On Dimensions of Reconstruction in Database Forensics

O.M. Fasan and M.S. Olivier

ICSA, Computer Science, University of Pretoria, South Africa
e-mail: mfasan@cs.up.ac.za, molivier@cs.up.ac.za

Abstract

Although very little amount of research has been done on database forensics, current research has tacitly focused on digital examination and reconstruction of databases from a number of dimensions. The general assumption is that only one of these dimensions needs to be handled during database forensics investigations. This paper analyses the dimensions in which research in database forensics has been focused on and uses these to reveal the different aspects of database forensics which are yet to be explored. The paper also elaborates on the tools and techniques currently being used in database forensics analysis process and highlights some of the challenges being faced in database forensics research and practice as they relate to the dimensions implied by current research in database forensics.

Keywords

Database forensics, Digital examination, Reconstruction, Forensics analysis.

1 Introduction

Database systems have become a core component of many computing systems and are often used to store critical and sensitive information relating to an organization or her clients. Unfortunately, the increase in the usage of databases to store volumes of information together with the increased relevance of the data on many databases in solving crimes have led to an increased number of attacks on databases and interests in investigating databases for artifacts that may assist in solving different crimes.

Database forensics is an emerging branch of digital forensics that deals with the identification, preservation, analysis and presentation of evidence from databases (Fowler, 2008). Although digital forensics has grown over the last decade from a relatively obscure trade-craft to an essential part of many investigations (Garfinkel, 2010), the same cannot be said of database forensics. Despite the large amount of research that has been done on digital forensics, database theory and database security, very little has been done on database forensics (Olivier, 2009) even though investigations involving databases have been explored in theory and practice.

Database forensics research has taken various directions as more researchers explore the field. Even though there is currently no defined underlying model for database forensics, the dimensions in which research has taken are worth exploring. The aim of this paper is to reflect on the research in database forensics and expose the directions in which research has been focused as well as how these directions relate to each other. The paper shows how the research directions form a basis for future research in database forensics and highlights the drawback in the structure of the research directions. The currently available tools for database forensics as they relate to these research directions as well as the techniques that should be employed in database forensics in general are described in the paper. An overview of the challenges being faced in database forensics research and practice are also discussed.

2 Database Forensics

Until recently, traditional digital investigations often excluded databases even though evidence can usually be found in them. Although the field is still in its early years, it is quickly becoming an important part of many investigations due to the increased volume of information that may be helpful in solving different crimes and the large number of risks associated with the information stored on many databases.

Of major importance in database forensics is the ability to retrace the operations performed on a database and reconstruct deleted or compromised information on the database. This requirement affects how data is collected and analyzed during the forensics analysis of a database. Although different aspects of database forensics have been explored by researchers over the past few years, the research works have taken directions that define various dimensions of reconstruction and investigation in database forensics. An inspection of these dimensions shows the research that has been done in relation to each dimension and reflects some of the aspects of database forensics that is yet to receive any research attention.

3 Dimensions of Reconstruction in Database Forensics

The little database forensics research that has been done seems divergent, often focusing on aspects that may only be relevant in a fraction of forensic investigations of databases. We argue that these apparently diverging strands of research are, in fact, related. More than that: we argue that they form different dimensions of a problem space, where most work done until now have focused on a specific dimension of this space. By realising that the individual contributions address a single dimension of a larger space, it becomes possible to focus on the larger problem by considering situations that involve more than one dimension. The following dimensions emerge from the literature:

1. Compromised database.
2. Damaged database.
3. Modified database.

Below, each of these dimensions is considered in more detail and illustrated by positioning previous research (from multiple authors) into a dimension. After that, we proceed to show that scenarios that entail more than one dimension are, in fact, likely and thus substantiate our claim that the database forensics problem is indeed multidimensional. Our conjecture, based on current research publications, is that the problem space will be three dimensional.

3.1 Research on Compromised Databases

We define a compromised database as a database where some of the metadata or some software of the database management system (DBMS) have been modified by an attacker even though the database is still operational. A major concern of database forensics in this situation is that an investigator cannot trust the information provided by the database being investigated. Olivier (2009) pointed out that although a database itself seems to be the best tool for collecting data for forensics analysis, the integrity of the data contained or the results obtained from queries cannot be trusted since the database might have been coerced into giving false information, for example, if the data dictionary has been modified. The problem was also identified by Litchfield (2007d) while discussing steps to perform in a live response to attack on an Oracle database. Beyers et al. (2011) consider four abstract layers of a DBMS and investigate various techniques that can be used for data collection when one or more abstract layers have been compromised. The major decision that has to be made when investigating a compromised database is whether to use the metadata as it occurs on the database being investigated or to try to get a clean copy of the DBMS.

3.2 Research on Damaged/Destroyed Databases

In contrast to compromised databases, the category of damaged or destroyed databases refers to databases where the data contained or other data files may have been modified, deleted or copied from their original locations into other places. These databases may or may no longer be operational depending on the extent of the damage done. Most of the research in database forensics falls into this category.

The series of papers by Litchfield (2007a; 2007b; 2007c; 2007e; 2007f; 2008) describe practical methods for recovering data from an Oracle database using different data sources on the database. Wright (2010) also explained technical methods for identifying when the data on an Oracle database has been modified and how to recognize vulnerabilities in the database. In his book on SQL server forensics, Fowler (2008) discusses the effects which database rootkits can have on the collection and analysis of data from a SQL server database. The book also highlights steps to be followed when database rootkits are detected.

Other research which falls into this category deals with the detection of database tamper, and data hiding in a database. Snodgrass et al. (2004) proposed a technique which relies on cryptographically strong one-way hash functions for detecting when the data on a database has been tampered with. Their idea was extended to deal with the forensics analysis of a data security breach in a subsequent paper by Pavlou and Snodgrass (2006), where the notion of corruption diagrams was introduced as a way of visualizing attacks and forensic analysis algorithms. Another algorithm for detecting tampering and determining what and when a database was tampered with was also presented by Pavlou et al. (2010). Although the work of Stahlberg et al. (2007) and that of Pieterse and Olivier (2012) seem to be anti-forensics, they both expose the areas in a database that should be checked for previously deleted data or any form of information hiding on a database during forensics analysis.

In order to investigate a damaged database, it may be necessary to employ techniques in file carving (Carrier, 2005) in regenerating destroyed files or underlying files in a database which may helpful in retrieving data of interest.

3.3 Research on Modified Databases

We refer to a modified database as a database which has not been compromised or damaged but has undergone changes due to normal business processes since the event of interest occurred. This category of databases is often of interest when a database is not directly involved in the crime being investigated but is used to store information that may assist in solving other crimes. Fasan and Olivier (2012) examine how the information in a database at an earlier time can be reconstructed even though several modifications of the data might have occurred. The authors introduced the notion of relational algebra log, value blocks and inverse relational algebra. The proposed reconstruction algorithm works by traversing the query log and performing series of forward or backward (inverse) queries on relations in the database based on the value blocks. In general, there are various locations where forensic data can be found on a database, especially in compromised or damaged databases. A brief overview of these locations is discussed in section 5.2.

3.4 Orthogonality of the Reconstruction Dimensions

At the moment, research in database forensics treats the dimensions discussed above as being orthogonal. Figure 1 shows the positioning of previous research in database forensics into a single dimension of the database forensics problem space.

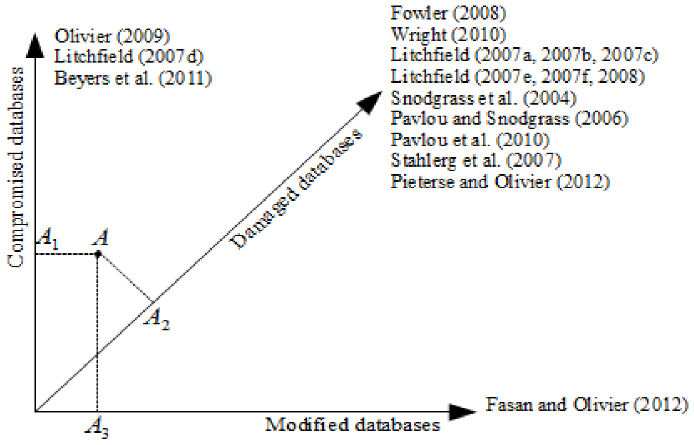

Figure 1: Dimensions of Database Forensics

We posit that a database being investigated may belong to one or more dimensions with varying degrees. For example, an investigation may be positioned at point *A* on this space (figure 1), such that the database being investigated has been compromised at A_1 degrees, damaged at A_2 degrees and modified at A_3 degrees. Thus, it is possible that someone compromises a database by modifying the metadata and then later destroys some of the data on the database in order to hide traces of the compromise. The database forensics problem space is at least three dimensional, with any given investigation positioned somewhere in this three dimensional space. This poses not only a research challenge, but also a challenge in practice: one may need heuristics during an investigation to determine where in this space to start the investigation. Moreover, it also raises questions about whether or not it is possible to quantify the degree of each dimension that occurs when investigating a database and the order to be followed in investigating each possible dimension. In deciding the techniques to be used in any investigation, we propose that questions relating to the trustworthiness of a database, the presence and/or authenticity of the data it contains and modification of the database should first be asked and investigation should be conducted in each dimension while putting into account the degree of that dimension. For example, if we are at the origin of the three dimensions (that is, where the database has not been compromised, damaged or modified) then data can be collected by simply querying the database but the same approach will not be applicable when the database has been compromised and/or damaged. Since there is no defined method for identifying a compromised database, it may be better to assume that we are at the origin of these dimensions when investigating time or mission critical databases and then make educated guesses from possibly contradictory information gathered based on the initial assumption.

The decision regarding the dimensions to be considered in an investigation and extent or degree of each dimension determines the tools that should be used, as well as the process that should be followed during a database forensics investigation.

4 Database Forensics Tools

The unavailability of tools for database forensics is a reflection of the little amount of research that has been done in the field. Database forensics tools should allow information from different sources on a database to be gathered and assist in investigating and recovering of information regardless of the dimensions involved in the investigation. They should also be able to determine when a database has been compromised or damaged. Currently, database forensics is done with tools which are not designed for forensics analysis or with the database system itself, through the query processor. Thus, most of the tools being used are only applicable to specific database management systems and are focused on investigating damaged databases. Quite a number of research are also on-going on the development of new tools.

Although many of the tools being used for database forensics have been helpful for collecting data and identifying transactions that indicate fraudulent activities, most of these tools are not accurate or precise enough to be used as a forensics tool. For example, the Oracle LogMiner was investigated for possible use as a forensics tool by Wright (2005). He reported that there are anomalies with the way it works which makes it inadequate for a forensics analysis. Other tools that have been used include audit features such as the SQL Server Audit and the Oracle Audit.

The database itself is also used as a forensics tool (Olivier, 2009) because it enables an investigator to search the data it contains using powerful queries. However, the database cannot be used as a tool to investigate itself when dealing with a compromised database as one is instantly faced with questions about the integrity of the database and how to ensure that the database has not been coerced into giving false result, for example, if the data dictionary has been modified. Moreover, the query processors on many databases often optimize queries in ways that cannot be controlled by the user. An investigator has to be certain that an optimized query is an exact representation of the original query, especially when dealing with damaged databases where data files on the database might have been modified. These constraints restrict the use of databases as a forensics tool for itself in many cases.

As the field of database forensics develops further, it is hoped that more tools specifically designed for the forensics analysis of databases will be developed. Litchfield (Westervelt, 2007) announced that he is in the process of developing an open source tool called Forensic Examiner’s Database Scalpel (FEDS) for database forensics a few years ago. Some of the major concerns that may be faced in the development of tools include the fact that the data model is typically hardcoded into database management systems and since such model that can be used as a forensics tool for itself does not currently exist, new models will have to be created (Olivier, 2009). Thus, more research is still required in order to develop tools that are specifically designed for database forensics. The development of these tools calls for an understanding of databases as well as the processes involved in database forensics analysis. In the following section, we describe the database forensics process and techniques applicable in various stages of the process.

5 Database Forensics Process

Research in digital forensics has led to the development of various techniques and process models. However, many of these techniques are not completely transferable to database forensics due to certain characteristics of databases which require them to be adapted for handling database forensics. This section discusses some of the techniques in digital forensics as it applies to database forensics analysis.

5.1 Data Acquisition and Preservation in Database Forensics

The acquisition of data from a modified database that has not been damaged or compromised is often done by simply querying the database. However, when investigating a database that has been compromised and/or damaged, there are three data collection methods that can be used: Live acquisition, Dead acquisition or Hybrid acquisition (Fowler, 2008). As with digital forensics in general, a live data acquisition occurs when the system being analyzed is still running while the analysis is being performed. The dead acquisition method involves copying of data from the system being investigated without using the system itself while the hybrid data acquisition method combines the key elements of both live and dead acquisition methods. Regardless of the method used, it is important to ensure that digital evidence is preserved and data is not unintentionally altered or destroyed.

The objective of the preservation phase is to reduce the amount of evidence that may be overwritten. Extreme care must be taken to guarantee that actions performed do not unintentionally alter data. Although data can be acquired from a modified database by querying the database, the execution of any query that can delete the information on the database must be avoided. Furthermore, no SQL statements should be executed in the case of a compromised or damaged database, regardless of the data acquisition method used as this will modify the data stored in memory and on the database data pages. It may also force internal data page split and the storage of new data in the caches, thus complicating the investigation process.

5.2 Collection and Analysis of Artifacts in Database Forensics

Forensic data often exists in several places on a database. It is important to know which data are important and prioritize evidence collection due to the volatility of some data. Apart from the data which can be collected from a modified database by executing queries, data can also be found in tools normally used by database systems, for example, in the execution plan cache and the transaction logs. An execution plan is a documentation of the most efficient way to execute data requests issued by database users and are stored in the plan cache for possible reuse. The plan cache can be used to identify database misuse by an insider or data damage on a database. The transaction logs help to determine previously executed queries on the database and this is often helpful in various investigations.

Other sources of forensics data in databases consist of files storing histories relating to the database. Some of these files are specifically reserved for the database, for example the database log file and data files while others such as the web server logs and the system event logs of an operating system are not explicitly reserved for the database server usage. It is important to consider the level of volatility of a file when deciding which data to collect first. Fowler (2008) gives a summary of the level of volatility of the various areas in which forensic data can be found on a database.

The analysis of collected data depends on the type of data, the specific DBMS and the specific situation being investigated. The analysis stage should take into account the dimensions involved in any particular investigation and where related information can be found. Another important criterion of the analysis phase of database forensics, is that previously deleted of modified data should be recovered and the actions performed by an intruder must be determined, particularly when investigating compromised or damages databases. The following section describes the steps that should be taken during database forensics investigations in general.

5.3 Database Forensics Investigation Process

There is currently no defined process model for database forensics. The available methods are focused on specific DBMSs. Wong and Edwards (2005) presented a patent method for the forensics analysis of an Oracle database. The method consists of generic steps that a forensic investigator may try to follow to discover more information about an operation that was performed on a database (Olivier, 2009).

Another methodology focused on a damaged SQL server database was presented by Fowler (2008). The methodology consists of four major steps: investigation preparedness, incident verification, artifact collection and artifact analysis.

Although the exact steps to be taken during database forensics will depend on the specific situation and DBMS being investigated, we propose that the database forensics investigation process should in general include the following steps:

a) Determining whether a database has been compromised, damaged or modified, or if an investigation involves more than one dimension; b) determining which acquisition method is most applicable; c) collection of volatile artifacts; d) Collection of non-volatile artifacts; e) preservation and authentication of collected data; f) analysis of collected data and determination of the intruder's activities; g) reconstruction of the database. Some of the various challenges involved in carrying out these steps and in database forensics research in general are discuss in section 6.

6 Challenges in Database Forensics

This paper reveals the lack of extensive research on many aspects of database forensics, which can be attributed to the different challenges involved in database forensics investigations. Olivier (2009) identified three dimensions in which a database needs to be considered during a forensics analysis. This inherent multidimensional nature and complexity of databases that is not yet completely understood in a forensics sense is a major contribution to this lack of research.

In addition, there are various challenges involved in data collection and analysis. First, is how to determine whether a database has been compromised, damaged, modified or has to be handled in more than one dimension. There is no heuristic on where to start an investigation in this three dimensional space. Moreover, deciding the degree of the dimensions involved in an investigation and the most appropriate data acquisition method in investigations with more than one dimension is a challenge since no research which offers a guideline has been done. Another challenge often encountered in database forensics is in the large volume of data that can be collected from a database. An examiner must determine which data are pertinent to an investigation in order to reduce the data set. The process of eliminating some data sources poses challenges such as the misinterpretation or over-interpretation of data due to the number of different file formats in various databases, resulting in the potential dismissal of valuable contents (Cohen, 2006). The information contained in some files is also sometimes encrypted.

Lastly, there is need for the development of a formal model for the reconstruction phase of database forensics, that can be applied regardless of the dimensions involved in an investigation. Also, since reconstruction may involve the recovery of data from proprietary formats, adequate vendor support is required (Olivier, 2009). Collaboration between researchers and database vendors is also required in order to have a good knowledge of the proprietary formats in individual database management systems and aid the development of database forensics tools.

7 Conclusion

Database forensics is still a new research area with very little research and few or no tools. The paper discusses the different dimensions in which database forensics research has been tacitly focused. It also shows that even though these dimensions are currently being addressed as being orthogonal, they should be treated as different dimensions of a single problem space. The paper raises some of the questions that need to be addressed when dealing with multiple dimensions in an investigation. An overview of tools, processes and challenges involved in database forensics research and practice, as well as the need for more research and development of tools which are not dependent on a specific DBMS are also discussed in the paper.

As future work, we plan to investigate the process of reconstructing information in compromised and damaged databases. Various scenarios of investigations involving more than one dimension of the database forensics problem space will be examined. The techniques to be employed during such investigations will also be explored.

8 Acknowledgement

This research was supported by the Organization for Women in Science for the Developing World (OWSD).

9 References

Beyers H., Olivier M., and Hancke G. (2011), "Assembling metadata for database forensics". In proceedings of IFIP international conference on digital forensics, pp.89-99.

Carrier, B. (2005), *File system forensic analysis,* Addison-Wesley Professional, Upper Saddle River, NJ, ISBN: 0321268172.

Cohen, F. (2006), "Challenges to digital forensic evidence", Fred Cohen and Associates. http://all.net/Talks/CyberCrimeSummit06.pdf, (Accessed 20 January 2012).

Fasan, O. M. and Olivier, M. S. (2012), "Reconstruction in database forensics", Presented at the IFIP WG 11.9 International Conference on Digital Forensics, South Africa.

Fowler, K. (2008), *SQL server forensic analysis,* Addison-Wesley Professional, Upper Saddle River, NJ, ISBN: 0321533208.

Garfinkel, S. L. (2010), "Digital forensics research: The next 10 years", *Digital Investigation,* Vol. 7, Supplement, pp.S64-S73. The proceedings of the tenth annual DFRWS conference.

Litchfield, D. (2007a), "Oracle forensics part 1: Dissecting the redo logs", NGSSoftware. www.databasesecurity.com/dbsec/dissecting-the-redo-logs.pdf, (Accessed 17 Feb. 2012).

Litchfield, D. (2007b), "Oracle forensics part 2: Locating dropped objects", NGSSoftware. www.databasesecurity.com/dbsec/Locating-Dropped-Objects.pdf, (Accessed 17 Feb. 2012).

Litchfield, D. (2007c), "Oracle forensics part 3: Isolating evidence of attacks against the authentication mechanism", NGSSoftware. www.databasesecurity.com/dbsec/Investigating-Authentication-Attacks.pdf, (Accessed 17 February 2012).

Litchfield, D. (2007d), "Oracle forensics part 4: Live responses", NGSSoftware. www.databasesecurity.com/dbsec/LiveResponse.pdf, (Accessed 17 February 2012).

Litchfield, D. (2007e), "Oracle forensics part 5: Finding evidence of data theft in the absence of auditing", NGSSoftware. http://www.databasesecurity.com/dbsec/ OracleForensicsPt5.pdf, (Accessed 17 February 2012).

Litchfield, D. (2007f), "Oracle forensics part 6: Examining undo segments, flashbacks and the Oracle recycle bin", NGSSoftware. www.databasesecurity.com/dbsec/oracle-forensics-6.pdf, (Accessed 17 February 2012).

Litchfield, D. (2008), "Oracle forensics part 7: Using the Oracle system change number in forensic investigations", NGSSoftware. www.databasesecurity.com/dbsec/oracle-forensics-scns.pdf, (Accessed 17 February 2012).

Olivier, M. S. (2009), "On metadata context in database forensics", *Digital Investigations*, Vol. 5, No. 3-4, pp.115-123.

Pavlou, K. and Snodgrass, R. T. (2006), "Forensic analysis of database tampering". In proceedings of the 2006 ACM SIGMOD conference on management of data, pp.109-120.

Pavlou, K. and Snodgrass, R. T. (2010), "The tiled bitmap forensic analysis algorithm", *IEEE Transaction on Knowledge and Data Engineering,* Vol. 22, pp.590-601.

Pieterse, H. and Olivier, M. S. (2012), "Application of data hiding techniques in a database environment", Presented at the IFIP WG 11.9 inter. Conf. on digital forensics, South Africa.

Snodgrass, R. T., Yao, S. S. and Collberg, C. (2004), "Tamper detection in audit logs". In proceedings of the 30th inter. conf. on very large databases - volume 30, pp.504-515.

Stahlberg, P., Miklau, G. and Levine, B. N. (2007), "Threats to privacy in the forensic analysis of database systems". In proceedings of the 2007 ACM SIGMOD international conference on management of data, pp.91–102.

Westervelt, R. (2007), "Black hat 2007: New database forensics tool could aid data breach cases", http://searchsecurity.techtarget.com/news/article/0,289142,sid14_ gci1266525,00.html. TechTarget, (Accessed 7 January 2012).

Wong, D. M. and Edwards, K. B. (2005), "System and method for investigating a data operation performed on a database", Number 2005028918, United States patent publication.

Wright, P. M. (2005), "Oracle database forensics using LogMiner", NGSSoftware, www.databasesecurity.com/dbsec/OracleForensicsUsingLogminer.pdf (Accessed 5 February 2012).

Wright, P. M. (2010), *Oracle Forensics: Oracle Security Best Practices*, Rampant Techpress, Kittrell, NC, ISBN: 0977671526.