Digital Investigation 24 (2018) S83—S92

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2018 Europe — Proceedings of the Fifth Annual DFRWS Europe

Controlled experiments in digital evidence tampering

Felix Freiling*, Leonhard Hösch

Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

Keywords:
Anti-forensics
Digital evidence
Forensic computing education

ABSTRACT

We report on a sequence of experiments performed with graduate level students on the tampering of digital evidence. The task of the study participants was to manipulate a given disk image so that it looked as if a website had been accessed and images downloaded in the past. Later, the same students had to distinguish their forgeries from a set of originals in which the images actually had been downloaded. During all parts of the experiment, efforts were recorded in project diaries. Overall, the results show that the tampering task was difficult since none of the forgeries was taken as an original. Furthermore, the analysis effort to detect forgeries consistently was below the effort to create the forgery even in the worst case scenario where the manipulator had full control over the evidence. It also required generally less effort to correctly classify an original than to correctly classify a forgery. Additionally, we derived results confirming that the effort to construct consistently manipulated evidence increases with decreasing control, i.e., the ability to precisely act upon the evidence.

Introduction

In the history of forensic science, there is a well-established tradition to document the experiences in handling and interpreting evidence, which in the last 150 years over-whelmingly has been *physical* evidence (Kirk and von Thornton, 1974; Groß and Geerds, 1977; Lee and Harris, 2000). Today, however, an increasingly large portion of evidence in criminal cases is *digital* evidence, i.e., evidence that is stored on or transmitted over digital media (Casey, 2011). There has been much philosophical debate about the "nature" of digital evidence and how it differs from the characteristics of physical evidence (Dardick et al., 2014; Paul, 2009). One of the resulting issues is the volatile nature of the binding between support and information, which makes digital evidence — at least in principle — more susceptible to manipulation. In the community of digital forensic analysis, there appear to be ambivalent opinions whether digital evidence can be perfectly tampered. Many appear to believe this to be the case, as expressed by Caloyannides (2003), who states that digital data can be manipulated at will, and depending on the manipulator's sophistication, the alteration can be undetectable, regardless of digital forensics experts' competence and equipment.

Digital forensics processes have tried to cope with this problem and long since established standard processes that try to contain the dangers of manipulating digital evidence. The most prominent of theses methods is the use of cryptographic hash functions to document the state of a string of bits within the chain of custody. But even though cryptographic hash values are an established part of digital forensics processes, they cannot help to detect manipulations that have occurred *before* evidence collection, be it either direct tampering by the suspect or evidence planting by corrupt law enforcement officers.

Related work

Questions of manipulation appear to surface in a non-negligible number of practical court cases. Briefly spoken, there are often two opposing views:

1. One side, usually the prosecutor, claims that the state of collected evidence is consistent with a particular hypothesis H_1 of how the crime happened.
2. The other side, usually the defendant, claims that the collected evidence was manipulated such that it appears as if hypothesis H_1 were true, but in fact a different and opposing hypothesis H_2 is consistent with the evidence, which includes the manipulation and a different sequence of events implied by H_1.

In the literature, this phenomenon has been termed the "Trojan Horse Defense" (Brenner et al., Henninger) where the defendant claims that not he or she committed the offense (such as

* Corresponding author.
E-mail addresses: felix.freiling@cs.fau.de (F. Freiling), leonhard.hoesch@fau.de (L. Hösch).

https://doi.org/10.1016/j.diin.2018.01.011

performing a cyberattack or downloading illegal documents) but rather a Trojan horse installed on the computer on behalf of a third party. Unfortunately, there is not much concrete advice in the literature on how to technically deal with such cases apart from general statements such as the following [8, p. 49]:

The investigator should use all available resources to determine if a remote person could have used the application to commit the crime or to install additional software that could have committed the crime.

Obviously, in such cases also other forms of evidence are critical, e.g., the fact whether the suspect has sufficient knowledge to manipulate, hide or wipe evidence from the system. But little is known on the actual effort it takes to change digital evidence such that the "true" hypothesis H_2 might be mistaken for the "false" hypothesis H_1.

Moch, (2015) reports on some preliminary experiments with students that were instructed to manipulate evidence as follows: Students were given two disk images I_1 and I_2, where I_2 resulted from I_1 by executing a certain action (like sending a message on ICQ). The task of the students was to mount I_1 and manually change the evidence such that the same evidence as that in I_2 was present but without actually performing the action that originally created the evidence. The results showed that it was extremely difficult to perfectly manipulate the evidence because manual operations (like invoking vi or touch) create other artifacts (such as swap files or timestamps of value zero).

Another experiment by Moch, (2015) exhibited a problem to swap the content of two files in an ext4 file system because the metadata (especially the inode number) remains unchanged. Therefore, the set of differences between the manipulated image and the target image I_2 (calculated using `idifference` (Garfinkel, 2012)) often was not empty although students had perfect control over the image and the manipulation process. In conclusion, perfect manipulations (defined as an empty difference set) were possible but needed an extreme effort and care for detail. Unfortunately, this effort was not quantified.

Related to the problem of evidence manipulation is the area of *anti-forensics*, meaning "any attempts to compromise the availability or usefulness of evidence to the forensics process" (Harris, 2006). Interestingly, the literature on anti-forensics has mainly focused on rather obvious and aggressive techniques, such as hiding or encrypting evidence (Berghel, 2007; McDonald et al., 1999), overwriting/wiping evidence (Foster and Liu, 2005; Savoldi et al., 2012) or attacks against investigative tools (Wundram et al., 2013). Maybe the most advanced area in the analysis of manipulated (or counterfeit) evidence is multimedia security, e.g., where methods of blind image forensics can be used to detect manipulations (Johnson et al., 2006; Lin et al., 2009). However, we are not aware of any literature with a similar intention focusing on non-multimedia files.

When speaking to experienced investigators, many might agree that digital evidence can *theoretically* be manipulated perfectly, but *in practice* it is very hard to do this and not make mistakes. So while this indicates that every manipulation appears to also leave traces that can be detected, we are not aware of work that has systematically explored the effort to perform targeted evidence manipulation.

Research goal and contributions

With the increase of the amount of digital evidence in court, it must be expected that also the number of attempts to counterfeit, manipulate or forge such evidence will increase. Therefore, expert witnesses in digital forensics should be prepared to react to efforts by any of the opposing sides in the spirit of the Trojan horse defense. In this direction it is not only necessary to question the competence and motivation of a suspect to forge evidence, but also to

1. look for evidence of manipulation and
2. in case no such evidence can be found, to understand the effort necessary to perform such perfect manipulations.

In analogy to the handling of physical evidence one can then argue, that if the effort for evidence manipulation is very high and there is no evident motivation or competence of the suspect to forge evidence, then it is more probable that there has been no manipulation than the opposite.

In this paper we study the effort to perform an evidence manipulation task by running a controlled experiment within a graduate level course on digital forensics at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) in Erlangen, Germany. Our focus was to study the class of manipulations that *make forgeries look like originals* and where no trace of tampering could be found. This is in contrast to many practical cases of tampering where evidence is blatantly overwritten or destroyed.

More specifically, the task involved to manipulate a given system disk image so that it appeared that files had been downloaded from a particular website in the past whereas in fact they had not been downloaded at that time. Independently we prepared a set of original evidence (i.e., evidence where the download action had actually happened). After performing the manipulation task, students had to investigate a randomly selected disk image and had to determine whether it was original or fake. During all of these activities, students were required to document their actions and log their effort in a project diary. The goal was to study the success probability of the manipulation attempts, the effort it takes and the factors that influence the quality of a forgery.

Overall, 14 students participated in the experiment and in this paper we report on the results of the analysis of the collected data. Since we were not aware of related work that performed similar experiments before, we could only state rough research questions instead of exact hypotheses to evaluate. Still, the following statements can be drawn from the data:

- All forgeries produced within our experiments were correctly classified as forgeries. This means that it appears to be generally hard to produce a convincing forgery in the given case.
- It required generally less effort to correctly classify an original than to correctly classify a forgery. This appears surprising since one would expect that forgeries can be classified as soon as obvious signs of tampering are found, whereas such signs are absent in originals.
- *Producing* a forgery generally requires more effort than *detecting* that a forged image is a forgery. This holds even in perfect manipulation environments where there are no restrictions on the tools and methods used to produce a forgery.
- Less control over the manipulation process (reduced toolset, more uncertainty over evidence) increases the effort to produce a forgery and reduces the effort to correctly detect a forgery as fake.

While our findings are limited and can be described as preliminary, we believe that the collected data, which is available online (Freiling and Hösch, 2018), will be helpful to shape further experiments in this relevant field in the future.

Paper outline

This paper is structured as follows: We first formulate the research questions along which the study was designed in Section 2. We then describe the experimental design of our study in Section 3. We report on the detailed quantitative (Section 4) and qualitative results (Section 5). We conclude in Section 6.