Review of Evidence Analysis and Reporting Phases in Digital Forensics Process

Prof. Dr. Asaf VAROL
Fırat University, Faculty of Technology
Elazığ, Türkiye
varol.asaf@gmail.com

Yeşim ÜLGEN SÖNMEZ
Fırat University, Faculty of Technology
Elazığ, Türkiye
yesimulgen123@gmail.com

Abstract— This paper reviews digital forensic phases and problems in evidence analysis phase and smart methods in this area. Among these phases, studies on the evidence analysis phase are examined. In the analysis of electronic evidences, use of smart methods and their development will contribute to information technology law and improvement of digital forensic devices.

Effective evidence analysis both provides easiness for digital forensic experts and helps jurists for accurate decisions. In this paper, digital forensic process and smart methods used in evidence analysis are examined. This literature survey discussed which new methods can be added to this process.

Keywords— Digital forensics phases, evidence analysis phase, smart methods.

I. INTRODUCTION

The fundamental purpose of digital forensics can be described as discovery, protection, collection, analyzing and presenting legal electronic evidences that are seen as potential evidences [1, 2]. Digital forensics aim to find digital evidence for a number of cases ranging from identification of a hacker resolution of a murder [3].

In digital forensics, the purpose is not to point out a person as guilty or innocent. It aims to present numerical evidences to forensic units in other form as complete and impartial interpretation of the evidence. Determination of whether a person is guilty or not is determined by judicial authorities as a result of the transfer of the evidence obtained through digital judicial processes to judicial units [4].

Some fields in digital forensics can be listed as data recovery, data annihilation, data conversion, encryption, decryption, finding under cover files, identifying criminals with the help of IP numbers [5].

II. DIGITAL FORENSICS PROCESS

Process of reaching legal electronic evidence from electronic evidence is called “digital forensic phases” [6]. Digital forensic phases are listed in Figure 1 [3, 4, 6, 7]; these phases are describing evidence which start with the crime scene investigation then collection of evidence, protection of evidence, analyzing evidence, reporting and presenting the evidence.

There is a starting point for every process [8]. The process starts with an alarm from the attack determination system, suspicious records on the firewall, warnings from the security system on the network, denunciation of an individual or denunciation of any crime cases [8, 9].

The purpose of value evaluation is to determine whether there will be a detailed investigation process or not [8, 9]. Later, procedures and protocols which will be applied in the crime scene are identified. People who responsible for the security of the crime scene are first responders and digital forensics specialists. Their trainings needed in this subject depend on protocols identifying the crime scene (such as video and photograph) beforehand [8, 9]. Later, protection and collection of data phases start. The Figure 2 shows these phases.

A. Identification and Collection of Electronic Evidences

In this phase, the purpose for experienced researchers is not to collect all virtual or physical evidences. They must decide what needs to be collected. Then they must create a document and finally perform the action [8].
B. Protection of Electronic Evidence

Within the scope of protection of evidence, it is required to denote where and in which conditions the evidences are collected in the crime scene. In other words, it is required to know the originality of collected evidences [6, 10]. Sending electronic evidences to the laboratory for investigation can be actualized through the conscious work of the police force in the identification and protection phases [11]. In Figure 3, the processes in the phase of protection of evidence are shown.

C. Recovery

Before starting complete analysis of the conserved digital evidences, it is necessary to discover deleted, hidden, transfigured data. It is necessary to discover data that is non-displayable with current operating system or file system, too. This is called data recovery. This process is not done on original evidences, it is done on their exact copies [8].

D. Decomposition

The purpose is to bring the data according to their specific characteristics in order to provide easiness for the research. For instance, since the child pornography cases [12], are usually based on visual digital data, files with the extension of gif, jpeg etc. are often brought together [8].

E. Reduction

Among the data collected, those that are directly related to the subject is selected. The selection criteria are carefully determined depending on whether the court can question the data collected. [8].

F. Organization and Research

It is necessary to organize group, to label reduced data and to place them meaningful units. The purpose is to ensure that researchers find and describe the data during the analysis and give reference to them in a meaningful way during the testimonial. A data index is created [8].
to make search in txt files, there are many programs (Powergrep etc.) available for these searches. If the evidence is being sought for crimes like pornography, image or video files should be searched as well. This process shortens the time.

Apart from this, it is possible to hide data in some image files via steganography method. By taking in consideration the qualification of the alleged crime and computer skills, use of devices that makes such controls can help to reveal evidences hidden in image.

BackTrack 5, code name “Revolution”, that is used at phases of protection and analysis of evidence is the most popular penetration test platform and released in 2011 [17]. Backtrack 5 devices which are named after “Backtracking” search algorithm, have a wide spectrum like having 12 categories ranging from password breakers to advanced penetration test devices and port scanners. Digital forensics investigations are conducted with the help of Forensics category in BackTrack 5 [17].

IV. REPORTING AND PRESENTING ELECTRONIC EVIDENCES

All processes applied to the evidence during the investigation and examination process should be indicated in the report, how the evidence is collected, what processes are used to make the copies, the devices, the operating system and the software used. These issues have critical importance for the report to be prepared. Reporting phase includes both technical and legal evaluation [6, 18].

There are two important issues that requires attention in digital forensics reports [9]. The first of them is to demonstrate that evidence integrity is maintained during the investigation process, the second one is to show that operations conducted are clear, transparent and repeatable, putting aside the exceptional situations [19].

After an investigation, expert reaches to investigation results with his/her opinion, proof or knowledge. Thus, it is necessary to explain which one is used [19].

Digital forensic experts should be able to convey their work on electronic evidences to the ones who have detailed information on the subject [19]. Hence, it is necessary to keep in mind this issue while preparing reports [6, 20]. Moreover, it would be beneficial to write the occupation of the person who prepared the report [21].

Due to the “direct evidence principle” in criminal procedure [22]”, evidences are considered collectively [22]. In other words, the court takes the case directly and makes a decision, adding witnesses, suspicions and other evidence [23].

V. SMART METHODS USED DURING THE EVIDENCE ANALYSIS

The Neural Networks [24], is a method used in digital forensics. For instance, hidden data in the analysis of audio files can be identified with smart bounce methods. There are studies which uses PNN (Probabilistic Neural Networks) technique [25].

Bayesian Networks [26] is used to work more efficiently on suspected evidences and eliminate possibilities in results. There is evidence based reasoning studies using open source language engineering software such as GATE and Bayesian Networks [27]. Bayesian Networks have studies based on three-layer structure. Figure 5 shows reaching to a conclusion through identifying the reliability of evidences through Bayesian Network [28].

![Figure 5. Hierarchical framework](image)

SOM (Self Organizing Map) Neural Network is used to visualize and group the data [29]. Figure 6 shows a classification with SOM and evidence chain.

SOM is used for subjects like identifying attacks, biometric systems and wireless network security. SOM, which is used by police during 2000s for sexual assaults, is also used in digital forensics. The data produced by digital forensics devices can be interpreted and analyzed with SOM. For instance, in some studies, two dimensional maps are produced by SOM based on graphical images, file names, extension, data of creation and time and electronic evidences are obtained accurately and reliably [30].
Use of multiple agents is a new paradigm for developing new smart agents’ software applications. Smart programs are the ones that map what they perceive on behaviors. They decide on behaviors by taking the information [31].

Smart agents that become special with the expert information are used in technical field. The purpose of smart agents is to automate the analysis and correlation of evidences and present the most important and relevant evidence by reducing the amount of big data. Correlation features help to find correlation between evidences that are overlooked by the expert person. In a big amount of data, it becomes harder to analyze the evidence and correlation. Digital forensics devices that were developed fall short as well and these devices do not have the ability of distributed processing. Hence, experts spend too much time [32].

For instance, hash sets are examined in a child abuse case [12]. First, smart software agent uses hash sets related to child abuse, so that it reduces the evidence range of the user. At the same time, Figure 7 shows the six special smart agent used in the system. These are:

- **HashSetAgents**, calculate MD5 hash value and compares them with the database that includes hash values obtained from different software related to more than 10 million child abuse cases.
- **FilePathAgents**, keep many file databases used in P2P (peer-to-peer) [12], VoIP and instant messaging investigations.
- **FileSignatureAgent**, investigates file headers (the first 8 byte values of the file), matches them with file extensions, identifies whether any of them keeps extensions or not. It keeps prefixes and names of the files as a list.
- **TimeScheduleAgent**, investigates access or change dates of incident happenings, software and system installments. It investigates restores and web browser usage.
- **WindowsRecordAgent**, shows important information extracted from files related to Windows records, system install date, configuration time zone and portable media.
- **KeyWordAgent**, searches for key words to be obtained from credit card numbers, URLs or email addresses and uses regular expressions [32].

Academic studies use metadata for big data, semantic web ontologies for different data sources, artificial intelligent techniques for data description [33, 34].
The first two phases of evidence capturing analysis are done through a series of hard drive and software devices. These devices are continuously developing in line with the technology and changes in devices.

Among digital forensics phases, the data analysis phase is support less. There are less software devices for this phase. There are less software devices available for text analysis and mining methods. Even though smart methods are strengthen for temporal reasoning for causal reasoning for utilization. These methods are utilized in smart methods for network operations. Among digital forensics phases, the data analysis phase is supported less. There are less software devices for this phase. Additionally, analysts reach to the insufficient training through appropriate devices.

For this reason, legal and verifiable smart methods and automatic reasoning techniques for the analysis of evidences reveal evidence sources. Even though smart methods are applied with different purposes in the field of digital forensics, these techniques are mainly related to the pattern recognition system, which identifies discrepancies in big amount of data such as image and email exchanges in multiple-media files and network operations.

Practical smart methods that might be beneficial should be utilized. These methods are response series programming and causal reasoning for possible scenarios in evidence analysis, temporal reasoning for time and time constraints, fuzzy logic for uncertainty in collecting data and neural networks for strengthening logical techniques with a strong classification.

In evidence analysis phase, new smart and/or statistical methods/models can be produced. This phase is very important for law and needs innovations.

REFERENCES


