A Log File Digital Forensic Model

Himal Lalla, Stephen Flowerday, Tendai Sanyamahwe, Paul Tarwireyi

Chapter 17

A LOG FILE DIGITAL FORENSIC MODEL

Himal Lalla, Stephen Flowerday, Tendai Sanyamahwe and Paul Tarwireyi

Abstract This paper describes a digital forensic model for investigating computer networks, focusing specifically on network log mining. A thorough examination of log files is needed to reveal the hidden actions of criminals in computer networks. The proposed model specifies the steps that forensic investigators can follow with regard to the extraction and examination of digital evidence from log files for use in legal proceedings.

Keywords: Digital forensic model, network forensics, log files

1. Introduction

Network forensics is an area of digital forensics where evidence is gathered and analyzed from network devices such as firewalls, switches, routers and network management systems. Cyber criminals typically use computers connected to the Internet to penetrate computer networks. The attack traffic has to pass through various network devices because the computers used to launch the attack and the targeted system are usually on different networks [8]. These network devices can provide investigators with digital footprints that could reveal details of the techniques and the identities of the individuals and machines responsible an attack [19]. If the network devices are configured properly, each attack action could leave digital footprints in the log files of the devices [8]. Log files provide significant evidence because they record all the activities that take place in an organization's computer network [19, 26]. After tracking down the machines used in an attack, investigators must correlate the logs found on the attack computers and those residing in the victim network [11]. Log file forensics involves all the procedures

involved in investigating the actions undertaken by the perpetrators of an attack.

The task of following digital footprints and mining digital evidence from network log files must involve well-defined forensic processes [21]. Network log mining is the process of discovering, extracting knowledge, and modeling and analyzing events recorded in the log files [12]. This paper describes a digital forensic model for computer network log mining. The model specifies the steps that forensic investigators should follow with regard to the extraction and examination of digital evidence from log files for use in legal proceedings.

2. Network Log Mining

A typical organization has many network devices that, if configured correctly, can generate and store log files of user activities [8]. In order to make sense of all the data provided in continuous streams by network devices, dedicated logging infrastructures (log file servers) have been developed to support the storage and management of logs [8]. Some of the techniques include console logging, buffered logging, terminal logging, syslog, Simple Network Management Protocol (SNMP) traps and the Authentication, Authorization and Accounting (AAA) protocol [29].

When users access an organization's website over the Internet, the log files of network devices record considerable data pertaining to user activities [10]. Each line of a log file typically lists the IP address, date and time of the access (timestamps), accessed object and referenced object [19]. Log files are an important source of digital forensic evidence because they usually connect events to points in time [26]. Indeed, log file data can be used to investigate network anomalies due to insider threats, data leaks and misuse of IT assets [12].

Log files can help identify network intruders [27]. In addition, they capture the behavioral patterns of users as they interact with computer networks, providing investigators with valuable insights into the *modus operandi* of cyber criminals [10]. However, comprehensive and well-defined log file forensic procedures are required to extract and analyze this evidence for use in legal proceedings [27].

The variety of digital forensic models that have been developed around the world demonstrate the complexity of evidence collection and analysis in digital forensic investigations [1, 21]. Prominent models include those developed by the Digital Forensic Research Workshop [20], Reith, Carr and Gunsch [23], Carrier and Spafford [7], Baryamureeba and Tushabe [3], Ieong [15] and Perumal [21]. However, what is missing is a model that focuses on the processes involved in network log forensics. The

available forensic models are generic in nature, leading to ambiguity about the specific procedures to be followed in log mining. Ultimately, this ambiguity can impact the admissibility of evidence presented in a court of law.

3. Proposed Model

The proposed model is shown in Figure 1. It has four basic layers, with laws and regulations serving as the base of the model (Figure 1). The four layers are: (i) preparation layer; (ii) discovery layer; (iii) testing layer; and (iv) elucidation layer. Each layer has unique processes that are supposed to be completed before the initiation of the next layer. The model has a top-down layout, which makes it simple and easy to follow. Processes within each layer follow the directions of the pointing arrows. Note that the precautions to be followed in forensic investigations are also part of the model.

3.1 Preparation Layer

Thorough preparation is essential in a technical investigation of network log files. Two processes are involved in the preparation layer: (i) approach formulation; and (ii) pre-incident preparation.

3.1.1 Approach Formulation. The organization must be aware of the need to conduct an investigation [9] and must ensure that the operations and infrastructure can sustain the investigation [3]. In order for the organization to be aware of the need, there must be a trigger, typically an event resulting from an illegal action. Law enforcement should be notified about the breach and all concerned parties must be informed about the possibility of a forensic investigation [22]. The legal and technical issues must be considered, along with the impact on business operations [24].

The main goal of the investigation is to recover admissible evidence from the log files without interrupting business operations. The associated forensic readiness plan has three goals: (i) recovering admissible evidence without interrupting business operations; (ii) keeping the cost of the investigation proportionate to the incident; and (iii) ensuring that the evidence has a positive impact on the outcome of the legal action. These goals serve to clearly define the relationships with the events and the impact on the other steps [25]. The approach formulation process must be robust to ensure the success of the investigation [23].

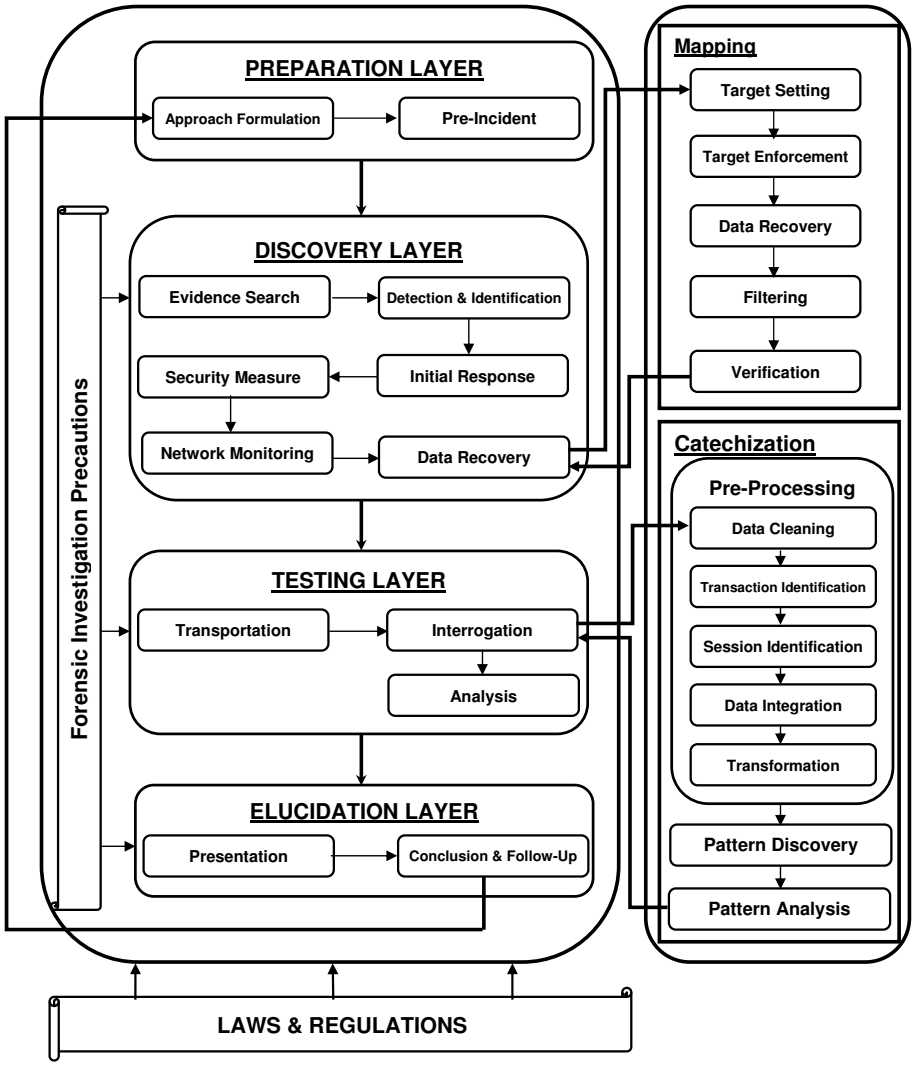

Figure 1. Log file digital forensic model.

3.1.2 Pre-Incident Preparation. Without proper preparation, the investigation may not be conducted in a systematic manner, leading to errors that could render the evidence worthless [9]. During the pre-incident preparation process, an initial understanding should be developed about the nature of the crime and the activities to be performed. The activities include building an appropriate team, assigning duties to team members, accumulating the materials for packaging evidence sources and retrieved log file entries, legal coordination and

general monitoring approval [7]. A thorough pre-incident preparation process leads to high quality evidence and contributes to the success of the investigation [22].

3.2 Discovery Layer

The second layer, called the discovery layer, follows the preparation layer (Figure 1). In this layer, tests and experiments are carried out on the network devices to identify the digital footprints of the suspects [8]. The goal is to discover a link between the suspects and the incident. The processes associated with this layer include evidence search, detection and identification, initial response, security measure implementation, network monitoring and data recovery.

3.2.1 Evidence Search. The systematic search for evidence is a process where the investigator surveys the physical and virtual crime scenes, deciding on the best methods for collecting evidence [23]. Evidence search involves the evaluation of the crime scene, formulating the relevant search plan and searching potential sources of evidence [9]. The investigator must identify key pieces of evidence that relate to the case [7]. Electronic equipment at the scene must be evaluated to determine if any expert assistance is required, individuals at the scene should be identified and preliminary interviews conducted [22]. Systems administrators must be interviewed for details about the system, applications, users and security measures [6]. At this point, if it is necessary to search items that are not listed on the warrant, then appropriate changes must be documented and a new warrant may have to be obtained; failure to do this correctly can result in evidence being deemed inadmissible by the court [24]. The evidence that is collected must be documented and a chain of custody must be established.

3.2.2 Detection and Identification. Before rushing to examine the log files of network devices, the forensic investigator must confirm the claim of intrusion into the organization's computer network using specialized techniques and tools [11]. Signature-based intrusion detection systems and/or anomaly-based intrusion detection systems can be used for this task. Signature-based systems rely on pattern-matching techniques; they contain a database of signatures of known attacks and attempt to match the signatures with the available data. Anomaly-based detection systems build a statistical model describing normal network activities; behavior that significantly deviates from the model is considered to be suspicious. After the detection and identification process

is completed, a warrant pertaining to the detected incident has to be obtained by law enforcement before the investigation can proceed [4].

3.2.3 Initial Response. During the initial response process, the investigator must brief the leadership of the organization about the results of the detection and identification process [7]. If the intrusion claim is false, then the investigation ends at this point. If the claim is true, then the initial response is to secure the incident site, which involves isolating the target computer from the network and maintaining the integrity of system log files [7].

3.2.4 Security Measure Implementation. The security measure implementation process is required when the investigator is a member of the victim organization's internal incident response team. The investigator must explain the vulnerabilities that were exploited and how network security can be improved [14]. This must be done right after the initial response so that systems administrators are aware of the vulnerabilities at the start of the forensic investigation [23].

3.2.5 Network Monitoring. The network monitoring process involves the monitoring and control of traffic in the organization's computer network [7]. The incident site must be secured to maintain the integrity of log files [7]. Also, the rate at which log files in the network devices are updated by new transactions should be reduced [16]. The log files of an organization can record approximately 240 million entries per day [8]; therefore, to capture proper metrics related to intrusions and disruptions, network traffic has to be restricted [19]. This reduces the laborious effort of searching for relevant entries among millions of log file entries during the data recovery process [19].

3.2.6 Data Recovery. During the data recovery process, investigators have to rummage for the digital footprints of cyber criminals in the log files. The investigators should focus on the log files of network devices as well as log file servers, if present [10]. This process, which is called mapping, comprises five sub-processes: (i) target setting; (ii) target enforcement; (iii) data recovery; (iv) filtering; and (v) verification.

Setting targets in network log mining requires investigators to document what has to be accomplished during the process of retrieving evidence [11]. In addition, all the steps must be numbered so that, if there is a change in personnel, the tasks will still be carried out in the correct sequence [11].

After setting the targets, the investigator must enforce the targets. This requires the investigator to ensure that all the documented aspects of the targets are achievable [11]. Having done this, the investigator must locate the batches of relevant log entries that can provide significant evidence about the crime.

The process of searching for batches of relevant log entries and the copying or imaging of the batches is called data recovery [27]. After the relevant log file entries are copied to their respective directories, the directories can be filtered into safe repositories such as digital evidence bags [26]. Finally, verification is conducted, which involves rechecking the log files to ensure that all the relevant log entries have been copied [12].

3.3 Testing Layer

The testing layer follows the discovery layer (Figure 1). It comprises three processes: (i) transportation and storage; (ii) interrogation; and (iii) digital evidence analysis. All three processes are typically conducted in a forensic laboratory.

3.3.1 Transportation and Storage. After the investigator is satisfied that the mapping process has been conducted in a comprehensive manner, the filtered evidence and seized devices are transported to a forensic laboratory for safe keeping and further analysis [3]. This step ensures the integrity of the evidence and reduces the risk of evidence tampering [7]. Proper safety measures must be maintained because the evidence can be destroyed while in transit due to shock, excessive pressure, humidity or temperature. The seized devices should be placed in anti-static bags to avoid damage or loss of evidence due to static electricity discharges. All the evidence should be stored in a climate-controlled environment with little or no electromagnetic radiation, dust, heat and moisture [22].

3.3.2 Interrogation. The interrogation process involves the examination of the retrieved log entries by the forensic investigator in order to acquire information relevant to the case [22]. During the interrogation process, an in-depth exploration of the filtered log files is conducted. This may involve the application of specialized digital forensic techniques to gather evidence and to further scrutinize the log file entries [1, 29]. The analysis of potentially large amounts of data is rendered feasible using layers of abstraction and, more specifically, by analyzing the network abstraction layer, which translates the lowest level data from a physical network to the data used by applications [7]. Multiple back-ups

of the filtered evidence should be created before the collected log entries are analyzed.

The process of interrogation must make the evidence visible by clarifying its originality and significance [3]. The interrogation process, which is associated with catechization, is broken down into three sub-processes: (i) pre-processing; (ii) pattern discovery; and (iii) pattern analysis [10]. Catechization involves self-reflection, learning, gaining new skills and knowledge about how the events took place.

The goal of pre-processing is to produce a structural, reliable and integrated data source for pattern discovery [10]. Pre-processing involves five tasks: (i) data cleaning; (ii) transaction identification; (iii) session identification; (iv) data integration; and (v) transformation [10].

Data cleaning eliminates irrelevant entries from the access log files, including entries that record "errors" or "failures," access records generated automatically by search engines, requests for picture files associated with requests for particular pages, and entries with unsuccessful HTTP status codes [19].

Transaction identification follows data cleaning. The goal of transaction identification is to create meaningful clusters of references for each individual who accessed the organization's network [10]. The log entries are also partitioned into logical clusters using one or a series of transaction identification modules [19].

The next task in pre-processing is session identification. A session covers the activities performed by a user from the time he logs into a computer network to the time he logs out [10]. Session identification helps understand how a cyber criminal maneuvered within the organization's computer network [19]. It involves segmenting the access log of each log file entry into individual sessions [19]. A session includes the IP address (source address), user ID, URLs of the accessed sites and the access times (timestamps) [10]. Session identification also helps establish if the computer used to access the network was located in an Internet cafe, a classroom or a residence [19].

Data integration (or data fusion) is performed after all the relevant transactions and sessions have been identified. In this task, the matching transactions and sessions are combined [19].

The final task in pre-processing is transformation. The transformation task checks that the pre-processing sub-process has been performed completely and correctly, so that a good foundation is laid for the pattern discovery and pattern analysis sub-processes that follow [10].

Pattern discovery follows the pre-processing sub-process. Pattern discovery is an important activity that involves the application of algorithms and techniques from areas such as data mining, machine learning,

pattern recognition and statistics [10]. Pattern discovery also involves thorough searches for passwords used to access resources, unusual hidden files and directories that were accessed, and file extension and signature mismatches [19]. Techniques such as statistical analysis, association rules, clustering, classification, dependency modeling and path analysis are used to analyze the pre-processed log file data to discover criminal activity in the network [10].

The final sub-process in the interrogation process is pattern analysis. Pattern analysis, also called reconstruction, has two major activities; (i) resolution; and (ii) backtracing [19].

Resolution seeks to extract salient rules, patterns and statistics by eliminating irrelevant data [10, 19]. Various tools and techniques may be used to facilitate the transformation of information into useful knowledge [10].

Backtracing, which follows resolution, involves the reconstruction of criminal activities in the organization's computer network [19]. It uses a source IP address acquired in session identification to trace back to the Internet Service Provider (ISP) and ultimately to the source computer [3]. Backtracing also enables the investigator to ascertain the password and user ID by referring to the timestamps in the log files of the source computer [8, 11].

3.3.3 Analysis. The analysis process involves technical reviews by the investigator of the interrogation process. The major activity in the analysis process is correlation. The correlation of events in log files involves identifying relationships between fragments of data, analyzing hidden data and determining the significance of the log files from the source computer and the filtered log files [11]. Reconstructing event data based on the extracted data and arriving at appropriate conclusions are also part of the correlation activity [17].

User IDs, passwords and user names from the source logs and filtered logs must be correlated to establish temporal relationships. Timestamps based on the Coordinated Universal Time (UTC) can provide proof of when the criminal activities occurred [5, 8]. UTC is a high-precision atomic time standard based on the earth's rotation. It enables log file events to be analyzed and correlated regardless of differences in the time zones of the source log files and filtered log files [8, 11].

Other key activities in the analysis process include timeframe analysis, hidden data analysis, application analysis and file analysis [28]. The results of the analysis process should be documented completely and accurately for use in legal proceedings [24].

3.4 Elucidation Layer

The elucidation layer is the fourth and final layer in the log file forensic model (Figure 1). This layer focuses on explaining the outcomes of all the processes in the investigation. The elucidation layer comprises two processes: (i) presentation; and (ii) conclusion and follow-up.

3.4.1 Presentation. The results of the investigation must be presented to corporate management and law enforcement officials as well as to judges and juries, attorneys and expert witnesses in legal proceedings [6]. The individual results of all the processes must be combined to provide a clear picture to the audience [14]. The results of the interrogation and the analysis processes must be reviewed in their entirety to elicit a complete picture.

Since opposing theories will also be presented in court, there is a need to provide substantiated exhibits of the events that occurred as well as support for the theory or model of the events that occurred [9]. A standardized model facilitates a proof of the legitimacy of the theory [9]. The model must be supported by the evidence and should be based on the applicable laws and regulations [13].

A report comprising an abstract of the various investigative processes and the findings must be prepared for submission along with the evidence [13]. Supporting materials such as the filtered raw log files entries, chain of custody documents and details of various items of evidence should also be readied for submission [17].

3.4.2 Conclusion and Follow-Up. The conclusion and follow-up process involves reviewing all the steps in the investigation and identifying areas that may need improvement [7]. The results and their subsequent interpretation can be used to further refine network log investigations [13]. The conclusion and follow-up process also involves the distribution of information to provide a basis for future investigations [2]. Also, the applicable policies and procedures regarding information sharing must be followed.

3.5 Precautions

Precautions must be observed from the discovery layer all the way through the elucidation layer. Following the precautions mentioned when describing the processes in each layer reduces the number of mistakes made during the investigation and their impact. Many of the precautions are well-known in the field of digital forensics, while others are unique to network log investigations. The main precautions are: (i)

avoid experiments on the original copies of log entries [7, 18]; (ii) account for any and all changes to an original copy of the logs (the responsible investigator must document his name and the nature and the time of the alteration) [7, 18]; (iii) observe all relevant IT best practices (e.g., ISO 27002, COBIT, COSO and PCI DSS); (iv) consult experts when an investigation becomes difficult or complicated; (v) maintain strict chain of custody procedures throughout the investigation [30]; and (vi) use good faith, and be diligent, conscientious and meticulous.

3.6 Laws and Regulations

Due to the nature of the digital forensic process and the applicable laws and regulations, mistakes can be costly. Therefore, it is vital to understand the impact of laws and regulations on the forensic investigation processes [15]. All the processes in the various layers must be conducted according to the prevailing laws and regulations, including the local and/or international regimes as the case may be. Indeed, laws and regulations must be considered carefully from even before the investigation begins to the time that the case is resolved. Otherwise, the victim organization may suffer considerable loss, penalties may be levied for evidence spoliation and criminals may go unpunished.

4. Conclusions

The proposed log file forensic model is intended for use in computer network investigations, especially those involving network log mining. The forensic model is specifically designed to enhance the admissibility and trustworthiness of evidence in legal proceedings. The model focuses on the extraction, analysis and correlation of data from log files. Also, it emphasizes the strict observance of precautions and applicable laws and regulations during all the phases of an investigation.

References

[1] K. Arthur and H. Venter, An Investigation into Computer Forensic Tools, Technical Report, Information and Computer Security Architectures Research Group, Department of Computer Science, University of Pretoria, Pretoria, South Africa, 2005.

[2] D. Ayers, A second generation computer forensic analysis system, *Digital Investigation*, vol. 6(S), pp. S34–S42, 2009.

[3] V. Baryamureeba and F. Tushabe, The enhanced digital investigation process model, *Proceedings of the Digital Forensics Research Workshop*, 2004.

[4] N. Beebe, and J. Clark, A hierarchical, objective-based framework for the digital investigation process, *Digital Investigation*, vol. 2(2), pp. 146–167, 2005.

[5] F. Buchholz and B. Tjaden, A brief study of time, *Digital Investigation*, vol. 4(S), pp. S31–S42, 2007.

[6] B. Carrier and J. Grand, A hardware-based memory acquisition procedure for digital investigation, *Digital Investigation*, vol. 1(1), pp. 50–60, 2004.

[7] B. Carrier and E. Spafford, Getting physical with the digital investigation process, *International Journal of Digital Evidence*, vol. 2(2), 2003.

[8] D. Casey, Turning log files into a security asset, *Network Security*, vol. 2008(2), pp. 4–7, 2008.

[9] S. Ciardhuain, An extended model of cybercrime investigations, *International Journal of Digital Evidence*, vol. 3(1), 2004.

[10] R. Das and I. Turkoglu, Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method, *Expert Systems with Applications*, vol. 36(3), pp. 6635–6644, 2009.

[11] D. Forte, The "art" of log correlation: Part 1, Tools and techniques for correlating events and log files, *Computer Fraud and Security*, vol. 2004(6), pp. 7–11, 2004.

[12] D. Forte, The importance of log files in security incident prevention, *Network Security*, vol. 2009(7), pp. 18–20, 2009.

[13] F. Freiling and B. Schwittay, A common process model for incident response and computer forensics, *Proceedings of the Conference on IT Incident Management and IT Forensics*, 2007.

[14] J. Giordano and C. Maciag, Cyber forensics: A military operations perspective, *International Journal of Digital Evidence*, vol. 1(2), 2002.

[15] R. Ieong, FORZA – Digital forensics investigation framework that incoporates legal issues, *Digital Investigation*, vol. 3(S), pp. S29–S36, 2006.

[16] B. Jones, Comment – Virtual neighborhood watch: Open source software and community policing against cybercrime, *Journal of Criminal Law and Criminology*, vol. 97(2), pp. 601–629, 2007.

[17] K. Kent, S. Chevalier, T. Grance and H. Dang, Guide to Integrating Forensic Techniques into Incident Response, NIST Special Publication 800-26, National Institute of Standards and Technology, Gaithersburg, Maryland, 2006.

[18] S. McCombie and M. Warren, Computer forensics: An issue of definitions, *Proceedings of the First Australian Computer, Network and Information Forensics Conference*, 2003.

[19] M. Munk, J. Kapusta and P. Svec, Data preprocessing evaluation for web log mining: Reconstruction of activities of a web visitor, *Procedia Computer Science*, vol. 1(1), pp. 2273–2280, 2010.

[20] G. Palmer, A Road Map for Digital Forensic Research, DFRWS Technical Report DTR-T001-01 Final, Digital Forensic Research Workshop, Utica, New York (`www.dfrws.org/2001/dfrws-rm-final.pdf`), 2001.

[21] S. Perumal, Digital forensic model based on Malaysian investigation process, *International Journal of Computer Science and Network Security*, vol. 9(8), pp. 38–44, 2009.

[22] A. Ramabhadran, Forensic investigation process model for Windows mobile devices (`www.forensicfocus.com/downloads/windows-mobile-forensic-process-model.pdf`), 2009.

[23] M. Reith, C. Carr and G. Gunsch, An examination of digital forensic models, *International Journal of Digital Evidence*, vol. 1(3), 2002.

[24] M. Rogers, J. Goldman, R. Mislan, T. Wedge and S. Debrota, Computer forensics field triage process model, *Proceedings of the Conference on Digital Forensics, Security and Law*, pp. 27–40, 2006.

[25] R. Rowlingson, A ten step process for forensic readiness, *International Journal of Digital Evidence*, vol. 2(3), 2004.

[26] A. Schuster, Introducing the Microsoft Vista event log file format, *Digital Investigation*, vol. 4(S), pp. S65–S72, 2007.

[27] B. Shebaro, F. Perez-Gonzalez and J. Crandall, Leaving timing-channel fingerprints in hidden service log files, *Digital Investigation*, vol. 7(S), pp. S104–S113, 2010.

[28] Technical Working Group for the Examination of Digital Evidence, Forensic Examination of Digital Evidence: A Guide for Law Enforcement, NIJ Special Report, NCJ 199408, U.S. Department of Justice, Washington, DC, 2004.

[29] E. Tug, M. Sakiroglu and A. Arslan, Automatic discovery of the sequential accesses from web log data files via a genetic algorithm, *Knowledge Based Systems*, vol. 19(3), pp. 180–186, 2006.

[30] P. Turner, Digital provenance – Interpretation, verification and corroboration, *Digital Investigation*, vol. 2(1), pp. 45–49, 2005.