Chapter 1

DIGITAL FORENSICS AS A SURREAL NARRATIVE

Mark Pollitt

Abstract Digital forensics is traditionally approached either as a computer science problem or as an investigative problem. In both cases, the goal is usually the same: attempt to locate discrete pieces of information that are probative. In the computer science approach, characteristics of the data are utilized to include or exclude objects, data or metadata. The investigative approach reviews the content of the evidence to interpret the data in the light of known facts and elements of the crime in order to determine probative information or information of lead value. This paper explores two literary theories, narrative theory and surrealism, for potential application to the digital forensic process. Narrative theory focuses on the "story" that is represented by text. At some level, a storage device may be viewed as a series of interweaving, possibly multi-dimensional, narratives. Furthermore, the narratives themselves, coupled with the metadata from the file system and applications, may form a meta-narrative. The literary theory of surrealism, the notion of disjointed elements, can be utilized to derive meaning from forensic evidence. This paper uses a technique known as surrealist games to illustrate the point.

Keywords: Digital forensics, narratology, surrealism

1. Traditional Approaches to Digital Forensics

Most digital forensic examinations are done in the context of an investigation. The items that are examined are collectively referred to as "digital evidence." The Scientific Working Group on Digital Evidence (SWGDE) [21] defines digital evidence as:

> "Information of probative value stored or transmitted in binary form."

Numerous methodologies have been proposed for digital forensics (see, e.g., [5–8, 13, 17, 19]). While the methodologies differ from each other,

they have in common the goal of preserving the integrity of the original evidence and extracting information of value to the case at hand. Implicit in every methodology is the notion that the original evidence is potentially massive in size and that the "important" information is a subset, often a very small subset of the original evidence.

Numerous software tools have been developed to support digital forensic investigations: examples include EnCase, Forensic Toolkit, ProDiscover and Sleuth Kit. Most tools utilize file system structure, file type and string searches, and hash value comparisons. Novel approaches such as data mining and social network analysis have also been proposed [2, 12]. Most of these approaches rely on the technical characteristics imparted to the data by the operating system and file system. Other approaches rely on the identification of discrete bits of information whose character can be predefined, e.g., string searches and data mining. But the results, even using novel techniques such as fuzzy searches [10], have been modest at best. We believe that the effective yield – the portion of pertinent information selected by forensic techniques – is becoming not more selective, but less effective. After the potentially probative information is extracted, it becomes an analytical exercise to evaluate the information in a contextual manner. In other words, how does the information fit into the narrative of the investigation? In many cases, we try to fit the digital pieces into the framework of the case, which is, in turn, framed by the presumptive fact pattern and the elements of the law as they apply to the pattern of facts.

2. Legal Issues

One of the hallmarks that distinguishes forensics from other applications of science is the requirement that the results be accepted as reliable evidence in a court of law. In the United States, the admission of scientific evidence is also tied to an expert witness who presents the evidence in court. An extensive body of law deals with the relationship between the evidence, the examiner and the testimony provided relative to the evidence. Two important elements are the empirical nature of the examination process and the ability of the examiner to explain the science as applied to the evidence. In the traditional model, it is important that the examination process be conducted in a "forensically sound" manner, i.e., all actions must be empirically demonstrable in both process and product.

Practitioners generally apply a scientific process to extract the evidence prior to its review for pertinence or investigative value. This tends to limit the use of novel digital forensic methodologies. If it is as-

sumed that only the resultant facts must be demonstrably factual, then it is possible to greatly expand the potential approaches for identifying probative evidence. For sake of argument, there is no reason why we could not find the information first and then ensure its reliability by referring to an empirical process. Note, however, that we are not suggesting that this approach is a "best practice" or that it is necessarily superior to the traditional process.

The remainder of this paper explores novel digital forensic approaches.

3. Guessing the Future

Novel approaches may prove fruitful given the potential environment of the future. For the purpose of this analysis, we make the following assumptions:

- The size of the storage media to be examined will continue to grow.
- Computing devices will be used for an increasingly large number of applications.
- A greater and greater proportion of data related to a person's life will reside in electronic storage.
- We will be able to tell many more things about a person from examining electronic media.

The first assumption is highly probable; several studies have discussed this trend (see, e.g., [16]). The second assumption seems to be borne out of the phenomenon of network convergence as evidenced by devices such as the iPhone. The data related to these multipurpose devices often resides in multiple locations. Unlike the traditional model of the desktop being the primary repository of data augmented by network storage, Web 2.0 applications such as shared calendars, blogs, wikis and social networking sites store data in a slew of application servers independent of the access devices. An important characteristic of many Web 2.0 applications is "personalization," which implies that a dataset closely represents the owner.

Credit/ATM cards, access cards, toll transponders, cell phone records and network connections produce digital recordings of a person's activities. Desktops, laptops and physical media such as flash cards store our most private information. Clearly, future generations will record much more of their lives than any previous generation and the recordings will be captured and stored electronically in the web of the future. All of this should not surprise us. But how will it impact digital forensics? Perhaps a different perspective will help illuminate the issue.

4. Hard Drive as Text

Recording the breadth of our lives in digital form can be viewed as producing "text." Not merely numbers, letters, words, sentences, paragraphs and pages, or even ASCII code, but text in the broadest sense of the recording of human thought, communications and activities. One of the many definitions for text is [22]:

> Something, such as a literary work or other cultural product, regarded as an object of critical analysis.

For the purposes of digital forensics, we can regard media as cultural products or artifacts and the sum of all media associated with an individual as a collective literary work. There are many texts within this work: letters, essays, emails, graphs, charts, diagrams, photographs, audio and video recordings. Program files are also texts, as are log files of computer and network activity. Computer media and, by extension, all the locations where an individual stores information constitute a digital anthology of that person's texts.

Many kinds of texts are contained in the digital anthology. Some of the texts are distinct, some overlapping and some redundant. The texts are of many forms. This paper focuses on one form of text, that of the "narrative."

5. Narrative Theory

Part of the definition of texts that we have considered above includes the notion of critical analysis. Students of the humanities have been conducting critical analyses of texts for hundreds of years. This activity is generally referred to as literary criticism. However, since the 1930s and, especially, due to the work of Claude Levi-Strauss [14], its use has expanded and it is now an important part of cultural anthropology. Modern cultural anthropologists and ethnographers such as Wesch [23] view electronic media as a cultural artifact and as a research instrument.

Mieke Bal, in her landmark text, *Narratology: Introduction to the Theory of Narrative* [1], defines a narrative simply as "a text in which a narrative agent tells a story." The notion of a story seems to be a topic of literary discourse than of forensic science. However, on reflection, it should be clear that forensic scientists and investigators are well acquainted with stories. In a very real sense, forensic scientists attempt, through their examinations, to determine the facts and circumstances that form the "story" of the crime. It involves the time, place, characters and action – the very things that make up a novel. Later, in the presentation phase, the examiner writes a report that communicates the "story" to the reader. When the examiner testifies at a trial, the

testimony conveys both the examiner's story and the story of the crime. In a real sense, forensic science is not about isolated, discrete facts, but a storytelling that communicates meaning. In our experience, forensic clients mostly ask questions about meaning, not about facts. Forensic examiners are often well equipped with facts but are not equipped to handle the meaning of the evidence.

Literary criticism is about the search for meaning. Scholars have for centuries been examining the process and the practice of human communication [18]. Many scholars have examined and continue to examine how meaning is extracted from texts. For the purposes of this paper, it would not be productive to discuss the multitude of critical approaches and schools of thought concerning the narrative. Instead, we will explore one methodology that might provide insight into how theories developed for literary criticism could be of value in a digital forensic setting.

6. Surrealism

Surrealism developed out of the despair of World War I and the rise of Dadaism. This philosophy sought new sources of inspiration from the world and the artist's mind. In his 1924 book, *Les Manifestes du Surrealisme*, Andre Breton [4] defined surrealism as:

> "Psychic automatism in its pure state, by which one proposes to express – verbally, by means of the written word, or in any other manner – the actual functioning of thought. Dictated by the thought, in the absence of any control exercised by reason, exempt from any aesthetic or moral concern."

While Breton studied psychiatry, he was not a Freudian psychoanalyst. Rather, he was trained in the French system of "dynamic psychiatry," which focused more on the elimination of the conscious than on the "unconscious" in the Freudian sense. Gibson [9] points out that Breton uses the term "depths of the mind" as opposed to the unconscious in *Les Manifestes du Surrealisme*. The significance of this concept will become apparent later in this exposition.

The initial focus of surrealism was the written word, although in time practically every form of art developed a surrealist school. The early stages of surrealism were characterized by the juxtaposition of seemingly disparate items, situations and ideas. Later, others, such as Benjamin [3], began to integrate the use of montage as a surrealist technique. In montage, isolated parts of a whole, such as sections of a photograph or portions of a projected film, were viewed in isolation. Surrealists would even use techniques such as viewing films through their fingers or a piece of cloth to disrupt the flow of the narrative [11]. They quickly discovered that these isolated pieces when placed together created a form

of narrative. The technique was seen as a method for creating art and for encouraging creativity. Eventually, it became both work and pleasure – when games were developed based on the notion of montage.

7. Surrealist Games

Surrealist games appealed to the adherents of surrealism for a number of reasons. Jean-Louis Bedouin described them as "a way of being serious without the worry of seeming so" [20]. It was the "automatic" aspect of the games that acted to suppress the conscious and free the unconscious, which it was hoped, would fuel creativity. Meanwhile, the "play" aspect served to minimize the motivation for conscious intervention in the resulting product [20].

Two popular surrealist games are Exquisite Corpse and Irrational Extension. Exquisite Corpse is a parlor game in which each player writes down a word from an assigned part of speech. These words are then combined into a phrase or metaphor. According to Ray [20]:

> "Since the philosophy of science has shown that all knowledge systems rest on a few basic metaphors, and that a new paradigm always proposes a metaphoric shift, this game might have more profound consequences than at first appears."

Irrational Extension games take a different tack. A movie is chosen as the "source" of the answers to questions posed by the person running the game. The questions are designed to be outside the scope of the source material, that is, they are designed so that there is no correct or even logical answer. The players are free to provide answers without the requirement of factual support. In addition to being humorous, these games point out holes or gaps in a narrative. By finding the questions not answered by a film, we "see" the things that our minds were likely thinking, but that were suppressed by our consciousness. Ray [20] likens this process to Freud's discovery that "resistance and repression were essential to the diagnostic process: they, in fact, pointed directly to the determining areas of a patient's experience, those leading to his symptoms."

Dove, a student of Ray, created a variant of the Irrational Extension game. In his version, the experimenter selects three "narratively important shots" (frames) and three randomly selected insignificant shots from a movie unfamiliar to the subject. All six are then presented to the subject in random order. The subject is asked a series of questions after each frame. Some of the questions are factual while others are speculative. According to Ray [20], this experiment "encourage[d] a sensitivity both to meanings communicated stereotypically and to those unconsciously." The experiment often produced unexpected results – the subject was

Figure 1. Photographs of women used in the experiment.

able to correctly identify important elements of the narrative as well as recognize important symbolism in the images.

Dove's game brings us full circle in that the power of the subconscious is used for cognition. On reflection, this should not be surprising. We are exposed to literally thousands of discrete pieces of information in our daily lives. If you were to look up from this page, you would be faced with myriad objects, sensations, responses, relationships and emotions. If you look back down, you will consciously only remember a small number of "facts," but your unconscious mind will use some of the information to contextualize the "facts;" the rest will evaporate. In a sense, we proceed through life as a series of "shots," each of which is interpreted and stored as a subset of the total image. According to Hammond [11]:

> "Stripped of their causal relations in the film, a rapid-fire of reported images emphasizes their latent content, their capacity to signify."

To test this notion, we performed an experiment in a graduate class. The class was divided into two groups, each of which was presented with one of two photographs (Figure 1). The photograph on the left, taken in Greek Cyprus, features an elderly widow dressed in the traditional black garb. The one on the right, showing a woman wearing a full-length abaya and a burqa, was taken in Alexandria, Egypt.

The students were asked five questions in the following order:

(a) What is this woman thinking?

(b) What is the next thing she will do?

(c) Why will she do it?

(d) Where is she?

(e) Why is she here?

The responses to the first three questions were predictably irrelevant as there is no information to support any objective conclusion. The answers to Questions (b) and (c) were determinable by the photographer, but not the students as they were not present at the scene. But the answers to Question (d) were objectively supportable. However, the students' responses were invariably incorrect because they did not have sufficiently detailed knowledge of the cultural contexts to answer Question (d) correctly. Since the students' responses to Question (e) were generally predicated upon their responses to (d), their responses to (e) were also incorrect.

This experiment is clearly not empirical. Rather, we wished to explore with the students how much we "know," how much we construct and when we project into the scene. This experiment seems to suggest that cognition, even subconscious cognition, requires some contextual information. The relative success of Dove's experiment may lie in his familiarity with plots, cinematic styles and history. The failure of our experiment may be due to a lack of geographic, cultural and religious background information. An additional issue may be that films are inherently narrative (a fact known to both Dove and his subjects) while photographs are not necessarily narrative. Likewise, the use of multiple "shots" in Dove's experiment provides more raw material for the subconscious.

8. Applying Surrealist Games

So how can such a seemingly unscientific methodology be useful in forensic science? Surrealists seek to disrupt the narrative to understand the meaning of the constituent elements of a work. In order to do so, they use techniques that eliminate the narrative and suppress the context. In forensics, we do the opposite – we attempt to use the data points to discover a context and document a narrative. However, two elements of surrealist games can be exploited for forensic use: narrative context and montage.

One of the things that makes the Exquisite Corpse and Irrational Extension games and Dove's extension somewhat effective is that there is a logical progression from structure to context. In an Exquisite Corpse game, the grammatical parts of speech that are to be contributed are specified. The context and any diegesis are prevented from being provided by the secrecy used to submit the words. In our view, individuals often have an internal consistency in their submissions, which may indicate an attempt at providing a personal context. Moreover, their responses tend to have a theme.

In an Irrational Extension game, the plot and the characters present a structure and a basic context. According to Ray [20]:

> "We think that after nearly two hours with the Smith family, or fifteen movies with the Hardys, we must know everything about them and their house and their neighborhood."

In a film or other narrative, the characters are introduced, the scenes unfold with all of their subtext and relationships develop. The reader or viewer tries to fill in the blanks to create a "complete" understanding. In cinema, the story is never complete, as historian Natalie Davis was told concerning the "Camel Principle." It is not necessary to include a multitude of details to convince the audience that the scene is in a desert; the mere presence of a camel will suffice [20].

Life is fuller and much more complex. There are many more "data points" with which to complete the narrative of even one person's life. In the context of a real person with a real life, that person has a complex web of complete narratives with subplots. In fact, the problem is not a lack of data and context, it is limiting it to a manageable amount that provides an accurate picture of the individual.

In a criminal investigation, we seek to identify all the data points that are pertinent to the "subplot" of criminal activity. The person has a "complete life," only some of which is criminal in nature. If we were to surveil a criminal suspect around-the-clock, we would see that most of his activities would not be relevant to the criminal case.

A criminal who uses a computer for a length of time records a great deal of his complete narrative (life) and likely a good part of the criminal subplot on his computer hard drive. Examination of the hard drive will allow for a complete – in the sense that it exists on the hard drive – narrative of the person and his activities. It is the task of the examiner to identify and extract the portion that documents the criminal behavior. For the purposes of this discussion, the narrative and context of the criminal subplot can be developed externally through a normal investigation and/or internally through an examination of the material contained within the computer's storage (hard drive).

The second aspect of surrealist games that can be applied to digital forensics is the notion of montage. As discussed above, our own recollection of our personal narrative is a series of memories (data points) that collectively represent our subjective history. Our lives can be compared to a series of snapshots and video clips. Cinema intentionally creates montages of sequenced scenes, which are designed to convey a narrative. Photographs are frozen moments in time, in effect, a time capsule.

A digital forensic examination report is very similar because it contains selected data points that are arranged into a narrative. This narra-

tive can be organized chronologically, topically or based on the structure of the examined media.

Since all the material on a hard drive is not included in the report, the "completeness" of the narrative is subjective and its effectiveness is measured against its "centrality" to the narrative or its pertinence to the case.

We can make use of these two aspects in the following way. We can construct a narrative externally via a traditional investigation. In this instance, we will be developing the context for the examination of the hard drive. Alternatively, we can attempt to construct the narrative and context solely by examining the content of the hard drive and any other stored electronic evidence. This, in effect, determines what the user was doing with the computer.

Next, we can select a set of data points from the evidence. Our selection of data points (files, emails, etc.) can be used to populate a narrative and/or to develop the context. The selection of these data points can be random or constrained. If it is random, it is reasonable to conclude that a larger number of data points will be necessary. If it is constrained, the effectiveness of the constraints will be a function of the centrality to the narrative. In other words, how likely is it that the constrained material contains pertinent information?

We use correspondence from the author's university email account as an example. On March 20, 2008 there were 238 emails in the account that were received during the period, March 1, 2008 through March 20, 2008. If you were to review the correspondence, you would conclude that the emails pertained to communications between the author and his superiors, peers and students. You would infer that the author was teaching two courses, enrolled as a student in two others, and active in several program committees and editorial boards. You would know about the author's research and outreach projects. You would even find a few emails from friends. The email correspondence provides a fairly complete picture of the author's professional life, since the author essentially limits the use of his university email account to professional purposes.

What if you only read eight out of every ten emails? There is a good chance that, while you might miss a few details, the narrative would still be substantially complete. Others might wish to challenge that assertion, but emails are often a series of communications and often quote previous text. As a result, the loss of 20% of the corpus is not significant; much of the information is replicated in other emails. Further, many of the "missing" details can be surmised from a careful reading of the remaining material, much like the subconscious information presented in film. We

have, in effect, a "montage of data" with sufficient context to correctly answer at least some of the pertinent questions.

But would you have any useful information if you were to read only 1% of the email corpus? Unless the context is thoroughly known and the place of the data in the context is defined, it is unlikely you would have useful information. Somewhere between the 80% montage and the 1% montage, there is a sliding scale or, perhaps, a polynomial expression, that determines when the information gathered is useful. Dove's experiment demonstrated that accuracy can be attained if contextually rich sets of information are used to form even a small montage. As Benjamin [3] states:

> "In the fields with which we are concerned, knowledge comes only in flashes. The text is the thunder rolling long afterward."

9. Conclusions

At one level, we may ask how this is any different from the way investigators already conduct the analytical aspect of forensic examinations. The answer may well be that we are consciously or subconsciously operating in this fashion. If so, surrealism, as explored in this paper, confirms this methodology as cognitively legitimate. It would benefit the digital forensic community to utilize these insights in order to perform examinations and analyses more effectively and efficiently.

The core concepts of narrative and montage are powerful tools of cognition, but they require context. Designing a system that exploits these concepts would require at least three things. First, the electronic data should be parsed into narratives. Second, these narratives and the investigative context should be coded to enable computational solutions. Third, it is necessary to understand how differing levels of narrative completeness impact the ability to make objective, sound conclusions. These requirements are not trivial and require very different approaches from those that have traditionally been used in computer science research.

The notion of utilizing surrealist gaming techniques as a digital forensic research method is not as farfetched as it might appear. One of the key issues is how to incorporate the concept of cognition into the mechanistic process of extracting information from a hard drive. Our understanding of that cognitive process currently lacks rigor. Surrealist techniques do not explain cognition, rather, they provide us with another way to "read" the hard drive. The combination of narrative and montage can inspire a new genre of powerful digital forensic tools. We must be mindful, however, of Adorno's criticism of Benjamin [15]:

> "Your study is located at the crossroad of magic and positivism."

References

[1] M. Bal, *Narratology: Introduction to the Theory of Narrative*, University of Toronto Press, Toronto, Canada, 1997.

[2] N. Beebe and J. Clark, Dealing with terabyte data sets in digital investigations, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 3–16, 2005.

[3] W. Benjamin, Theoretics of knowledge; theory of progress, *Philosophical Forum*, vol. 15(1-2), pp. 1–40, 1984.

[4] A. Breton, *Les Manifestes du Surrealisme*, Jean-Jacques Pauvert, Paris, France, 1972.

[5] S. Bunting and S. Anson, *Mastering Windows Network Forensics and Investigation*, Sybex, Alameda, California, 2007.

[6] B. Carrier, *File System Forensic Analysis*, Addison-Wesley, Boston, Massachusetts, 2005.

[7] H. Carvey, *Windows Forensic Analysis*, Syngress, Rockland, Massachusetts, 2007.

[8] E. Casey, *Digital Evidence and Computer Crime*, Academic Press, Boston, Massachusetts, 2004.

[9] J. Gibson, Surrealism before Freud: Dynamic psychiatry's "simple recording instrument," *Art Journal*, vol. 46(1), pp. 56–60, 1987.

[10] J. Guan, D. Liu and T. Wang, Applications of fuzzy data mining methods for intrusion detection systems, *Proceedings of the International Conference on Computational Science and its Applications*, pp. 706–714, 2004.

[11] P. Hammond, *The Shadow and its Shadow: Surrealist Writings on the Cinema*, City Lights Books, San Francisco, California, 2000.

[12] M. Hoeschele and M. Rogers, Detecting social engineering, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 67–77, 2005.

[13] W. Kruse and J. Heiser, *Computer Forensics: Incident Response Essentials*, Addison-Wesley, Boston, Massachusetts, 2001.

[14] E. Leach, The social theory of Claude Levi-Strauss, *British Journal of Sociology*, vol. 33(1), pp. 148–149, 1982.

[15] H. Lonitz (Ed.), *Theodor W. Adorno and Walter Benjamin: The Complete Correspondence, 1928-1940*, Harvard University Press, Cambridge, Massachusetts, 1999.

[16] P. Lyman and H. Varian, How much information? University of California, Berkeley, California (www2.sims.berkeley.edu/research/projects/how-much-info-2003), 2003.

[17] A. Marcella and D. Menendez, *Cyber Forensics: A Field Manual for Collecting, Examining and Preserving Evidence of Computer Crimes*, Auerbach Publications, Boca Raton, Florida, 2007.

[18] W. Ong, *Orality and Literacy*, Routledge, New York, 2002.

[19] M. Pollitt and S. Shenoi (Eds.), *Advances in Digital Forensics*, Springer, Boston, Massachusetts, 2005.

[20] R. Ray, *The Avant-Garde Finds Andy Hardy*, Harvard University Press, Cambridge, Massachusetts, 1995.

[21] Scientific Working Groups on Digital Evidence and Imaging Technology, SWGDE and SWGIT Digital and Multimedia Evidence Glossary (www.swgde.org/documents/swgde2008/SWGDE_SWGITGlossaryV2.2.pdf), 2007.

[22] TheFreeDictionary.com, text (www.thefreedictionary.com/text).

[23] M. Wesch, Digital ethnography (mediatedcultures.net/about.htm).