The Ultimate Tool For Fighting Crime

Mefail Tahiri
Docent at State University of Tetova

Ejup Rustemi
Master at State University of Tetova

Abstract

Forensics involves the investigation of evidence by following scientific methods within the regulations of the law. Digital forensics applies these principles to the process of digital evidence recovery. It is a sub-discipline of forensics that can be traced back more than two decades.

***Key words**: Forensics, , cybercrime, digital forensics, security.*

Fraudulent activities account for billions of dollars lost in the insurance, banking, healthcare, retail, transportation, manufacturing, and communications industries each year. Likewise, fraudulent activity riddles our federal and local governments; virtually every industry is vulnerable to fraud.

Flexibility remains a critical aspect for quickly responding to changing fraud patterns. It is crucial to dynamically expose new patterns of fraud without having to reprogram, retrain, or reinvent the underlying systems. Most important is to expose the fraud before it impacts the operations or business foundations. Keep in mind that before patterns are classified they first have to be discovered.

There have been a multitude of new technologies introduced into the antifraud marketplace over the past several years, including link analysis and other systems for detecting non-obvious relationships and associations. Perhaps even more important are the refined analytical methodologies that help to interpret the complex networks and patterns presented by these technologies. Better understanding of the data will

inevitably lead to better pattern detection, and ultimately, lower fraud incidence. Once a pattern has been exposed, it is up to the affected company to act on that knowledge by changing business processes to flag related or similar occurrences of the pattern. Remember that there are always exceptions to the rule, and there are exceptions to the exceptions. (*Data mining for intelligence, fraud and criminal detection; Christopher Westphal; 2009*)

Digital forensics, in essence, answers the when, what, who, where, how and why concerning a digital crime (Beebe and Clark, 2004). When conducting an investigation on a computer system, for example, the 'when' refers to the time interval the activities took place during. The 'what' concerns the activities performed on the computer system. The 'who' concerns the person responsible, the 'where' refers to where the evidence is located, the 'how' addresses the manner in which the activities were performed, and the 'why' seeks to ascertain the motives behind the crime.

Today, an increasing amount of criminal evidence resides on a computer, even if the majority of such evidence is still used to commit traditional or conventional crimes. An example would be a threatening letter sent in electronic format, such as an e-mail or text file, instead of a traditional paper format.

Digital evidence by definition is information of probative value stored or transmitted in digital form (Pollitt, 2001). It is somewhat unique when compared to other forms of documentary evidence. For instance, it may be found in unusual locations, ones that are unknown to general computer users. It is also fragile in nature and can easily be altered or destroyed.

Network traffic, for example, is a source of digital evidence that presents numerous challenges (Casey, 2004b). This is due to the limited opportunity for capturing network traffic. Adequate data capturing systems must be in place as data travels through a network otherwise the opportunity is lost. Furthermore, when dealing with network traffic, it is often difficult to locate and extract specific items from the large number of flows on a network.

Sources of digital evidence include (but are not limited to):

• compact discs (CDs);

• computer systems;

• digital cameras;

• digital media devices such as the iPod

• digital versatile discs (DVDs);

• flash drives;

• floppy disks;

• game consoles such as the Xbox 360 (Microsoft Corporation, 2006);

• memory cards;

• mobile phones;

• network devices such as routers and switches;

• network traffic;

• notebooks, and others.

Forensics involves the investigation of evidence by following scientific methods within the regulations of the law (Vacca, 2002). Digital forensics applies these principles to the process of digital evidence recovery. It is a sub-discipline of forensics that can be traced back more than two decades (Inman and Rudin, 2001). It can be classified into two key areas, namely, computer and network forensics. In general, computer forensics deals with data in a computer (Reith et al., 2002), whereas network forensics deals with data that may be spread over several databases residing on computers in one or more networks. (*Data visualisation in digital forensics disertation; 2007*)

More often, the investigation will comprise a detailed examination of the crime scene, the collection of forensic material and recording of witness and victim statements. This level of process is not just reserved for serious crimes, but is routine practice even for relatively minor ones.

It is apparent that regardless of the level of investigation, a single recorded crime will generate a considerable amount and diversity of information. The challenge is not only to store this information, but to use it to facilitate the investigative process. The features of one case at a particular point in time may have little value but as we shall see below, the ability to retrospectively interrogate such a comprehensive crime database is a powerful investigative tool.

KDD

Knowledge discovery from databases (KDD) is the non trivial extraction of implicit, previously unknown and potentially useful information from data (Fayyad and Stolorz, 1997; Fayyad, 1996). The KDD process begins with analysis of data stored in a database or data warehouse

and ends with production of new knowledge. Fayyad (1996) describe knowledge discovery as a process with five distinct stages: data selection, data pre-processing, data transformation, data mining and interpretation.

The first phase of any KDD process involves the selection of a sample of data from a database of records. Decisions must first be made regarding the nature of the problem of interest in order to assess its suitability for the KDD process. This phase is equivalent to sampling in statistical circles and involves selecting which records to include and which to omit. There are two distinct considerations; how to select records and how to select variables.

Data may need to be transformed in order to discover useful knowledge. Transformation can involve changing the categories of values a variable may have. It can take one of three basic forms:

a) the decomposition of the data set into smaller parts where each part will be the subject of an independent data mining exercise,

b) the aggregation of variables and/or values to form a simpler, more general data set,

c) transforming values of variables in some way.

Following the acts of terrorism of September 11 2001, there has been an international focus upon the avoidance, detection and prosecution of acts of terrorism. Intelligence authorities have been very keen to embrace the use of data mining to help prevent acts of terrorism occurring.

The problem with data mining in this and other criminal domains is that information is often stored as free text and in unrelated datasets. For example prior to September 11 2001, very few people would have taken much notice of students who wished to learn to fly airplanes but showed no interest in landing the planes. This issue only became important once it had become clear that the perpetrators of the acts of terrorism of September 11, 2001 had no interest in landing planes.

Over the past few years, Computer Forensics has been supported by several theories and methods developed in order to find evidence quickly on seized computers as far as specific crimes such as murder, child abductions, missing persons, death threats etc. are concerned. In such cases the need for the timely identification, analysis and interpretation of digital evidence is

crucial since it could be the difference between life and death for the victim.

The forensic investigation of digital evidence is predominantly employed as a post-incident response to an activity that cannot be defined definitely as legal or to an incident that does not comply to the organizational norms and policies. While the presence of physical forensic investigation model has matured through the years of its presence, refined as revised globally, the involvement of digital evidences have made its presence felt in the recent years. In 1995, M Pollitt, suggested a four step process that mapped admission of documentary evidence in the court of law to admission of digital evidence, giving a concrete base for dealing with potential digital evidence; The process steps included were acquisition, identification, evaluation and admission.

However in 2001, DFRW (Digital Forensic Research Workshop) came up with a framework which involves identification, preservation, collection, examination, analysis, presentation and decision. This framework is the basis for all the proposed models that followed till date.

Stage 1: Preparation: The main focus is acknowledging the role of digital storage device(s) in the identified or untoward incident. This step recognizes the presence or absence of the digital forensic investigation. All suspected physical storage devices are to be physically secured to prevent tampering. The concerned authorities are to be notified about the presence of possible evidence(s) and the need for examination of the same, and hence permission to access the device. In case the evidence needs to be removed from the premises or site of the activity, steps for obtaining the necessary permissions for the removal are to be identified and executed. On the whole, based on the nature of the incident or crime, the investigation steps are to be chalked out.

Stage 2: Collection and preservation of digital device: The device collection phase opens with the identification of the ownership of the device along with the identification of supposed users of the device. All the digital devices and any other supporting evidences about the usage of these devices, that are present at the scene of crime are to be confiscated for data collection. In case the physical device is password protected, the software necessary for accessing the device contents is identified and verifying that it does maintain the integrity of the data as it works on accessing the device. The device contents

should be duplicated or imaged maintaining the integrity of the data in the device. Each step of the activity should be documented.

Stage 3: Data extraction and preprocessing: The device/disk that has been imaged or duplicated is to be accessed and examined for the presence of any hidden or encrypted data and system related data. Required software tools are to be used to decrypt or access the data. These tools should not tamper the original data. Ensure that nothing will/shall be written on to the device that is under scrutiny.

Based on the nature of the incident, the investigation is to be categorized as goal based or non-goal based. The data should be extracted from the digital device and the steps for the preprocessing the data is to be outlined, justifying the reason for the same. The software required for the process is to be identified. All through the stages, concern about maintaining the integrity of data should be the key focus and each step is to be validated before executing. Documentation of the activities carried out should be precise and justified as this would act as the base document for justifying the integrity of the presence or absence of evidence leading to the crime.

Stage 4: Data examination and analyses: Before the data is subjected to examination and analyses, the data is to be cross checked for authentication and integrity. The analyses that can be carried out on the extracted data, based on the nature of the data, are to be considered along with the required tools to perform the same. On justifying the analysis methodology, the actual analysis is to be carried out until stable results are achieved. Interpretation of data is the most difficult step, while at the same time the most important step in this flow.

Stage 5: Reporting and documentation: Though this has been cited as the stage 5, it is a continuous process, which needs to be reviewed and updated finally, before presentation in the court of law, for completeness and accuracy. The validity and the acceptance of the process or methodology in the scientific community should also be explored. Documentation of the analyses, conclusions and assumptions if any, are also of importance. The limitations of the procedures/analysis carried out are to be outlined clearly.

Stage 6: Presentation in the court of law: The main focus of this step is to prove the presence or absence of digital evidence,

from the digital devices collected from the scene of the incident under examination, in the court of law. While computer forensics is highly technology specific, people handling law in the court of justice are not technology specialists. Hence it is very important for technology specialists to understand the ramifications of the legal world and at the same time, communicate effectively and clearly the complete digital investigation process, emphasizing on the analysis of the findings. The documentation of the entire process may also be submitted in the court of law to cross-examine the steps adopted during the investigation process. While this may suffice the needs of the court to arrive at a decision, it may sometimes be required to complete further analysis or redo a phase, as required by the court, to support any issues.

Digital Forensics Tools

In the Literature, Computer Forensics Tools are basically classified as:

1. Hardware forensic tools, and

2. Software forensic tools

Hardware forensic tools can be used for Single-purpose components or complete computer systems and servers.

Software forensic tools are used for Command-line applications and GUI applications. The development of a variety of specialist commercial and freeware tools began in 1980s and 90s. These can generally be broken down into three categories as follows.

General forensic tools: Tools allowing a wide variety of investigation, particularly keyword searching, on digital media.

Specialist forensic tools: Which focus on a specific piece of forensic material for investigation perhaps images, or internet artefacts. Often relying on output from one of the general tools.

Case Management tools: These are used to track, audit and report on cases.

The unique need of computer forensics have resulted in the creation of computer forensic tools in the form of computer software. These tools ensure that digital evidence is acquired and preserved properly to maintain the integrity of digital evidence. For example, copying and pasting data onto another storage medium may not be admitted in a court of law as forensically sound evidence. This is because the process of copying and pasting data can modify it, for example, altering the timestamps of the data. As a result, a typical digital

investigation requires the making of an exact bit by bit (or bit stream) copy of all the data on a storage medium. This exact bit by bit copy is called an image and the process of making an image is frequently referred to as imaging.

Computer forensic tools focus primarily on digital evidence recovery, in other words, on recovering residual data from a piece of media. These tools usually have limited abilities to assist in the analysis of the recovered data. The presentation of data offered by computer forensic tools is deceptive at times. The reason is that the dimensionality, complexity and volume of data still exist because the computer forensic tools merely present it to investigators. The digital investigators still have to examine the presented data and draw conclusions.

At present, computer forensic tools are not ideal for the following tasks:

- Association: identifying correlations among data.
- Classification: discovering and sorting data into groups based on similarities of data.
- Clustering: finding and visually presenting groups of facts previously unknown or left unnoticed;
- Forecasting: discovering patterns and data that may lead to reasonable predictions.

Main Forensics Techniques

Imaging

One of the first techniques used in a digital forensics investigation is to image, or copy, the media to be examined. Though this seems to be a straightforward step at first, modern Operating Systems (OSs) perform many operations on file systems when connected, such as indexing or journal resolution. Without care, media can be modified, however slightly, and the integrity of the evidence can be compromised.

Hashing

To quickly identify a file and to provide authenticity that an image or file was not modified, the forensic community adopted cryptographic hashing. Modern hashing functions use one way Cryptographic functions to obtain a hash. The uniqueness of the hash depends on the cryptographic function used. MD5 hashing was developed in 1991 by Ron Rivest and was rapidly adopted by the forensics community.

Carving

One category of tools in the digital forensic toolkit is called file carvers. These tools allow the Scanning of disk blocks that don't belong to current files to find deleted data. Carvers use known header and footer signatures to combine these 'unused' nodes into the original files that were deleted. Carving can recover deleted but not overwritten files as well as temporarily cached files on media. An analysis of carving techniques was performed by Mikus in 2005. Recent advances in carving allowing fragmented files to be recovered with more accuracy.

Data mining & soft Computing has several applications in digital forensics. These include identifying correlations in forensic data (association), discovering and sorting forensic data into groups based on similarity (classification), locating groups of latent facts (clustering), and discovering patterns in data that may lead to useful predictions (forecasting).While this technique is ideal for association, classification, clustering and forecasting, it is also particularly useful for visualization.

Importance of Data Mining

Digital Forensic Techniques	Data Mining Techniques	Tool
Data Recovery, data generation and preprocessing	Statistical Test Analysis Bartlett's test of sphericity Kaiser-Meyer-Olkin (KMO)	Recuva FTK Encase Sleuth kit/Autopsy ProDiscover
Data analysis	Clustering – K-means, EM, Hierarchical Clustering	Weka
	Classification – Supervised learning - Decision Tree, Neural Networks, SVM, Naïve Baiyesian	Weka
	Unsupervised learning – PCA, Karnohuen Map	-
	Frequent Pattern Mining/Association rule Mining - Apriori, Eclat	Weka
	Named Entity recognition	LingPipe
	Visualization	CyberForensicTimeLab
	Statistical Analysis and Anamoly Detection	EMT/MET

	Recursive data mining	-
	Phishing	Invisible Witness
	Regression	-

Conclusion

The emergence of tools designed specifically for public safety and security analysis is an exciting trend that we hope continues. While data mining and predictive analytics can be very intuitive, the ability to match an analyst's style of investigation and query, is somewhat limited in their usage in some directly applied setting.

However we cannot in any way underestimate the help given by these tools. As mobile devices, such as smartphones, tablets or phablets are selling in outstanding numbers, crime, and especially, cybersecurity is something very serious, and finding ways to deal with that is the main goal of computer forensics.

References:

- *Data mining for intelligence, fraud and criminal detection; Christopher Westphal; 2009*
- *Data visualisation in digital forensics disertation; B. K. L. Fei; 2007*
- *Data Mining based Crime-Dependent Triage in Digital ForensicsAnalysis; R. Bertè, F. Marturana, G. Me, S. Tacconi; 2012*
- *In The Trenches: Computer Forensics and Data Mining; John Mallery (PowerPoint Presentation)*
- *https://www.piriform.com/recuva*
- *http://www.cs.waikato.ac.nz/ml/weka/*