is beyond the scope of this paper.

4.4 Reconstruction

Inman and Rudin define reconstruction as the "ordering of associations in space and time." Reconstructing a series of events is more common in the engineering fields than in the physical and biological sciences. It is, perhaps, more common in digital forensics than other fields because of the dates and times stamped on metadata pertaining to data, files, file systems and network communications. It is important to recognize, as Inman and Rudin do, that time is often a relative value or ordering rather than a definitive value.

In cases involving the creation and/or alteration of documents or images, the files and file systems may provide information about sequences of events if not the exact dates and times of the events. Comparing file or e-mail metadata may permit the "normalization" of dates and times from multiple computers within a margin of error. Using monitor software, it is possible to observe and document changes to files and file systems that result from the execution of computer code. Generally, the more data points considered and the more consistent the metadata, the more probable that the specific event sequence is correct.

5. From Principle to Question

Inman and Rudin state:

> "Before the criminalist ever picks up a magnifying glass, pipette or chemical reagent, he must have an idea of where he is headed; he must define a question that science can answer."

This seemingly simple statement in many ways defines the forensic case management problem. It is important to understand how to define an examination as one or a series of investigative or legal questions, which are translated into scientific questions (to use Inman and Rudin's terminology). This suggests a two-part process: defining the legal/investigative questions and then – and only then – defining the digital forensic (scientific) questions.

While this seems obvious, it is not how many examinations are developed. Often, the investigator provides a case synopsis to the examiner and asks the examiner to study the evidence and provide any and all information that might be useful. Sometimes, the examiner will think, even before the investigator has finished speaking, about what could be done. This results in an examination being designed based on what could be done instead of on the specific information that should be located.

The alternative proposed here is to begin by defining the legal or investigative questions that the investigator thinks could be answered from the information contained in the evidence. The examiner may well need to discuss the questions with the investigator, continuously refining the requirements and providing feedback on what is possible, likely and remote. Time spent developing the investigative questions pays off in the ability of the examiner to translate them accurately into an efficient examination plan that is responsive to the legal/investigative questions and that is supported by science. An important part of this discussion is for the examiner and investigator to mutually understand the tasking and the limitations on the potential results. The latter is important for several reasons. Over-reliance on low probability results is misleading, and it may become the weak link in a courtroom presentation. Expending a great deal of examiner effort to produce information of limited value is a poor use of resources. Experience has demonstrated that the process also helps manage investigative expectations.

Once the legal/investigative questions are finalized, the examiner can begin to develop the scientific questions. It is here that the forensic processes discussed above become relevant. Most investigative/legal questions can be translated directly into one or more of the four processes.

For example, several forensic questions can be created to answer whether or not information concerning a particular person is present in a specimen. What name(s) should be searched? Where will information about the person(s) be located? Are there any temporal constraints on when this information might appear? Having answered these questions, the next step is to select a technique or tool that can locate the information.

The above is an example of the identification process. A classification question would involve locating all the images relevant to a certain investigation. Matching an image located online or on another computer to an image found on the specimen computer is an example of individualization.

Investigators could benefit by connecting cameras to images, users to accounts and activities, computers to network connections, and devices to computers. Each of these involves the specification of an association question. Malware, intellectual property and intrusion investigations often rely on the presentation of a sequence of events and the demonstration of cause and effect; these would require the framing of reconstruction questions. When investigative questions are translated into questions based on forensic processes, examiners can develop efficient and objective tests that yield definitive conclusions.

Perhaps the most valuable aspect of this process is that it provides a definitive end to an examination. Many forensic examinations languish because the examiner does not know when the case is finished. If an examination is designed based on what is possible, the examination will never be completed because it is always possible to do more. However, if the questions are defined at the outset, the examination is done when all the questions have been answered. Note that it does not matter what answers are obtained, just that they are accurate.

6. Conclusions

Traditional forensic science has developed an effective and relatively efficient process that has stood the tests of time and the courts. Digital forensics practitioners can learn much from this process. Incorporating the development of forensic questions into the examination process ensures scientific objectivity while simultaneously assisting in case management. Managers can use this approach to leverage their limited resources. Educators can also utilize the approach to ensure compete and consistent results from training programs.

References

[1] V. Baryamureeba and F. Tushabe, The enhanced digital investigation process model, *Proceedings of the Fourth Digital Forensic Research Workshop*, 2004.

[2] N. Beebe and J. Clark, A hierarchical, objectives-based framework for the digital investigation process, *Proceedings of the Fourth Digital Forensic Research Workshop*, 2004.

[3] B. Carrier, Defining digital forensic examination and analysis tools using abstraction layers, *International Journal of Digital Evidence*, vol. 1(4), 2003.

[4] B. Carrier and E. Spafford, Getting physical with the digital investigation process, *International Journal of Digital Evidence*, vol. 2(2), 2003.

[5] K. Inman and N. Rudin, *Principles and Practices of Criminalistics: The Profession of Forensic Science*, CRC Press, Boca Raton, Florida, 2001.

[6] T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, Illinois, 1970.

[7] H. Lee and H. Harris, *Physical Evidence in Forensic Science*, Lawyers and Judges Publishing Company, Tucson, Arizona, 2000.

[8] G. Mohay, A, Anderson, B. Collie, O. de Vel and R. McKemmish, *Computer and Intrusion Forensics*, Artech House, Boston, Massachusetts, 2003.

[9] A. Mozayani and C. Noziglia, *The Forensic Laboratory Handbook: Procedures and Practice*, Humana Press, Totowa, New Jersey, 2006.

[10] National Archives and Records Administration, The Future, College Park, Maryland (www.archives.gov/about/history/building-an-archives/statues/statue-future.html).

[11] National Institute of Standards and Technology, National Software Reference Library, Gaithersburg, Maryland (www.nsrl.nist.gov).

[12] B. Nelson, *Guide to Computer Forensics and Investigations*, Thompson Course Technology, Boston, Massachusetts, 2006.

[13] M. Pollitt, An ad hoc review of digital forensic models, presented at the *Second International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2007.

[14] M. Reith, C. Carr and G. Gunsch, An examination of digital forensic models, *International Journal of Digital Evidence*, vol. 1(3), 2002.

[15] R. Saferstein, *Forensic Science Handbook, Volume II*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[16] P. Stephenson, Modeling of post-incident root cause analysis, *International Journal of Digital Evidence*, vol. 2(2), 2003.

[17] C. Welch, Flexible standards, deferential review: Daubert's legacy of confusion, *Harvard Journal of Law and Public Policy*, vol. 29(3), 2006.