The Trojan Made Me Do It: A First Step in Statistical Based Computer Forensics Event Reconstruction

Megan Carney and Marc Rogers
CERIAS
Purdue University

Abstract

The current study was exploratory and represents a first attempt at a standardized method for digital forensics event reconstruction based on statistical significance at a given error rate (α = .01). The study used four scenarios to test the ability to determine whether contraband images located on a system running Windows XP were intentionally downloaded or downloaded without the user's consent or knowledge. Seven characteristics or system variables were identified for comparison; using a stepwise discriminant analysis, the seven characteristics were reduced to four. It was determined that a model consisting of two characteristics-- the average of the difference between file creation times and the median of the difference between file creation times -- was the best model for discriminating the intentional action at α = .01. The implications of this finding and suggestions for future research are discussed.

Introduction

The Trojan horse defense has presented the field of forensics with a difficult challenge. It is now necessary to provide a method that reliably reconstructs events to determine guilt or innocence. Currently, testimony by computer forensics experts may leave the jury confused, since it often requires explaining details of the investigative process (Carrier & Spafford, 2004; Rogers, 2003). A standardized method would bring computer forensics closer to the kind of evidence presented for drug and DNA testing; a probability that such a substance was found at the scene and belongs to a given individual with a measurable error rate (Carrier & Spafford, 2003; Whitcomb, 2002). This work presents the first attempt at a standardized method for event reconstruction with statistically determined accuracy and error rates.

Often, individuals indicted for crimes based on digital evidence claim that a Trojan horse or virus installed on their computer was responsible. Cases in England and the United States (e.g., *Regina v. Greene, Regina v. Caffrey*) have already proven the effectiveness of this defense. The field of computer forensics must answer this challenge with a method of event reconstruction that will allow investigators to determine true guilt or innocence in a way rigorous enough to satisfy the *Daubert* criteria. As yet, there is no reliable way to counter the Trojan defense.

It is troubling to think that individuals guilty of possessing child pornography could be set free. It is even more troubling to think of innocent people convicted of crimes they didn't commit; already there are unconfirmed reports of protection rackets where criminals demand money in return for not planting evidence on the victim's computer.

Some may object to the previous statement that there is no reliable way to determine guilt or innocence; certainly, qualified examiners can tell. To a point, this is true. A good examiner should be able to determine intent from any number of factors (e.g., Is there evidence of a Trojan having existed in the registry or on the hard drive or in the logs?). However, the examiner is not the final audience. For criminal proceedings, the findings of the examiner must be admissible and they must be communicated to a judge or jury in a way that can be understood (Casey, 2002; Smith & Bace, 2003; Sommer, 1997).

Currently there is no established standard method for conducting a computer forensic examination (Carrier & Spafford, 2003; Whitcomb, 2003). Each examiner must explain the sequence of actions they took, and what evidence if any, was found. They must explain why they chose this sequence rather than another, what affect another sequence would have had on the results, etc. (Carrier & Spafford, 2004; Smith & Bace, 2003). This often involves explaining concepts such as slack space, the continued presence of deleted files, and timestamps (Mandia, Prosise, & Peppe, 2003). These concepts are foreign to most outside the field. Furthermore, these are only the concepts that must be understood for a simple case, such as the presence of illicit material on a hard drive. More complex cases involving network attacks may require knowledge of network protocols and architecture (Mandia, Prosise, & Peppe, 2003) .

For this reason, expert testimony by those in computer forensics often leaves the jury confused (Smith & Bace, 2003). Juries are doubly confused if both sides have employed experts. In the case of the denial of service attack against the port of Houston, the teenager indicted was acquitted despite expert testimony that a Trojan likely never existed on his computer, and that the alteration of the logs in the way the defendant claimed was near impossible.[1]

Drug and DNA testing are equally complicated. Each drug has a specific test that is appropriate. There are multiple types of DNA testing, depending on what substance is being tested for DNA and how much of it is available (Connor, 2004). If a lab technician had to explain the role of carrier gases in a mass spectrometer to explain how illegal steroids were found in an athlete's blood convictions would be difficult to obtain. Similarly, if, in order to prove that an individual's blood was found at the scene of the crime, the role of polymerases in cutting DNA into manageable segments had to be explained, it would be difficult for a jury to trust the evidence without a degree in biology.

[1] http://zdnet.com.com/2100-1105_2-5092745.html

Yet, juries trust DNA and drug. These methods are standardized and peer-reviewed. The error rates are measurable. If two examiners are given the same test to run on the same sample, provided the equipment is working and the procedure is followed, the result will be the same within the error rate.

In order to be recognized as a mature scientific discipline, computer forensics must be able to meet the legal and scientific criteria. This is currently not the case (Whitcomb, 2003). This study is a first attempt at using statistics in order to provide an empirically based method for evaluating possible events that could account for the presence of evidence on a suspect's system.

Method

The hypothetical situation this study examined was a case in which five illicit images were found on an individual's computer. The research question was whether an investigator could determine if images were downloaded intentionally or without the owner's knowledge based on characteristics located in the operating and file system. Four possible scenarios were considered; in three of these scenarios, the suspect is innocent (no intentionality), in the fourth, the suspect is guilty (intentional behavior):

> **Scenario 1:** The user visits a website with popups that contain illicit images, but immediately closes the windows.
>
> **Scenario 2:** The user downloads and unzips an archive file that seems innocent, but contains illicit images.
>
> **Scenario 3**: An attacker is remotely controlling the user's computer using a program like BackOrifice or RealVNC. The attacker downloads the illicit images and saves them to the user's home directory.
>
> **Scenario 4:** The user visits the website containing illicit images and saves them to his home directory. The user then views his home directory and saves one of the images to a floppy disk. The image saved on the floppy disk is opened once from the disk.

Each scenario was acted out in three trials using a 10 GB master image of a Windows XP install. The master image also had WinZip and RealVNC installed and some background activity present. Web browsing of some innocuous sites such as news sites and Ebay was done to add some entries to the cache folders. A document was created with Word to simulate the normal use of the computer. This background activity was done to test the ability of the characteristics listed below to be useful even when the computer had been used for other activities.

Based on the literature reviewed, it was predicted that the following characteristics (variables) may help an investigator to determine whether images

located on a suspect system were the result of unintentional or intentional activities:

- Average of the difference between file creation times.
- Mode of the difference between file creation times.
- Median of the difference between file creation times.
- Number of references to contraband items stored on local disk in the Recent Folder.
- Number of references to contraband items saved to/opened from external devices in the Recent Folder.
- Number of thumbnails that exist for contraband images.
- Number of images created within five minutes of visit to contraband website.

The first characteristics chosen were based on the amount of time that passed between the creation of the files. Human response time is much slower than automated processes. Files downloaded by a Trojan - similar to scenario 2 - would be created rapidly. Files downloaded by a human agent - situations 3 and 4 - would be created much more slowly. Since it was unclear which aggregate measure of the difference between file creation times would be the best, the average, mode, and median were examined.

Another possible measure of intent is the number of references to the contraband images stored on the local disk in the Recent Folder for the user. If the user has opened a document and the Recent Folder has not been cleaned out, there will be a reference stored for that item. Also, the Recent Folder will contain references to any files recently saved to or opened from external devices, like floppy drives.

When a directory containing images is viewed and certain user settings are enabled, a thumbnail will be created for each image in the directory. If the user has downloaded the images unintentionally, it is less likely that he or she will have viewed the directory. In scenario 1, the images will be stored in a directory used for temporary Internet cache and in scenario 2, if the images are buried sufficiently, the user will not have viewed that directory.

The last measure chosen was the number of images created within five minutes of the last access to the site that possibly contained the contraband images. This measure is intended to distinguish between situations where the user has visited the site and where the user has not.

Each of the characteristics was measured across the trials (see Tables 1, 2, & 3).

Results

A stepwise discriminant analysis was performed using the identified characteristics as predictors of intentional behavior. The initial analysis identified only four variables as significant for distinguishing between the scenarios. The four variables identified were the average of the difference between file creation times (Average), median of the difference between file creation times (Median), number of thumbnails that exist for contraband images (Thumbnails), and number of images created within five minutes of visit to contraband website (Web) (see Table 4). However, given the limited number of trials used in this exploratory study, the maximum number of variables that could be combined into a predictive model without violating the assumptions of the test was two (Tabachnick & Fidell, 2001). The analysis was conducted again with that restriction, and the average of the difference between file creation times and the median of the difference between file creation times were identified as being significant (see Table 5). Using these variables, 100% of the cases were classified correctly and the cross-validated accuracy rate was 83.3%.

Two discriminant functions were calculated, with a combined $\chi 2$ (6) = 33.90, $p < .01$. After the removal of the first function, there was no longer a strong association between scenarios and predictors. The first discriminant function maximally separates intentional from non-intentional behavior (see Table 6).

Discussion

The findings indicate that it is possible to determine, with a given statistical significance and accuracy rate, which situation created images (evidence) located on a suspect system. The ability to determine the veracity of the defense's explanation for the existence of contraband images is extremely important. This determination can greatly assist a judge or jury in eliminating reasonable doubt regarding guilt or innocence (Smith & Bace, 2003). The use of well-known statistical methods (e.g., discriminant analysis, logistic regression etc.) provides an empirical foundation for determining the veracity of alternate explanations. The use of statistical analysis and a documented protocol is the first step toward allowing a computer forensics investigator to testify on the stand that there is a 99% probability events happened in a certain sequence.

Caution should be used when interpreting the results of this pilot study. Because this study was exploratory, a limited number of trials were conducted. The limited number of trials did not allow a full model to be tested, due to the assumptions of the discriminant analysis test (Tabachnick & Fidell, 2001). A larger number of trials would allow for the testing of models with more variables and possibly improve classification and significance levels. In addition, a larger number of trials may expose any irregularities in the measurements due to operating system software. It is difficult to say without extensive code review or testing what exactly

happens to an object as it is modified by the operating system. Large enough trials would reduce this uncertainty to acceptable levels.

Furthermore, the choice of characteristics to study is sensitive to particular situations and may depend on underlying software or hardware. If this method were to be used to classify hard drive images of multiple operating systems, even across different versions of Windows, extensive testing would have to be done to determine if the characteristics are the same across each operating system. By way of example, initially, a registry key was going to be used to determine whether the image file had been opened but the registry key that worked with Windows 2000 didn't exist in Windows XP.

These limitations are not uncommon in other forensic sciences. The situation in drug testing is similar; each drug has specific tests that are appropriate (Connor, 2004).

Applied Uses

The method proposed in this study could be extended to many different types of cases. Consider the hypothetical case of an individual accused of sending illegal, unsolicited email. The individual claims that a Trojan program existed on her computer and while it was her email address, she was not aware of what was happening. The first step is to build a number of alternate scenarios for how the Trojan could have been installed and operated without her knowledge. Add to this list of scenarios the sequence of events investigators believe would occur if she were guilty. Evidence used could include more than simply the hard drive image here; records from the ISP would be useful in establishing a timeline.

The second step is to have a large number of trials done for each scenario, being careful to recreate the environment in which the crime occurred accurately. Next, computer forensics experts would list some characteristics of the hard drive or network traffic records they would expect to be different between scenarios. There are registry keys that record how often a user accesses a certain program by clicking on it or by choosing it from the start menu, and the last time of access. If the key that records this shows that a spam program has been run a hundred times in the past three months, it may provide a clue for intent. File creation times could also be useful. If the action is occurring while she can provide alibis for being away from the computer, it may provide evidence for the defense. Any characteristic that can be measured numerically and might be relevant is a candidate.

In the third stage, a statistical method, such as discriminant analysis, is used to determine which characteristics are useful, what the level of significance is, and how often the model is correct. Discriminant analysis is a good fit for this problem, as in stepwise discriminant analysis one characteristic is added in each

step. The characteristic chosen is always the variable that is most useful (i.e., the one that lowers the significance the most) (Tabachnick & Fidell, 2001).

Once the discriminant model has been created, evidence gathered from the suspect's computer and ISP could be measured for the same characteristics and by using the discriminant functions as demonstrated earlier, classified with a known level of significance and accuracy.

Conclusion

This paper has presented the first attempt at a standardized method for event reconstruction which has measurable accuracy and significance. The ultimate goal is to enable an expert witness in court to testify that there is a 99% chance the illicit images were or were not placed there intentionally. Most established forensic sciences have standardized processes for determining the sequence of events. Studies have been done to determine how long a body will take to decompose under given conditions, at what angle and from what height a drop of blood fell, etc. There is much work left to be done for computer forensics to reach that point, but doing so would take computer forensics one step closer to being an established forensic science.

About the Authors

Megan Carney (mcarney@cerias.purdue.edu) graduated with a Master's degree in Computer Science from Purdue University in May 2004. While a graduate student, she spent a summer at NASA Office of Inspector General Computer Crimes Division at the Jet Propulsion Laboratory in Pasadena, CA. In addition to this research, she has publications in the area of intrusion detection systems. Her website is http://www.megancarney.com/geeks.

Marc Rogers Ph.D. CISSP, CCCI, (mkr@cerias.purdue.edu) is an Associate Professor in the Department of Computer Technology at Purdue University. Dr. Rogers is an ex-police detective, and has extensive experience in computer forensics. His website can be found at http://www.cerias.purdue.edu/homes/mkr/.

References

Carrier, B., & Spafford, E. (2003). Getting physical with digital forensics investigation. *International Journal of Digital Evidence* (Fall 2003).

Carrier, B., & Spafford, E. (2004, February 2004). *Digital crime scene event reconstruction.* Paper presented at the American Academy of Forensic Sciences 56th Annual Meeting, Dallas.

Casey, E. (2002). *Handbook of computer crime investigation : forensic tools and technology*. San Diego, Calif.: Academic Press.

Connor, T. (2004). Forensic Toxicology. Retrieved March 15, 2004, from http://faculty.ncwc.edu/toconnor/425/425lect14.htm.

Mandia, K., Prosise, C., & Peppe, M. (2003). *Incident response & computer forensics* (2nd ed.). Berkeley, Calif. ; London: Osborne.

Rogers, M. (2003). The role of criminal profiling in computer forensic investigations. *Computers and Security, 4*.

Smith, F., & Bace, R. (2003). *A guide to forensic testimony: The art and practice of presenting testimony as an expert technical witness*. Boston: MA: Addison Wesley.

Sommer, P. (1997). *Computer forensics: An introduction*. Retrieved June 3, 2003, from http://www.virtualcity.co.uk/vcaforensics.htm

Tabachnick, B. & Fidell, L. (2001). *Using multivariate statistics, 4th edition*. Needham Heights, MA: Allyn & Bacon

Whitcomb, C. (2002). A historical perspective of digital evidence: A forensic scientist's view. *International Journal of Digital Evidence* (Spring 2002).

Table 1. File Creation Times

Scenario	Trial	Average	Mode	Median
1	1	.25	0	0
1	2	.5	1	.5
1	3	0	0	0
2	1	.5	0	0
2	2	.5	0	0
2	3	.5	0	0
3	1	8.75	8	8.5
3	2	10.25	9	9
3	3	21.25	NA	17
4	1	5.25	6	5.5
4	2	5	5	5
4	3	4.5	5	4.5

Note: Time in seconds

Table 2. Number of References

Scenario	Trial	Local	Ext
1	1	0	0
1	2	0	0
1	3	0	0
2	1	0	0
2	2	0	0
2	3	0	0
3	1	5	0
3	2	5	0
3	3	5	0
4	1	4	1
4	2	4	1
4	3	4	1

Table 3. Number of Images

Scenario	Trial	Thumbnails	Web
1	1	0	5
1	2	0	5
1	3	0	2
2	1	0	0
2	2	0	0
2	3	0	0
3	1	5	5
3	2	4	5
3	3	4	5
4	1	5	5
4	2	5	5
4	3	5	5

Table 4. Stepwise Discriminant Analysis

	Wilks' Lambda	F	df1	df2
Average	.011***	206.05	3	7
Median	.006***	380.83	3	7
Thumbnails	.008***	294.64	3	7
Web	.110***	18.88	3	7

*** $p < .001$

Table 5. 2 Factor Discriminant Model Test

	Wilks' Lambda	F	df1	df2
Average	.216***	9.97	3	8
Median	.149***	15.22	3	8

*** p < .001

Table 6. Discriminant Functions

Test of Functions	Wilks' Lambda	Chi-square	df
1 through 2	.01***	33.90	6
2	.63	3.68	2

*** p < .001