Chapter 2

PROTECTING THIRD PARTY PRIVACY IN DIGITAL FORENSIC INVESTIGATIONS

Wynand van Staden

Abstract The privacy of third parties is an important issue in digital forensic investigations. The typical steps followed during an investigation require that the equipment used during the commission of the crime be seized and analyzed in a manner that complies with accepted investigative policy and practice. The analysis of the data on the seized equipment provides the investigator, who may not necessarily be associated with a law enforcement agency, with the opportunity to access personally identifiable information of persons or entities who may not be linked to the crime; this is especially true in multi-user environments.

This paper considers the key aspects surrounding privacy protection of third parties during the *post mortem* data analysis phase of digital forensic investigations. It proposes a framework that helps protect privacy without reducing the effectiveness of an investigation. The design includes a profiling component that analyzes a search corpus and a filtering component that calculates the diversity in the search results. Depending on the sensitivity level, the search results are either filtered out or are presented to the investigator.

Keywords: Privacy-enhancing technology, third party privacy protection

1. Introduction

Crimes involving computer equipment are investigated using procedures created specifically to surmount the difficulties associated with the digital nature of the crimes, and also to conform to the requirements for investigative procedures as may be required by law. In the discipline of digital forensics, there are typically two types of procedures, one geared for live investigations (investigations of crimes that are in

G. Peterson and S. Shenoi (Eds.): Advances in Digital Forensics IX, IFIP AICT 410, pp. 19–31, 2013.

progress), and the other targeted for *post mortem* investigations (investigations of crimes after they have been committed).

This paper is concerned with *post mortem* investigations, with a focus on multi-user environments. The investigative steps include the preservation, collection, examination and analysis of digital evidence [10, 15].

The collection phase involves the gathering of all digital data that may be relevant to the case. It includes the seizure of computing devices and the collection of data from storage media. The seized data normally pertains to the suspects and victims as well as third parties whose data and activities may have no bearing to the investigation. When an investigator examines collected data, he may stumble upon sensitive data relating to third parties. Even when filtering techniques are used, intrinsic or extrinsic connections between suspects/victims and third parties may result in third party data being scrutinized. This is an example of a third party privacy breach, where the act of investigating a crime inadvertently results in the loss of privacy for entities that are unrelated to the case.

This paper presents a privacy-enhancing framework that attempts to prevent – or at least minimize the risk of – third party privacy breaches. The approach draws on the basic privacy principle of information self-determination. In general, a privacy-enhancing technology (PET) protects access to personally identifiable information pertaining to the users of a computing system. During a digital forensic investigation, the reasonable expectation of privacy by entities that have no bearing to the case should not be discarded without due consideration [18]. Therefore, the framework is designed to protect third party privacy without preventing justifiable access to private information. The framework, which employs a filtering approach for weeding out data that is not relevant to an investigation, provides subjects in a multi-user environment with a reasonable expectation of privacy without reducing the effectiveness and speed of investigations.

2. Background

This section briefly discusses the main issues related to digital forensics, privacy and privacy-enhancing technologies (PETs).

2.1 Digital Forensics

Digital forensics is the branch of forensic science [10, 15] that is concerned with the "preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of evidence in a

digital context" [15]. Scientifically-sound and validated methods are required to be used in every phase of a digital forensic investigation.

During the collection phase, all digital information that may be relevant to the investigation is gathered [10, 17]. In practice, this often means creating bit-for-bit copies of the seized storage media. This is done to preserve the storage media content without any alteration. Typically, this is accomplished using a write-blocking device when copying the storage media.

Since all the data residing on the seized storage media is potential evidence, the investigator can be expected to access and examine all the data. In some jurisdictions, such as The Netherlands [18] and South Africa [13], data collected for investigative purposes is referred to as "police data." This data belongs to the police for investigative purposes. In both jurisdictions, the police are allowed to examine all police data without prejudice. Thus, the privacy rights of all the entities about whom there is data on the storage media are suspended during an investigation.

The analysis phase of a digital forensic investigation involves the systematic search of the collected information for data that is relevant to the case. This may involve searching through documents and other related files, searching the slack space and searching for data that may be hidden using anti-forensic techniques such as steganography [10, 17]. Since a large quantity of data has to be examined, a variety of digital forensic tools have been developed to support the analysis phase. Some of these tools (see, e.g., [8]) provide search filters that eliminate data that is considered to be noise (i.e., not relevant to an investigation). The filters typically use text strings (in documents or file metadata), file types and file timestamps [7].

Other techniques for filtering non-relevant data include text mining [2] and machine learning approaches such as self-organizing maps [7, 8]. Text mining [21] returns data that is important based on some relevance scale. Self-organizing maps [12] cluster data using various features to assist the investigator in forming opinions about the relevance of data. Noise may be filtered using the thematic clustering of search results [3].

The practice of filtering potential red herrings during the investigative process is well established. Filtering often yields false positives, i.e., data that may not be relevant to an investigation, which requires effort to classify and discard later in the investigation. False negatives are also a problem because they correspond to relevant data that is hidden from the investigator by the filtering tool.

2.2 Privacy

Efforts at protecting an individual's right to privacy have been undertaken for more than one hundred years [20]. Modern interpretations of the right to privacy are enshrined in the European Union Data Protection Directive [6] and the South African Protection of Personal Information Bill [13], that when enacted, will be on par with the European Union directive. Both these instruments deem privacy important enough to prohibit the cross-border flow of personal information to countries that do not have equivalent privacy laws. To reduce the trade impact of these laws, countries with strict privacy laws generally sign safe harbor treaties with non-compliant countries.

The principles governing the protection of private information were first listed by the Organization for Economic Cooperation and Development (OECD) [14]. The OECD principles, which are incorporated in the European Union and South African instruments, include collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation and accountability. When implemented as technological safeguards for privacy protection, these principles are commonly translated to the right to information self-determination [9, 19] – the right of an entity to determine which entities have data about it, what data these other entities have, what these entities may do with the data, and to which entities they may give the data [19].

Legislation typically provides a reasonable expectation of privacy to individuals who entrust their data to enterprises and service providers who make positive statements about their intent to protect user privacy. The exemption of these laws during an investigation may take individuals by surprise – in particular, the fact that personal information that is not relevant to the case may be perused by an investigator during the course of the investigation.

2.3 Privacy-Enhancing Technologies

Privacy protection systems rely on PETs to protect the privacy of individuals. These systems focus on purpose specification and use limitation [9, 19]. While the systems address only two of the eight principles of privacy protection, the other six principles are normally taken care of as a matter of operating procedure. Privacy protection systems ensure that personal data is not disseminated to entities who are not supposed to get the data, and certainly not to entities who may desire to use the data for unauthorized purposes [19].

PETs are effective and, as such, are promising candidates for protecting personal privacy during digital forensic investigations. The challenge is to protect personal privacy without reducing the effectiveness and speed of investigations.

3. Preventing Third Party Privacy Breaches

As discussed above, filtering techniques can be used to exclude content that is deemed to be not relevant to an investigation. A privacy aware system used during an investigation would incorporate enhancements to basic filtering techniques. The important point is that the system would restrict the results to data that has limited potential to constitute a breach of third party privacy.

Croft and Olivier [5] articulated the notion of privacy aware digital forensics. In particular, they proposed the sequenced release of data to protect the privacy of individuals under investigation. Their proposal pivots on hypothesis testing before any additional data is released. In other words, the system only releases data if it knows that the entity querying the system already has some prior data that is relevant.

The approach presented in this paper allows the privacy of third parties to be taken into account. The system evaluates the results of a query and either releases the results to the investigator or reports that the query may result in a privacy breach. In the latter case, the investigator is requested to provide a more focused query.

Based on the operating principles of a PET listed above, it could be argued that the framework presented here is not a PET because no explicit purpose specification and use limitation are present. However, it is safe to assume that the purpose of the data access is the investigation, and that the data is stored and accessed for the purpose of the investigation.

The PET forces the investigator to present focused questions when searching for evidence. The search results are provided to the investigator only after it is determined that the query is focused. Clearly, this approach should not encourage data sequestration that may hinder an investigator (i.e., the system reports false negatives). In the case of false positives, the guiding principle is accountability. If personal data that is not related to the investigation is accessed, the PET should record the access, enabling the affected entity to know that a privacy breach did in fact occur.

A PET used in an investigative process should have the following characteristics:

- The PET should not encumber the investigator. The investigator should not have to perform a significant number of additional tasks during an investigation.
- The PET should reasonably sequester data that is determined as being not relevant to the investigation. It should specifically limit data related to parties who are not under investigation.
- The PET should provide the investigator with the freedom to access data that is deemed to be false negative.
- The PET should support accountability and openness. It should record accidental and intentional accesses of private data that bears no relevance to the investigation.
- The PET should not undermine an investigation by adding significant time delays to established investigative techniques such as filtering and text mining.

4. PET Design

Our PET design for supporting privacy aware digital forensic investigations incorporates four principal components (Figure 1). The first is a profiling component that creates a profile of the data and stores the results in a meta-database; the profile is created just in time, when a new document is queried and the profile is reused after it is created. The second component is a query front-end that enables an investigator to perform searches (e.g., partial string and file type searches). The third component is a response filter that determines if the query results could potentially lead to a privacy breach. The fourth component is an audit log that is used for auditing compliance with the privacy principles.

Note that the meta-database that stores profiling results and caches *ad hoc* profiles, and the audit log that stores search queries and PET configuration changes must be tamper-proof.

5. Screening Search Results

As stated above, an investigator may gain access to sensitive data concerning third parties during the course of an investigation. This problem is mitigated by incorporating a filter that forces the investigator to focus search queries on the case at hand. The solution engages the

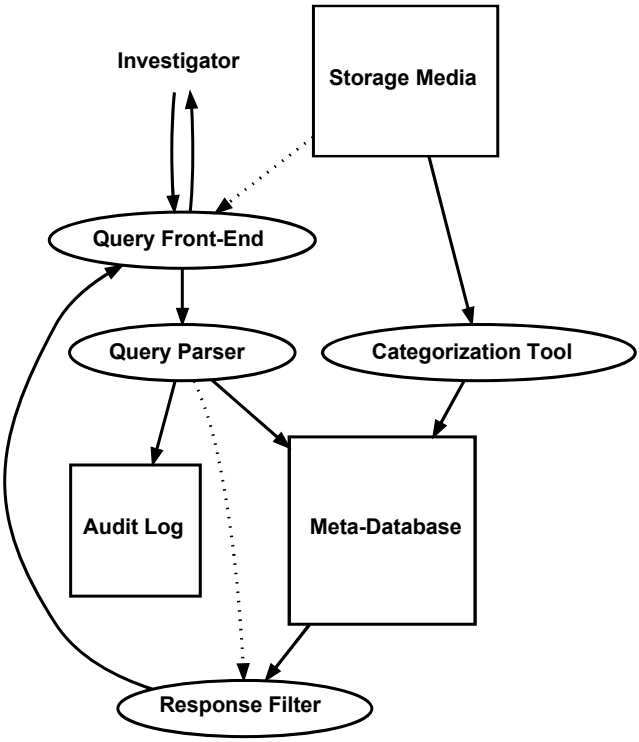

Figure 1. PET design for digital forensic investigations.

profiling component, a metric that assesses the diversity of search results and a sensitivity measure that determines the leniency to be provided to queries that return diverse search results.

5.1 Profiling

The profiling component uses n-grams [4], an established profiling technique that delivers good results. The n-gram technique relies on Zipf's law, which states that the n^{th} most common word in human text occurs with a frequency that is inversely proportional to n [4]. The implication is that two texts that are related should have similar frequency distributions. Using this principle, the text is tokenized and each token is deconstructed into n-grams where $2 \leq n \leq 5$ [4].

A frequency table is used for each n-gram in a text sample. After a document is scanned, a ranked list is created of the most frequently occurring n-grams. The number of entries in the rank table can vary; however, according to Cavnar and Trenkle [4], domain-specific terminol-

ogy only begins to appear around rank 300. The n-gram technique is fast and provides a usable profile for text.

After a profile is generated, it can be compared with another profile using a distance measure between the two profiles (or texts). Texts that are similar have a smaller distance measure than texts that are dissimilar. The next section describes how the distance is calculated.

5.2 Distance Calculation

Cavnar and Trenkle [4] use the rank-difference of n-grams as a distance measure:

$$\sum_{i}^{|P_k|} |r_{k(i)} - r_{j(i)}| \tag{1}$$

where P_k is the rank profile for text k and $r_{k(i)}$ is the rank of n-gram i for text k.

This distance measure takes content into account, but it ignores some important aspects. First, it is reasonable to assume that similar documents that have different owners are the result of two individuals working on the same project. Therefore, a difference in document owners should be considered when evaluating the similarity between documents. However, in an investigation, another user could be an accomplice of the suspect, so the existence of different owners should not skew the results too much.

Another aspect is the logical location of documents in a storage medium. This is important when two or more individuals work on a project, but only one of the individuals has been involved in malicious acts while working on the project. The goal here is to protect innocent individuals by carefully presenting the documents they own as separate and distant from the documents owned by the suspect.

The attributes used to determine the distance measure employ features used in anomaly detection by Fei, *et al.* [7]. In their approach, the documents that are clustered closely together using the features are worthy of further investigation.

The following features are considered for inclusion when determining the distances between the documents returned during a search for evidence: document text, document creation date and time, document owner and the absolute logical location of the document in the storage medium. Note that we do not propose a fixed set of features because the specific features used in distance calculations would depend on the type of investigation being conducted.

Thus, the distance between two documents in a search result corpus is calculated as:

$$D(j,k) = R(j,k) + |\delta_j - \delta_k| + P(j,k) + O(j,k) \tag{2}$$

where $R(j,k)$ is the rank-difference distance between j and k, δ_i is the creation date and time for i, $P(j,k)$ is the difference in the logical locations of the files, and $O(j,k)$ is the ownership distance. Furthermore, $P(j,k)$ is calculated as the number of directory changes needed to navigate from one location to the other.

The ownership difference $O(i,j)$ relies on a maximum value to create additional distance between the suspect and other users:

$$O(j,k) = \begin{cases} 0 & \text{if } j_o = k_o \\ \tau & \text{if } j_o \neq k_o \end{cases} \tag{3}$$

where τ is a maximum value assigned to $O(i,j)$.

After an investigator issues a query, the profiles of the documents are constructed (or retrieved). The distances between all the documents are then calculated. This information is used in the diversity calculation, which is discussed in the next section.

5.3 Diversity Classification

After a query is executed, the documents in the search results are clustered to determine the diversity of results using a hierarchical clustering scheme. Single-linkage clustering is used. However, the algorithm is only applied once, yielding an intuitive clustering result that provides a good indication of document similarity. This hierarchical clustering scheme also clusters documents very rapidly [11].

For a corpus S consisting of all the documents that are returned in a result:

$$y = argmin_{z \in S} D(x,z) \quad \forall x \in S, x \notin C_i \tag{4}$$

where $D(x,y)$ is a distance measure between x and y. Note that if $y \in C_i$ for some i, then $C_i \cup \{x\}$ else $C_j = \{x,y\}$ with $j = |C| + 1$.

Informally, for each document in the result, the cluster is found that most closely matches it, and the document is added to the cluster. If a document is found that most closely matches the document, but this document is not yet in any cluster, then a new cluster is created with the two documents.

After clustering is complete, the PET computes the diversity of the corpus using Shannon's entropy index [16]:

$$D_S = -\sum_{i}^{C} p_i log_2 p_i \tag{5}$$

where C is a cluster, i is a document in cluster C, and p_i is the probability of choosing document i from cluster C_i. The higher the entropy, the more diverse the corpus. Ideally, the search result would have returned one cluster of documents (i.e., all the documents are closely related).

This ideal search result, representing the lower bound on the diversity of the corpus, is called the "uniformity norm." It is calculated as:

$$N_S = -\sum_{i}^{S} p_i log_2 p_i \tag{6}$$

Ideally, a search result should be on par with ideal uniformity. In other words, the query is focused enough to return only documents that are closely related to each other. If this is not the case, the query should be rejected.

The proposed PET should support and not interfere with legitimate investigative efforts. To assess this, we introduce a sensitivity level indicator l as a threshold for determining if a search result is too diverse. If the diversity indicator $D_S > l$, then the query is defined as being "too wide" and the results are dropped. Note that l is range limited because $N_S \leq l$, which requires an ideal query result.

A query that returns results that are too diverse is referred to as a "wide query." On the other hand, a query that is close to the uniformity norm is referred to as a focused query. Note that all wide queries are reported and no results are returned.

A wide ranging implication with regard to threshold adjustment is that the threshold may be used without due regard for privacy and could be disregarded altogether. Specifically, if the threshold is set too low, then the attempt at preventing a third party privacy breach is disabled and all the results are returned. Cross-checks can serve to avoid this situation by requiring authorization from a higher authority before the threshold is adjusted. The audit logging component discussed in the next section is used to record threshold adjustments.

6. Audit Logging

The audit log records wide queries to ensure that only questions that are related to the investigation are asked and answered. Note that queries are not logged to have a "smoking gun" that the investigator issued unfocused queries, but rather to have documentary proof if the investigative process is questioned in the event of a privacy breach.

The audit log also records changes to the sensitivity level of the results filter, making it possible to trace privacy breaches at a later point in time. Because the queries are recorded and the data and meta-database

are kept intact, privacy breaches can be traced by examining the results obtained from queries logged in the system, similar to the auditing compliance work on Hippocratic databases done by Agrawal, *et al.* [1]. Additionally, the audit log provides a transparent recording of the investigative process.

7. Usability

Any digital forensic tool that casts doubt on the integrity of the evidence that it collects or processes is inappropriate for use in an investigation. The framework described in this paper is intended to be an extension of how an investigator searches for relevant data. It merely augments the analysis phase, attempting to prevent sensitive data about uninvolved third parties from being accessed. The framework does not change any data on the media being inspected, nor does it hinder the investigator or investigation. If the investigator uncovers data pertaining to a third party, the sensitivity level can be adjusted to reveal less data. Alternatively, the filter can be switched off completely, enabling the investigator to access all the data.

8. Conclusions

Preserving the privacy of uninvolved third parties is an important issue during a digital forensic investigation. The possibility of a privacy breach is especially high in multi-user environments where a suspect may have worked on group projects with other innocent users. The privacy protection framework described in this paper is highly effective during the *post mortem* data analysis phase of an investigation. One of its primary features is that an investigator is required to issue focused queries related to an investigation. If a query is deemed to be too wide, the framework prevents the investigator from accessing the search results. Should the investigator have concerns about missing relevant data, a sensitivity threshold can be adjusted to return results even if the query is considered to be unfocused. An audit log is maintained to ensure that queries and their results are saved for compliance purposes if a privacy breach were to occur. The framework operates on the same principle as tools that filter data that are not relevant to investigations. The important difference, however, is that the framework takes privacy concerns into account when filtering query results.

Our future research will evaluate the usability of the privacy protection framework in actual investigations. Also, it will examine the effectiveness of the framework when querying slack space data and file carving results.

References

[1] R. Agrawal, R. Bayardo, C. Faloutsos, J. Kiernan, R. Rantzau and R. Srikant, Auditing compliance with a Hippocratic database, *Proceedings of the Thirtieth International Conference on Very Large Databases*, pp. 516–527, 2004.

[2] N. Beebe and J. Clark, Dealing with terabyte data sets in digital investigations, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 3–16, 2005.

[3] N. Beebe and J. Clark, Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results, *Digital Investigation*, vol. 4(S), pp. S49–S54, 2007.

[4] W. Cavnar and J. Trenkle, N-gram-based text categorization, *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*, pp. 161–175, 1994.

[5] N. Croft and M. Olivier, Sequenced release of privacy-accurate information in a forensic investigation, *Digital Investigation*, vol. 7(1-2), pp. 95–101, 2010.

[6] European Parliament and Council of the European Union, Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, EU Data Protection Directive 95/46/EC, Brussels, Belgium, 1995.

[7] B. Fei, J. Eloff, M. Olivier, H. Tillwick and H. Venter, Using self-organizing maps for anomalous behavior detection in a computer forensic investigation, *Proceedings of the Fifth Annual South African Information Security Conference*, 2005.

[8] B. Fei, J. Eloff, H. Venter and M. Olivier, Exploring forensic data with self-organizing maps, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 113–123, 2005.

[9] S. Fischer-Hubner, *IT Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms*, Springer-Verlag, Berlin-Heidelberg, Germany, 2001.

[10] P. Gladyshev, Formalizing Event Reconstruction in Digital Investigations, Ph.D. Thesis, Department of Computer Science, University College Dublin, Dublin, Ireland, 2004.

[11] S. Johnson, Hierarchical clustering schemes, *Psychometrika*, vol. 32(3), pp. 241–254, 1967.

[12] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin-Heidelberg, Germany, 2001.

[13] Minister of Justice and Constitutional Development, Protection of Personal Information Bill, Pretoria, South Africa (`www.justice.gov.za/legislation/bills/B9-2009_ProtectionOfPersonalInformation.pdf`), 2009.

[14] Organization for Economic Cooperation and Development, OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, Technical Report, Paris, France, 1980.

[15] 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.

[16] C. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, vol. XXVII(3), pp. 379–423, 1948.

[17] Technical Working Group for the Examination of Digital Evidence, Forensic Examination of Digital Evidence: A Guide for Law Enforcement, Technical Report, National Institute of Justice, Washington, DC, 2004.

[18] R. van den Hoven van Genderen, Cyber Crime Investigation and the Protection of Personal Data and Privacy, Discussion Paper, Economic Crime Division, Council of Europe, Strasbourg, France, 2008.

[19] W. van Staden and M. Olivier, On compound purposes and compound reasons for enabling privacy, *Journal of Universal Computer Science*, vol. 17(3), pp 426–450, 2011.

[20] S. Warren and L. Brandeis, The right to privacy, *Harvard Law Review*, vol. IV(5), pp, 193–220, 1890.

[21] I. Witten, Text mining, in *Practical Handbook of Internet Computing*, M. Singh (Ed.), Chapman and Hall/CRC Press, Boca Raton, Florida, pp. 14-1–14.22, 2005.