

Ontology-based Privacy Preserving Digital Forensics Framework

Xuejiao Wan¹, Jingsha He¹, Na Huang¹ and Yonghao Mai²

¹*School of Software Engineering
Beijing University of Technology
Beijing 100124, China*

²*Department of Information Technology
Hubei University of Police
Wuhan, Hubei 430034, China
jhe@bjut.edu.cn*

Abstract

Along with the rapid growth of the number of intelligent mobile devices as well as applications in recent years owing to the development of network information and telecommunication technologies, personal privacy faces a unprecedented level of risk. This seems to be especially inevitable in digital forensic investigations. Consequently, there is an urgent need to balance the desire for privacy preservation and information extraction in the process of digital forensic investigations. This paper focuses on the issues related to personal privacy protection by proposing a digital forensics framework based on the theory of ontology. In addition to introducing the basic concepts that will lead to the establishment of the framework, we will also present some experiments that can be viewed as the demonstration of the effectiveness of the framework in privacy protection in the process of digital forensic investigations. The framework can thus serve as the foundation in the design of digital forensic tools and systems that can respect the privacy of individuals.

Keywords: *Forensics; Ontology; Privacy; Android*

1. Introduction

As intelligent mobile devices grow in complexity and maturity in recent years, smart phones and other intelligent mobile devices have experienced a wide-spread use and thus quickly become an important part of the sources for getting digital evidence and hence the main focus of digital forensics research. California has extended privacy protection to the contents of cell phones, settling a judicial split by prohibiting police from searching cell phones in the incident leading to arrest [1]. A growing number of courts believe that a warrant is necessary because of the ubiquity of mobile devices so that the majority of such devices are Internet connected smartphones that contain text messages, pictures, videos, emails and other sensitive and personal information. In Hawaii, New York, Oregon and Washington, just to name a few, it is now the requirement that police use a search warrant to track a person's movement with a GPS or other electronic tracking devices. The office of the privacy commissioner for personal data in Hong Kong has suggested a number of privacy practices to highlight the importance of privacy protection of personal data [2]. However, the above practices are mainly non-technical means in defining and enforcing privacy requirements during the process of digital forensics.

To respond to privacy requirements as technologies change our society, some attentions have started to be placed on developing technical means to balance the relationship between convenience and privacy. However, privacy protection has not received insufficient attention in current research on digital forensics. In this paper, we propose a novel privacy preserving digital forensics framework aiming to bridge the gap between the

existing digital forensic disciplines and privacy policies. It worth mentioning that ontology provides a strong foundation for the development of privacy preserving methodologies in the area of digital forensics. In this paper, we propose ontology description model and a privacy preserving digital forensics framework as an effective solution to the significant challenge digital forensics investigators faces.

The remainder of this paper is organized as follows. Section 2 discusses some prior work on the issues of privacy preservation in digital forensics. Section 3 presents the proposed ontology-based privacy preserving digital forensics framework in which we will investigate how ontology can be designed to represent the particular forensics scenarios and then propose the privacy preserving digital forensics framework to meet the demands of specific scenarios. Section 4 serves to demonstrate the effectiveness of the proposed ontology through conducting some experiments using some real data. Finally, Section 5 concludes this paper with an analysis of different types of digital forensics scenarios and some discussions on future development prospect.

2. Related Work

The combination of smartphones and the Internet has become very indispensable to the users. However, such dependency also makes it easy to leak personal private information. Phone calls, emails, text messages and other forms of communication would provide to observers with so many pieces of the puzzle that they could be used as the clues for deducing personal identities and other critical information that can used to break into the devices to get more information about the users. Android has been perceived as a relatively insecure mobile operating system due to its open source nature, but a company named Silent Circle has relabeled it "PrivatOS" through making some significant changes into it that are both visible and behind the scenes [3]. An Android phone known as Blackphone is claimed to be the first smartphone to place "privacy and control directly in the hands of its users" while being used a serious digital device that could carry a wide variety of user personal information [4].

In order to semantically interpret the input data and perform data transformations aiming to minimize the loss of semantic content for unbounded categorical attributes, Martínez et al. [5] proposed a new masking method relying on the knowledge in ontologies. Ntantogian et al. [6] revealed a set of critical observations regarding the privacy of Android mobile applications and devices. In [7], the authors analyzed and retrieved files from several storage media devices by using software called DriveSpy while working in the DOS command mode focusing on floppy drives, which the authors could show to be able to grasp the detailed structure of an entire file system by manually examining a floppy disc sector by sector within a few hours.

Privacy requirements may define constraints on data types, forensic scenarios and forensic scopes. Kost and Freytag [8] described a formal approach to technically evaluate an information system with respect to privacy protection in the design and implementation. Gupta [9] proposed the Privacy Preserving Efficient Digital Forensic Investigation (PPEDFI) framework to extract the evidence files and rank them based on their relevance of being used as the evidence. In [10], the authors suggested that policies for privacy protection be abstracted as trust attributes in a semantic way. Kanbe et al. [11] proposed RFID privacy protection ontology that consists of goal ontology, function ontology and technology ontology. However, although there are currently a lot of discussions on privacy preservation in the context of digital forensics, there are hardly any general methods that have been developed to be able to deal with the privacy issues in digital forensics. Therefore, the purpose of this paper is to introduce ontology into the privacy issue and customize it to assess the feasibility and the impact on the protection of personal privacy in several real digital forensics scenarios. Then, an ontology-based privacy preserving digital

forensics framework is proposed with an analysis on its feasibility as well as its effectiveness in providing the necessary protection on private information in digital forensics.

3. The Ontology-based Framework

As digital mobile devices are used extensively to exchange information between users, privacy concerns has arisen due to potential leakage of personal and private information. Digital devices are usually used to store information with common applications that may touch sensitive data about users, such as photographs taken, electronic accounts established, chat records sent and received, clocks and calendars scheduled, contacts entered into the address book, and social network usernames and passwords. Furthermore, mobile devices are often considered to be very personal and private from the viewpoint of performing the role of personal organizers and being carried around and used by individuals. Stirparo et al. [12] analyzed how mobile applications are used to manage user data after it is loaded into the volatile memory of the devices to raise the awareness of the research and development communities on the poor attention that is generally paid in the secure development of mobile applications. Nevertheless, a technical approach for privacy analysis is largely missing in which a framework will certainly help to fill the gap.

3.1. Disclosure of Privacy Information in Digital Forensics

For the purpose of showing what could be revealed when browsing through the data in a captured device just like what an forensic investigator would do, in this part, we test 5 archives download from the Oxygen Forensic Website to emulate real connected devices. The data are analyzed using Autopsy, a digital forensics platform and graphical interface to the open-source tool The Sleuth Kit (TSK). Figure 1 illustrates a view of the tree structure of the files stored in the five devices discovered by Autopsy, where we can see clearly that personal data such as SkypeAccounts, ImportantData and InternetHistory are readily available to the investigator.

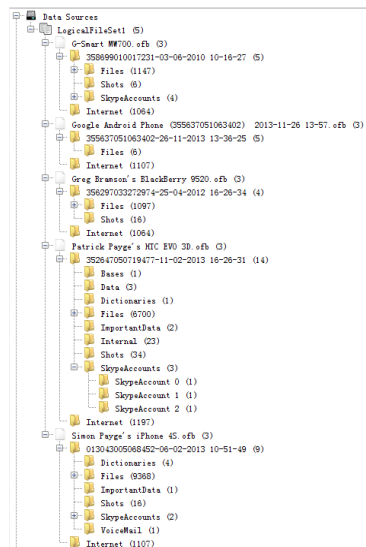


Figure 1. Forensic Data of the 5 Devices using Autopsy

Figure 2 shows that a large amount of forensic data including images, videos, audios and documents exist in the digital devices. Apparently, sensitive data related to personal

privacy is revealed. Figure 3 shows the images stored in the forensics digital devices at hand. Figure 4 shows the Skype chat records and user account information. Similarly, we can see from Figures 5 and 6 that important private data becomes available in the forensics views window.

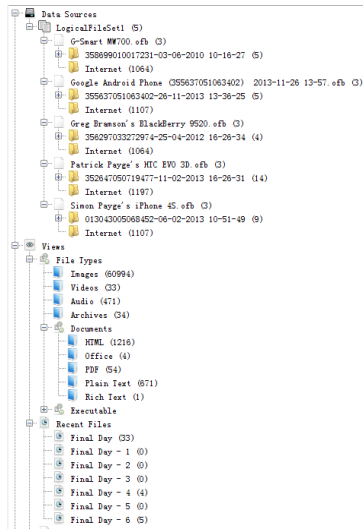


Figure 2. Types of the Extracted Files

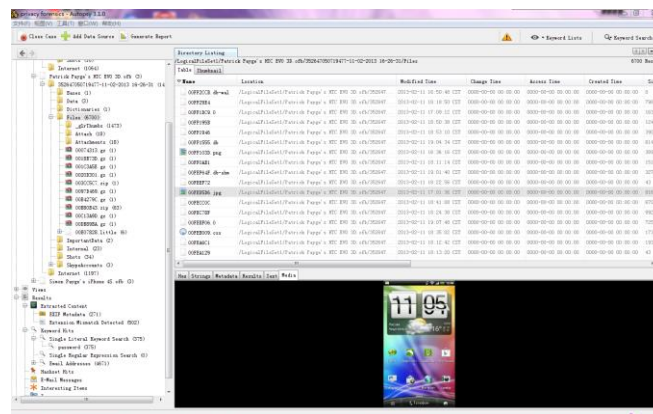


Figure 3. Images Stored in the Mobile Devices

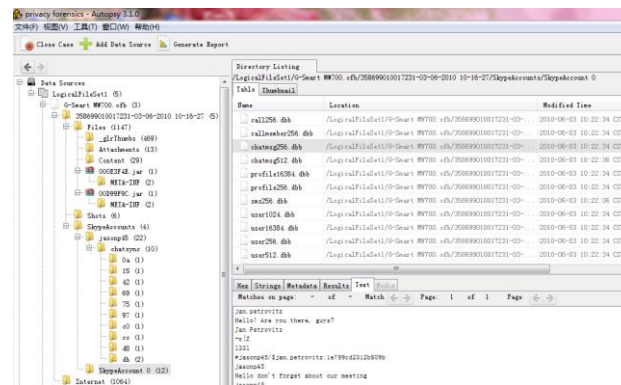


Figure 4. G-Smart Phone SkypeAccount Context

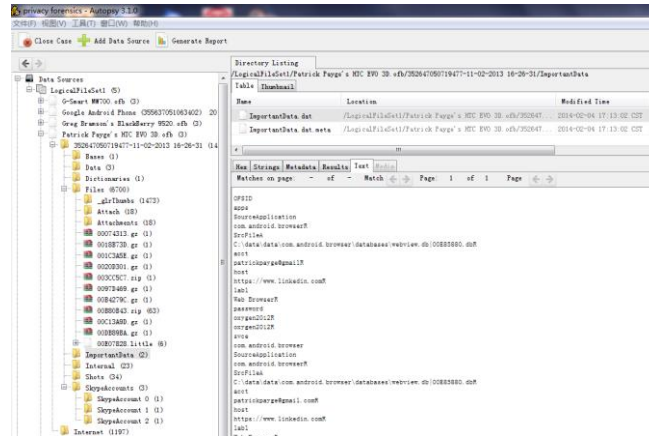


Figure 5. Important Data in the HTC Mobile Phone Used in the Experiment

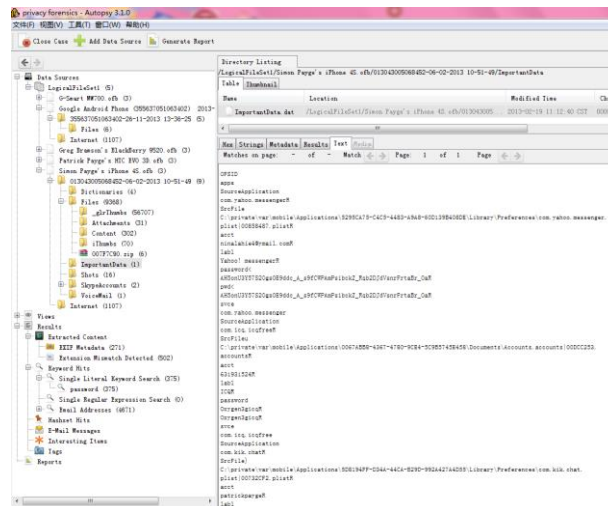


Figure 6. Iphone 4s Important Data

3.2. The Ontology for Digital Forensics Involving Privacy

Protégé is an open-source ontology design tool for building intelligent systems with a highly configurable user interface that includes RDF/XML, Turtle, OWL/XML, OBO and other formats available for ontology uploads and downloads. Therefore, in our study, we use Protégé as the tool to design the ontology for the privacy preserving digital forensics framework. In addition, there are three ontology definitions customized below to assist the design to illustrate the ontology.

Definition1: The attributes of the ontology for privacy in digital forensics are defined by the triplet (C, RC, HC): (1) C: the concepts involved in the digital forensics process. (2) RC: the morphisms between concepts. (3) HC: classification levels of the concepts.

Definition2: Cato= (C', RC') is a category of ontology for digital forensics. (1) Cato: ontology category. (2) C': the object of a category including classes defined in the ontology and the relationships between them. (3) RC': the morphisms between objects including classes and hierarchical relationships between objects.

Definition 3: The path of concepts in the ontology is supposed to be a path through the hierarchy that starts from a specified concept node and ends at the root node.

According to the above notions of ontology, we can form a set of ontologies. Figure 7

illustrates how we can classify forensics processes through applying Protégé as the ontology design tool. The figure also depicts the hierarchy of the mentioned ontology in detail.

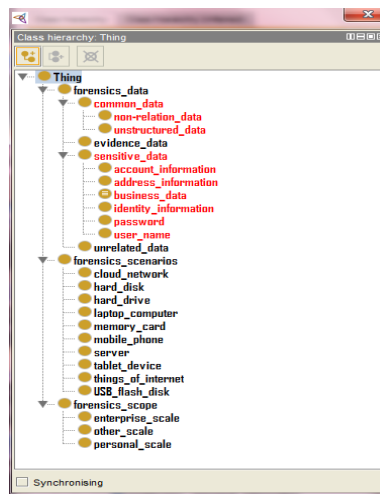


Figure 7. Digital Forensics Privacy Ontology

Finally, Figure 8 shows the ontology that is composed of 4 levels in horizontal direction. Moreover, digital forensics is divided into three branches as we can see in the figure. In general, the digital forensics regarded as an object is divided into three categories including forensics scope, forensics data and forensics scenarios.



Figure 8. Privacy Ontology of Digital Forensics

3.3. The Ontology-based framework for privacy preservation in digital forensics

Based on the ontology presented above, Figure 9 is our proposed framework for digital forensics for privacy protection. The ontology-based privacy usually consists of sensitive information such as username, password, personal account photograph, etc. The suggested framework can achieve the goal of protecting privacy through means of identifying authentication and the scenario as well as protecting sensitive information and images.

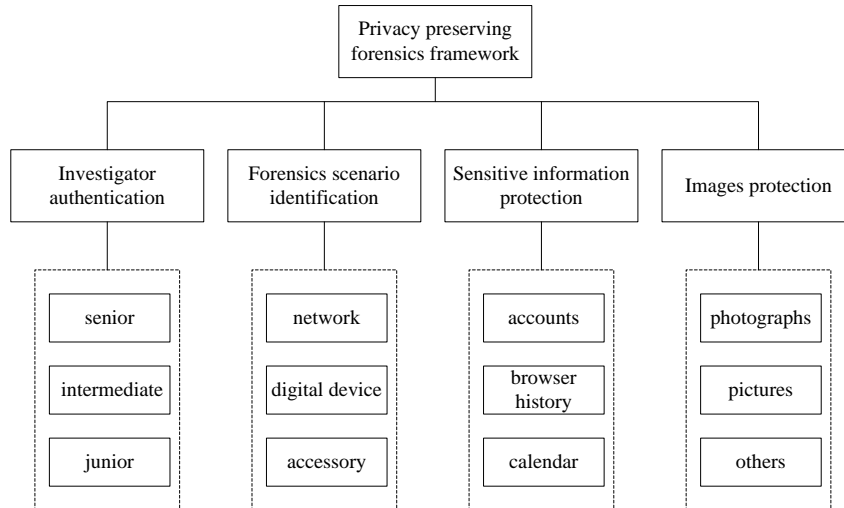


Figure 9. Ontology-based Privacy Preserving Digital Forensics Framework

With such protection, senior level investigators could be assigned permissions to view protected information and images without too much limitation, middle are able to visit permitted information through limiting the forensics scenario, and junior level investigators could be prohibited from viewing protected data for privacy protection in the process digital forensics.

4. Analysis of Effectiveness

Wynand van Staden [13] presented framework for enhancing privacy to prevent – or at least to minimize the risk of – third party from privacy breaches, which includes an analyzing component to profile a search and a filtering component to calculate the diversity in the search results. Nevertheless, Wynand overlooked the variety of forensics scenarios that are supposed to impact the result of forensics significantly. Figures 13-18 show how using our proposed framework, sensitive information can be properly protected when digital forensics is performed by investigators at different levels.

Figures 10 and 11 show how protecting Skype accounts can be realized through implementing the proposed ontology-based framework. Sensitive information is protected by considering the authentication of different investigators together with the identification of forensics scenarios.

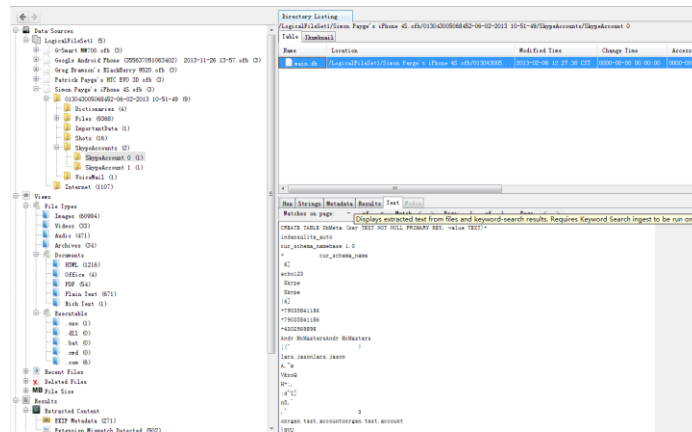


Figure 10. Senior View of SkypeAccounts

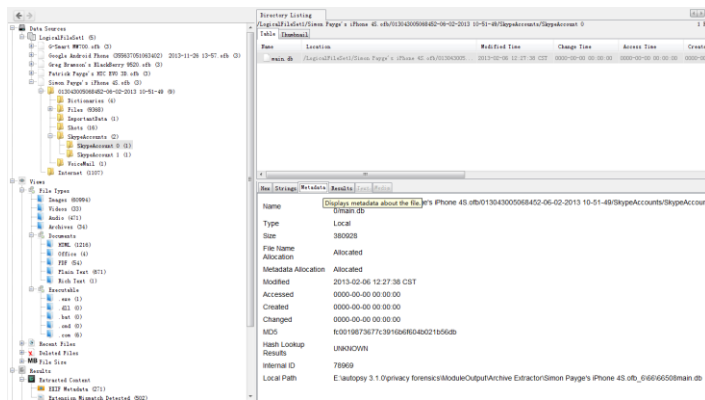


Figure 11. Junior or Intermediate View of Skype Accounts

The set of experiments can thus illustrate and verify the effectiveness of the proposed ontology-based framework for privacy preservation in digital forensics. We can see that the lack of proper policies of protecting privacy in digital forensics process will lead to serious privacy issues. On the contrary, the proposed privacy preserving digital forensics framework can serve to balance the privacy requirements and the needs of digital forensics investigations.

5. Conclusion

In this paper, we presented a privacy preserving forensics framework based on ontology by abstracting the privacy attributes in digital forensics scenarios. Privacy protection in the proposed ontology-based digital forensics framework is then realized through allocating rights at different levels of the hierarchy. We discussed the necessity of preserving sensitive information in the process of investigating digital devices and applied the proposed ontology to classify information involving privacy in three aspects: data types, forensics scenarios and forensics scopes. We also analyzed the effectiveness of the framework using real data. Such a privacy enhancing digital forensics framework would enable better categorization of digital forensic disciplines and assist in the development of methodologies and specifications that can offer direction in preserving privacy in digital forensics. Finally, forensic specialists, judicial surveyors and researchers with an interest in the areas of digital forensic tools, privacy discipline and development of forensics methodologies should find the framework in this paper very constructive.

Acknowledgements

The work in this paper has been supported by National Natural Science Foundation of China (61272500) and Beijing Natural Science Foundation (4142008).

References

- [1] H. Fakhoury, "A National Consensus: Cell Phone Location Records Are Private", <https://www.eff.org/deeplinks/2014/07/constitutionally-important-consensus-location-privacy>, (2014).
- [2] Office of the Privacy Commissioner for Personal Data, "Six Data Protection Principles (DPP) of the Personal Data (Privacy) Ordinance", <http://www.pcpd.org.hk/english/ordinance/ordglance.html>, (2014).
- [3] R. Granger, Stanford Researchers Team Up with Defense Firm Rafael and Create 'Gyrophone' A New Privacy App., [http://hereisthecity.com/en-gb/2014/08/18/new-smartphone-app-exposes-privacy-invasion-on-your-phon e/](http://hereisthecity.com/en-gb/2014/08/18/new-smartphone-app-exposes-privacy-invasion-on-your-phon-e/), (2014).
- [4] A. Souppouris, "Blackphone: An Android Phone That Puts Privacy First", <http://www.theverge.com/2014/2/24/5441642/blackphone-silent-circle-geeksphone-pre-order-launch>, (2014).
- [5] S. Martínez, D. Sánchez and A. Valls, "Ontology-Based Anonymization of Categorical Values", Proceedings of the 7th International Conference on Modeling Decisions for Artificial Intelligence, (2010); Perpignan, France.
- [6] C. Ntantogian, D. Apostolopoulos, G. Marinakis and C. Xenakis, "Evaluating the Privacy of Android Mobile Applications under Forensic Analysis", Computers & Security, vol. 42, (2014).
- [7] L. Batten and L. Pan, "Teaching Digital Forensics to Undergraduate Students", IEEE Security & Privacy, vol. 6, no. 3, (2008).
- [8] M. Kost and J. C. Freytag, "Privacy Analysis Using Ontologies", Proceedings of the 2nd ACM Conference on Data and Application Security and Privacy, (2012); San Antonio, TX.
- [9] A. Gupta, "Privacy Preserving Efficient Digital Forensic Investigation Framework", Proceedings of the Sixth International Conference on Contemporary Computing, (2013); Noida, India.
- [10] F. Gao, J. S. He, S. P., X. Wu and X. Liu, "An Approach for Privacy Protection Based on Ontology", Proceedings of the 2nd International Conference on Networks Security, (2010); Wuhan, China.
- [11] M. Kanbe and S. Yamamoto, "Ontology Alignment in RFID Privacy Protection", Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, (2009); Fukuoka, Japan.
- [12] P. Stirparo, I. N. Fovino and I. Kounelis, "Data-in-Use Leakages from Android Memory - Test and Analysis", Proceedings of the 9th International Conference on Wireless and Mobile Computing, Networking and Communications, (2013); Lyon, France.
- [13] W. Staden, "Protecting Third Party Privacy in Digital Forensic Investigations", Proceedings of the 9th IFIP Working Group 11.9 International Conference on Digital Forensics, (2013); Orlando, FL.

Authors



Jingsha He, he is currently a professor and Ph.D. advisor in the School of Software Engineering at Beijing University of Technology in Beijing, China. He received the B.S. degree in computer science from Xi'an Jiaotong University in Xi'an, China in 1982 and the M.S. and Ph.D. degrees from the University of Maryland at College Park in Maryland, U.S.A. in 1984 and 1990, respectively. Prior to joining Beijing University of Technology (BJUT) as an endowed professor of the City of Beijing in 2003, Prof. He worked for several hi-tech companies in the United States such as IBM Federal Systems (Senior Computer Scientist), MCI Communications Corp. (Senior Engineer), GRIC Communications, Inc. (Principal Engineer) and Fujitsu Laboratories of America, Inc. (Senior

Member of Research Staff) where he carried out research and development work in the areas of computer networking and information security. Prof. He's main research interests include computer and network security, network measurement and wireless communication protocols and security. His academic accomplishments include over 200 publications in scholarly journals and major international conferences, 12 U.S. patents and more than 30 China patents, 18 software copyrights, and 6 authored books.



Xuejiao Wan, she is currently a Master degree candidate in the School of Software Engineering at Beijing University of Technology in Beijing, China. Her main research interests include digital forensics and computer and network security.