Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 56 (2015) 370 – 375

Procedia Computer Science

International Workshop on Cyber Security and Digital Investigation (CSDI 2015)

Privacy Levels for Computer Forensics: Toward a More Efficient Privacy-preserving Investigation

Waleed Halboob [a,b,*], Ramlan Mahmod [a], Nur Izura Udzir [a], Mohd. Taufik Abdullah [a]

[a] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[b] Center of excellence in Information Assurance (CoEIA), King Suad University, Riyadh, Saudi Arabia

Abstract

Computer forensics and privacy protection fields are two conflicting directions in computer security. In the other words, computer forensics tools try to discover and extract digital evidences related to a specific crime, while privacy protection techniques aim at protecting the data owner's privacy. As a result, finding a balance between these two fields is a serious challenge. Existing privacy-preserving computer forensics solutions consider all data owner's data as private and, as a result, they collect and encrypt the entire data. This increases the investigation cost in terms of time and resources. So, there is a need for having privacy levels for computer forensics so that only relevant data are collected and then only private relevant data are encrypted. This research paper proposes privacy levels for computer forensics. It starts with classifying forensic data, and analyzing all data access possibilities in computer forensics. Then, it defines several privacy levels based on the found access possibilities. The defined privacy levels lead to more efficient privacy-preserving computer forensics solution.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Computer Forensics; Privacy protection; Privacy Levels; Cryptography; Forensic imaging

1. Introduction

The currently widely used procedure for collecting digital evidence in computer forensics involves the creation of a bit-by-bit image from the data owner's physical storage and then later analyzing the bit-by-bit image at a Computer Forensics Laboratory (CFL). Using this procedure, all of the data found in the storage of the data owner (suspect, victim, or any related party to the crime) are collected and analyzed. In fact, this procedure has been proven to be a non-practical solution because of increases in the quantities of storage and data commonly owned, which increase the

* Corresponding author. Tel.: +966-503792189; fax: +966-11-469 5237.
E-mail address: wmohammed.c@ksu.edu.sa

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.07.222

investigation cost in terms of the required time and resources [1,2]. The problem becomes worse when dealing with a server's storage because of the huge amount of data involved and many users not related to the crime under investigation.

In addition, this procedure creates a significant problem when the data owner's privacy is a concern. Collecting only relevant data is a key point for privacy preservation. Recently, a selective imaging concept has been proposed to gather only data relevant to the crime, which would reduce the investigation cost. However, selectively imaging only the relevant data is still not a sufficient solution for privacy preservation in computer forensics, and many other requirements must be addressed such as specifying privacy policies, collecting only relevant data, using cryptographic techniques, taking into account existing privacy act(s), and auditing the investigation process [3,4,5,6,7,8].

However, collecting the relevant data while ignoring irrelevant data is a key point for privacy preservation in computer forensics, as mentioned earlier, and then only private relevant data are encrypted. But, this needs having some privacy levels that which data have to be collected or not, and encrypted or not. Existing privacy-preserving computer forensics solutions [9,10,11,12,13,14] are not efficient because they require the collection and encryption of all the data. This is because collecting and encrypting all data is not an acceptable solution especially when dealing with shared storages or servers.

In this research paper, privacy levels are proposed for computer forensics. However, the privacy levels help for classifying the forensic data into relevant and non-relevant as well as classifying the relevant data into private and non-private. Hence, at the end, only relevant data are collected, and only private relevant data are encrypted during the collection process. This leads to improve the investigation efficiency by reducing the required costs (in terms of time and resources).

The rest of this paper is structured as follows. Section 2 presents a brief review of the related works on privacy preservation in computer forensics. Section 3 presents the proposed privacy levels for computer forensics. Section 4 discusses the proposed privacy levels. Finally, Section 5 concludes the paper with possible future work.

2. Related Work

Several researches [3,4,5,6,7,8] have studied the conflict between privacy preservation and computer forensics. These studied suggest specifying accountability and before-the-fact privacy policies, along with using cryptographic techniques for preserving the private data during the evidence acquisition phase. Furthermore, they suggest that existing computer forensics tools, such as EnCase [15], Forensic ToolKit [16], must consider existing privacy acts. In addition, cryptographic techniques can be used to provide a privacy protection.

However, existing solutions can be classified into two branches: policy-based and cryptographic approaches. The policy-based approaches can be further classified into two approaches namely policy statements and privacy policies [17,18]. In fact, the goal of using the policy-based approaches is to let the data owner know how his private data should be collected, used, and disclosed. The cryptographic approaches [9,10,11,12,13,14] protect the data owner's private data during the investigation process. They encrypt all relevant and irrelevant data (even private or not) using some cryptographic techniques such as searchable encryption technique. All data owner's data are considered relevant and private, means the entire data are collected and encrypted. This increases the investigation cost in terms of time and resources. As discussed earlier, there is a need for collecting only relevant data as well as encrypting only private relevant data.

In Halboob *et al.* [19], we proposed a computer forensics framework that includes a definition for some privacy levels for computer forensics. Nevertheless, it has been proven that one level is not practical. Though, in this research different and refined privacy levels are defined.

3. The Proposed Privacy Levels

The privacy levels are generally used to explain the levels of privacy protection that the data collector must provide. Recent studies on computer forensics consider all the data owner's data as private. Therefore, they encrypt and encrypt the entire data. In fact, considering all the forensic data to be private requires protecting (e.g., encrypting) all these data, which consumes more time for data encryption and decryption.

In general, to define privacy levels, the targeted data (e.g., forensic data) should be classified into several groups. These groups can be used to determine all the access possibilities, which lead to defining the required privacy levels. The data can be classified based on several factors, including the privacy, relevancy, etc.

In computer forensics, there is a conflict because the data owner has no right to prevent the investigator from accessing his private data. However, the data owner has a right to decide whether his data, or a set of his data, are private and has a right to ask for privacy protection under any privacy act or policy applied in the area that the computer crime occurred. At the same time, the investigator has a right to collect any data, private or not, relevant to the crime. Thus, classifying the forensic data into some groups is a cooperative task between the data owner and investigator. Though, to define privacy levels for computer forensics, several steps should be taken into account, which are:

- Classifying the forensic data into groups based upon their privacy and relevancy.
- Analyzing all the data access possibilities upon the classified data's groups.
- Defining the privacy levels.

3.1. Forensic Data Classification

Forensic data are classified here using two factors: *privacy* and *relevancy*. From the data owner's point-of-view, the forensic data can be classified either as private or non-private. In general, all data can be considered by default to be private, but this classification (private and non-private) is from a computer forensics perspective.

More specifically, when the data owner claims that sets of his/her data are non-private, it means non-private to the investigation authority. For example, the data that are freely disseminated such as movies, clips, newsletters, books, etc. can be considered to be non-private (or public data). Thus, there is no need to consider these types of data to be private to reduce the investigation time or, in other words, to minimize the required encryption time.

In fact, the data owner can decide whether a set of his own data is private or not based on the investigation's privacy protection status, which determines whether or not the investigation authority protects the collected data of the data owner during the investigation process. The privacy protection provided can be through policy means (e.g., ethics rules, policies) or the use of other information security solutions such as auditing, access control, etc. The data owner builds his/her trust upon the privacy protection granted by the investigation authority. In other words, the trust degree is about how much the data owner trusts the investigation authority in terms of the privacy protection solutions provided. In computer forensics, some investigation authorities have some privacy protection granted as an internal access authorization such as ethical rules, policies, access control, and auditing.

At the end, a data owner may consider a set of his data to be private or not based on several factors such as the data sensitivity, investigation authority, privacy protection granted, etc. If the investigation authority provides some privacy protection guarantee to the data owner, the data owner may trust the investigation authority based on this guarantee and classify his data as non-private. Based on this trust degree, the data owner makes his/her decision during the classifying of his/her data as private or non-private.

From the investigator's point-of-view, forensic data can be classified as either relevant or non-relevant. The relevant data are any data considered by the investigator to be related to the crime being investigated. The forensic data relevancy is determined based on the investigation's goal and scope, which limit the investigator's access to certain data only. In addition, the data relevancy can be determined based on the crime being investigated. The investigator may know (based on the reported crime) that the requested digital evidence is, for example, a photo or email. Thus, there is no need to consider other types of data as relevant. In anyway, at the end the forensic data has the following four possible groups:

- Non-private and non-relevant
- Non-private and relevant
- Private and non-relevant
- Private and relevant

3.2. Forensic Data Access Possibilities

Based on the forensic data groups defined in the previous section, all the data access possibilities in computer forensics are analyzed and listed in Table 1. There are three access possibilities namely no collection, direct collection, and privacy-preserved collection.

Table 1. Forensic data access possibilities in computer forensics

No.	Forensic Data		Data Access Possibilities
	Relevancy (0: non-relevant; 1: relevant)	Privacy (0: non-private; 1: private	
1	0	0	No collection
	0	1	
2	1	0	Direct collection (no privacy concern)
3	1	1	User's choice (direct or in a privacy-preserving manner).

In addition, as shown in the above table, there are three forensic data access possibilities:

- If the forensic data (or a subset of forensic data) are classified by the investigator as non-relevant (private or not), then these data (or this subset of data) should not be collected at all. This means selective imaging must be used to collect only relevant data.
- When the forensic data (or a subset of forensic data) are classified as non-private and relevant by the data owner and investigator, respectively, these data (or this subset of data) can be collected directly. This means the data are collected without any privacy preservation.
- If the forensic data (or a subset of the forensic data) are considered to be relevant and private by the investigator and data owner, respectively, the data owner has a right to choose whether his data can be collected directly or in a privacy-preserving manner.

3.3. Defined Privacy Levels

As illustrated in Table 1, there are three different possibilities for accessing the forensic data: no collection, direct collection, and privacy-preserving collection. Thus, three privacy levels are defined for computer forensics based on these defined access possibilities. These levels are listed in Table 2.

TABLE 2. Privacy levels for computer forensics

No.	Level Name	Description
1	Direct Accessible Data (DAD)	Data are relevant and non-private so it can be directly imaged and analyzed.
2	Privacy-Preserved Accessed Data (PAD)	Data are relevant as well as private. Therefore, a privacy preservation technique(s) needs to be applied during the data imaging and analysis.
3	Non-Accessible Data (NAD)	These data are not relevant to the investigated crime and are not accessible at all.

4. Discussion

In this Section, the proposed privacy levels are discussed, by presenting a scenario to show how they can be enforced while investigating. To enforce the privacy levels, the forensic data must be classified first into relevant and

non-relevant as mentioned earlier. Also, the relevant data need to be classified into private and non-private. In fact, selecting the relevant or private forensic data requires the application of a pre-analysis process before imaging any data. Normally, the analysis step comes after collecting the forensic data and is executed at a Computer Forensics Laboratory (CFL). As needed in the selective imaging concept, selecting which data are private and relevant requires the application of a pre-analysis step before collecting or imaging the forensic data.

Therefore, for relevant or/and private data selection, current and suitable forensic data recovery tools are suggested here. According to Turner [20], the forensic data can be selected using three selection methods, namely manual, semi-automatic, and fully automatic selections.

Some recovery tools such as EnCase [15], FTK Access Data [16], Winhex (X-Ways) [21], and CnWRecovery [22] have been tested, and it has been shown that they can be used for this purpose. In the other words, both manual and semi-automatic forensic data selection are totally supported by these tools while the fully automatic forensic data selection still needs more research works [23,24]. However, these tools support full storage or disk scanning and analysis and allow the data owner or/and investigator to search for a set of data with several search types and options. The search results can be reported using csv files, and the listed data files (inside the reports) can be processed later, even by other applications.

Depending on several factors such as the relevant privacy act, investigation time, etc., the data owner may consider all his data to be private or non-private. In addition, the investigator may consider all the forensic data to be relevant or non-relevant. In this case, there is no need to execute the pre-analysis selection process.

Additionally, a very important issue is that this classification can easily be flexible. The data owner can consider all his data to be private or non-private. In this case, the investigation cost in terms of time can differ because considering all the data to be non-private means there is a need to encrypt all the data. The encryption process is an extra time cost during data imaging. In terms of resources, encrypting all the data may lead to an increase in the storage required for the collected data depending on the encryption scheme used. In anyway, if the data owner chooses to consider all his data as private or non-private so, in this case, there will be only two privacy levels which are DAD and NAD, or PAD and NAD.

Moreover, the investigator can consider all the found forensic data to be relevant. If this happens, it increases the required investigation time and resources. In this case, only two privacy levels will be considered namely DAD, and PAD.

5. Conclusion

In this paper, privacy levels are proposed for privacy-preserving computer forensics solutions. The privacy levels are specified based on existing access data possibilities found in computer forensics. They can help for improving the investigation efficiency when the privacy is a concern. Our ongoing work is developing a privacy-preserving computer forensics framework based on these privacy levels. Further work is required for defining such levels and policies for network forensics, and supporting fully automatic data selection for selecting the private and/or relevant data.

REFERENCES

1. Stüttgen, J. (2011). *Selective Imaging: Creating Efficient Forensic Images by Selecting Content First.* Diploma Friedrich Alexander Universität Erlangen Nürnberg.
2. Stüttgen, J., Dewald, A., & Freiling, F. C. (2013). *Selective Imaging Revisited.* Paper presented at the Seventh International Conference on IT Security Incident Management and IT Forensics, Nuremberg (Nürnberg), Germany.
3. Burmester, M., Desmedt, Y., Wright, R., & Yasinsac, A. (2002). *"Security or Privacy, Must We Choose?"*. Paper presented at the Symposium on Critical Infrastructure Protection and the Law.
4. Bui, S., Enyeart, M., & Luong, J. (2003). Issues in Computer Forensics: Santa Clara Universoty Computer Engineering, USA.
5. Saboohi, M. (2006). Collecting Digital Evidence of Cyber Crime.
6. Adams, C. W. (2008). *Legal Issues Pertaining to the Development of Digital Forensic Tools.* Paper presented at the Third International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE '08), Oakland, California, USA.
7. Fahdi, M. A., Clarke, N. L., & Furnell, S. M. (2013). *Challenges to Digital Forensics: A Survey of Researchers & Practitioners Attitudes and Opinions.* Paper presented at the Information Security for South Africa, Johannesburg.
8. Saleem, S., Popova, O., & Bagillib, I. (2014). Extended Abstract Digital Forensics Model with Preservation and Protection as Umbrella Principles. *Procedia Computer Science, 35*(2014), 812 – 821.

9. Croft, N. J., & Olivier, M. S. (2006). Sequenced Release of Privacy Accurate Call Data Record Information in a GSM Forensic Investigation. Paper presented at the Information Security South Africa, Pretoria, South Africa.
10. Croft, N. J., & Olivier, M. S. (2010). Sequenced Release of Privacy-accurate Information in a Forensic Investigation. Digital Investigation, 7(1-2), 95-101.
11. Law, F. Y. W., Chan, P. P. F., Yiu, S. M., Chow, K. P., Kwan, M. Y. K., Tse, H. K. S., & Lai, P. K. Y. (2011). *Protecting Digital Data Privacy in Computer Forensic Examination* Paper presented at the IEEE Sixth International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), Oakland, CA
12. Hou, S., Uehara, T., Yiu, S. M., Hui, L. C. K., & Chow, K. P. (2011). *Privacy Preserving Confidential Forensic Investigation for Shared or Remote Servers* Paper presented at the Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Dalian
13. Hou, S., Uehara, T., Yiu, S. M., Hui, L. C. K., & Chow, K. P. (2011). *Privacy Preserving Multiple Keyword Search for Confidential Investigation of Remote Forensics.* Paper presented at the Third International Conference on Multimedia Information Networking and Security, Shanghai, China.
14. Hou, S., Yiuy, S.-M., Ueharaz, T., & Sasakix, R. (2013). A Privacy-Preserving Approach for Collecting Evidence in Forensic Investigation. *International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2*(1), 70-78.
15. Software, G. (2014). EnCase Forensics V7. Retrieved 4 February, 2015, from https://www.guidancesoftware.com/products/Pages/encase-forensic/overview.aspx
16. AccessData (2013). Forensic ToolKit (FTK 5.5). Retrieved 5 March, 2014, from http://accessdata.com/solutions/digital-forensics/forensic-toolkit-ftk
17. Srinivasan, S. (2006). *Security and Privacy in the Computer Forensics Context.* Paper presented at the International Conference on Communication Technology (ICCT'6), Guilin.
18. Srinivasan, S. (2007). Security and Privacy vs. Computer Forensics Capabilities Information *Systems Control Journal, 4*, 1-3.
19. Halboob, W., M. Abulaish, and K.S. Alghathbar. *Quaternary Privacy-Levels Preservation in Computer Forensics Investigation Process*. in *The 6th International Conference for Internet Technology and Secured Transactions (ICITST-2011)*. 2011. IEEE.
20. Turner, P. (2006). Selective and Intelligent Imaging using Digital Evidence Bags. *Digital Investigation, 3*(1), 559-564.
21. X-Ways (2015). WinHex: Computer Forensics & Data Recovery Software, Hex Editor & Disk Editor. Retrieved 5 May, 2015, from http://www.x-ways.net/winhex/index-m.html
22. CnWRecovery (2015). CnWRecovery. Retrieved 5 May, 2015, from http://www.cnwrecovery.co.uk/index.html
23. Halboob, W., et al. *An Efficient Computer Forensics Selective Imaging Model*. in *The 8th FTRA International Conference on Further Information Technology*. 2014. Gwangju, Korea: Springer.
24. Halboob, W., et al., *An Ordered Selective Imaging and Distributed Analysis Computer Forensics Model.* Journal of Applied Sciences, 2014**. 14(**21): p. 2704-2712.