*2009 International Conference on Computer Engineering and Applications*
*IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore*

# Efficient Model for Detection Data and Data Scheme Tempering with Purpose of Valid Forensic Analysis

Jasmin Azemović $^{1+}$, Denis Mušić $^{2}$

[1] Faculty of Information Technologies, Univesity "Džemal Bijedić"
Mostar, Bosnina and Herzegovina, jasmin@fit.ba
[2] Faculty of Information Technologies, Univesity "Džemal Bijedić"
Mostar, Bosnina and Herzegovina, denis@fit.ba

**Abstract.** Accelerating development of information technologies also causes the growth of computer crime which has the main goal to profit illegally, industrial espionage, forgery and falsification of data. Therefore, a demand for data security is needed more than ever. This paper describes a model which ensures effective detection of authorized and unauthorized modification of database scheme and data itself. At the moment when data modification becomes a criminal act, this model represents a foundation for forensic analysis of collected digital evidence. A research described in the paper has been applied on access control methods which use SQL triggers. Special emphasis is laid on detection of forgery over data which were created during the process of access control using hashing methods.

**Keywords:** data tampering, forensic analysis, database, computer crime.

## 1. Introduction

Modern information technologies are involved in every aspect of human life. That fact is forcing us to look at computer crime from different angle.

Business process produces big amount of data in government agencies, universities and enterprises on daily basics. Therefore a secure environment for storing a data is imperative. Cases in which data where maliciously modified (data tampering, data fraud and unauthorized data gathering) can produce serious and long term consequences. Data tampering can be done with unauthorized access and in some cases through authorized users. Results of that action can be unpleasant for business and their clients [1].

There are cases when malicious data modifications result with police actions and involving justice. Finale goal of investigation process is reviling identity of crime committer because those details can be very important. First steps in investigation of committed crime is collecting and analyzing evidence. Products of that process are facts that can help in solving a crime. Looking from aspect of "classic" crime actions, that process is based on forensics.

Computer crime is also a crime, but with different consequences. In those cases we use digital or computer forensic to collect and scientifically examine information systems in all aspects. Outcome of that process is to determine details about digital criminal activity. Digital forensic is the most important part of investigation process.

That is because collected facts need to be presented in court of law. Process of collecting, analyzing and preserving digital data is based on scientific methods. Only that evidence is valid. Figure 1 illustrates phases of digital forensics. Digital evidence is defined like any data stored on computer or transmitted over network that can prove (support) or drop theory about who has commit a digital crime [2].

[+] Corresponding author. Tel.: + (387 61 741 556); fax: +(387 36 570 730).
*E-mail address*: (jasmin@fit.ba).

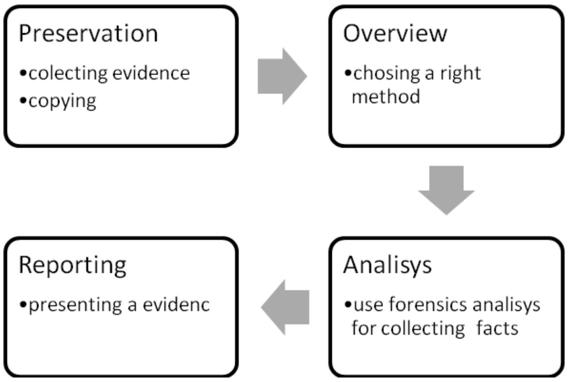


Fig. 1: Phases of digital forensics

Digital evidence collected in investigation process needs to satisfy following elements:

- Admissible
- Authentic
- Complete
- Reliable
- Believable

## 1.1. Examples of data tampering

There are many reasons why someone should attempt to use unauthorized access and make malicious data modifications. Good example is HIS (Hospital Information System). After examining a patient, a physician gives a diagnosis and prescribes therapy with or without using medications. Mistakes in that process can produce serious consequences and even death of a patient. In both cases medical personnel is highly “motivated” to falsify data in information system.

Let’s now look at this problem from database level. Investigation process and forensic analysis in this segment is extremely difficult. To get a clearer picture, we need to examine this hypothetical situation. There was a data breach in a bank information system. Unauthorized access to clients banking accounts is resulting with money problems to a certain number of clients. Somehow, they have a minus on credit card accounts. Bank personnel are unable to do anything with solving a case on the internal level. Clients decide to press charges. Persecution department send a team of digital forensics. They have a task to collect all evidence about suspicious transactions, examine them and make a report to the court of low. The team needs to follow a precisely defined procedure to provide valid court evidence (Fig 1). But in early phase of investigation the team has confronted with very serious problems. The first phase is collecting data and copying. In a case of bank information system those tasks are very difficult to realize. Here are some key points:

- Bank information system is based on distributed database architecture
- Size of database is over 1 TB
- To copy all disks with images of data, system needs to be stopped
- Bank policy does not allow stopping a system
- Back up media are not enough, because team those not know a time frame of CE (corruption event)

The corruption event (CE) could be due to an intrusion, some kind of human intervention, a bug in the software (DBMS or the file system or somewhere in the operating system), or a hardware failure, either in the processor or on the disk. There exists a one-toone correspondence between a CE and its corruption time (tc), which is the actual time instant (in seconds) at which a CE has occurred. [1].

The previous example shows us that it is not possible to collect valid digital evidence and in same time keep business continuity of bank. Section 5 (Data Breaches Analysis) of this paper goes in to deeper details and statistical analysis.

## 2. Audit Logging

It is a well known fact that modern information society has a strong need to provide secure data storage environment. Keep in mind that a single change in database can make a criminal from an ordinary citizen. This research is motivated by that fact.

There is only one direction to provide valid forensic analysis of database. That process requires creating an audit logging in all aspects of information system. Based on that, data investigation process can reconstruct what has really happened. Depends on application architecture (web, desktop, combined) there is a wide range of data that is necessary to be collected. For example: user name, IP address, time stamp etc. But not all data is required in each situation. IP address of internet provider is not needed in classical desktop environment. Vice versa situation can be crucial in determining geographical location of criminal. Basic task of audit logging is to give answers to three questions (Fig 2)

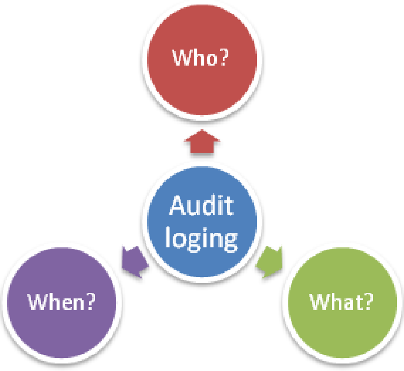


Fig. 2: Tasks of audit logging

Depending on how much details are necessary about changed data; audit logging can be divided in two groups: simple and advanced access control.

Simple access control collects basic data about actions in information systems. Basically those are very simple implementations and we are not going to describe in more details. Simple access control gives answers to “Who” and “When” questions. Key point is that simple access control is not enough to provide data for valid forensic analysis. Bottom line is that every system requires, at least this access control.

On the other hand there are systems where integrity, data precisions, privacy and security are on the first place. They are ideal candidates for advances access control. In this environments “Who” and “When” are not enough. The biggest issue is the question “What”. Table 1 show a list of activities for implementing advanced access control.

| Who | When | What |
|---|---|---|
| Identity | Log in time | Data about data modifications |
| Location | Log out time | Data about data scheme modifications |
| Operating system and application | Timestamp | Hash checksum |

Table 1: Advanced access control elements

### 2.1. Who

The first and basic element in this category is identity of user. Every information system has its own user implementation. But all of them are based on real user identity. No matter what is the tool for accessing data, application or direct access to database layer, user account is required. autentification and authorization are the first clues of "who".

The second element is a location of user. The location is important no matter whatever user access the system from web or desktop environment. Digital fingerprint of user is needed, but also digital signature of software has the same importance (operating system, web browser, telnet etc.).

## 2.2. When

No data collected in investigation process are valid without time dimension. From the aspect of database, time frames of user's activity can provide or deny user alibi. Date and time of login, logout and the period of activity are required in this segment of audit logging

## 2.3. What

This component is the major difference between simple and advanced access control. It plays the key role in this process. It can contain information like: tampered data (old data), modified (new data) and validation element. Validation element is a security mechanism for providing extra layer of protection from users with high access rights (administrators, DBA, power users etc.). If we are looking from digital forensic point of view this is the key element. Without this, the forensic analysis is irrelevant.

# 3. Data Tamper Detection Model

We propose data tamper detection model whose task is to successfully provide elements for collecting and securing validity of collected data. (Fig 3).

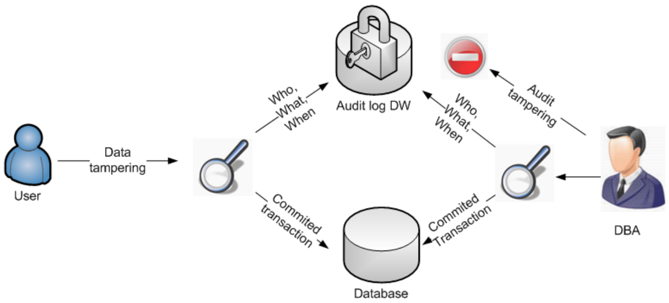


Fig. 3: Tamper detection model

a) A user makes an unauthorized or authorized data tampering

b) Information about that activity are collected and stored in data warehouse (DW)

c) Validation components secure collected audit logs

Let's examine this model in more details

a) Any data modification in the background of information system is a collection of SQL statements (INSERT, UPDATE, DELETE). User is not aware of this process. Application layer through business logic calls stored SQL procedures. From user's perspective, those can be Save, Delete or Add command in specific application.

b) Each transaction passes through layer for detecting data modification. This layer is positioned in the last segment of information system. That is a database itself. Any other position can be manipulated or compromised. Detection layer is implemented with SQL triggers.

Triggers are programming routines and special implementation of SQL code. They are executed automatically, based on a defined event. An event can be local (database level) or global (server level). Existing RDBMS (Relational Database Management System) can operate with DML (Data Manipulation Language) and DDL (Data Definition Language) triggers. DML can be executed on any data modification event (INSERT, UPDATE, DELETE) on database tables. Some of interesting DML triggers are AFTER and INSTEAD OF groups. From the aspect of audit logging and "What" component, AFTER triggers are used in our model. Special group of triggers are DDL. They can be executed on specific events when someone tries to change or create some database object (CREATE, ALTER and DROP statements). It is very rear that a user application can modify database structure. But what if the administrator tries to destroy evidence of certain activity? DDL trigger can provide evidence of that activity.

c) During storing audit logs, this model implements an extra layer of security using hash cryptographic method. Every new record is passes through this layer. So, standard (Who, When and What) are secured with hashing function. During INSERT operation into Audi table, trigger evaluates two hash values and stores with every record. Figure 4 describes this algorithm in more details [6].

For this purpose, we have added two special columns called HReserved (Horizontal Reserved) and VReserved (Vertical Reserved).

Whenever there is an insert operation in the Audit table, we need to calculate two hash values - a horizontal or a row hash, and a vertical or a column hash. The row or the horizontal hash is stored in the HReserved column and contains a hash value of all the columns in the row except HReserved and VReserved column values. Any change or a modification in any given rows will result in a mismatch of the hash value and therefore can be detected.

The column or the vertical hash is stored in the VReserved column and contains a hash value based on the HReserved values of the last two rows as well as the current row. This interwoven hashing mechanism will ensure that if one particular row is deleted from the Audit table, the detection algorithm can find a mismatch by the existence of other two rows immediately preceding the deleted row.

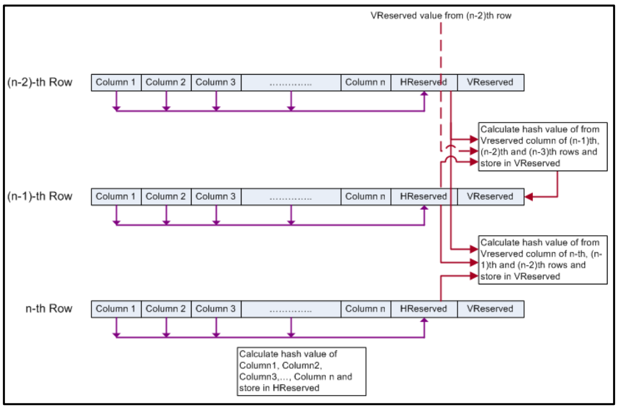

VReserved value from (n-2)th row
(n-2)-th Row
Column 1
Column 2
Column 3
.............
Column n
HReserved
VReserved
Calculate hash value of from Vreserved column of (n-1)th, (n-2)th and (n-3)th rows and store in VReserved
(n-1)-th Row
Column 1
Column 2
Column 3
.............
Column n
HReserved
VReserved
Calculate hash value of from Vreserved column of n-th, (n-1)th and (n-2)th rows and store in VReserved
n-th Row
Column 1
Column 2
Column 3
.............
Column n
HReserved
VReserved
Calculate hash value of Column1, Column2, Column3,..., Column n and store in HReserved

Fig. 4: Algorithm for protection audit logs

## 4. Future and related research

Our model is successful in the area of detecting malicious and none-malicious data tampering. Also we successfully protect audit logs from data falsification or data destroying. However, to cover all aspects of forensic analysis this model can be improved in the following areas.

1. Proposed model can be improved with intention detection. Intention is a serious issue in classic forensic analysis [7]. In digital environment, especially database, it is very hard to say that someone has tried or will try to make some data modification. For example: what if user makes a query or run report to collect all social security numbers of users in database? From the access control perspective, nothing suspicious has happened. But that can be the first step in selling social security numbers on the black market from where can be used by identity thefts. Very good model that can be applied is MalDViP (Malafide Intention based Detection of violation in Privacy) [7] and CBR (Case Based Reasoning).
2. Privacy is becoming an important feature in modern society. The rapid advances in information technology and the emergence of privacy-invasive technologies have made informational privacy a critical area to be protected. Privacy-enhancing technology aims at making privacy protection guidelines and laws an integrated part of the technology. Thus, an information system is designed to embed components that allow monitoring compliance of the system to privacy rules. [9] We propose

to define the intended purpose of personal information as a chain of acts on this type of information. It should include audit logging and forensic analysis.

3. Performance can be a big issue because overhead of audit logging enforcement. Very important question is: What is the overhead cost introduced by audit logging enforcement? Intercepting every query execution in system with thousands of transaction per second is affecting system performance. [10]. Real life experiment should answer this question

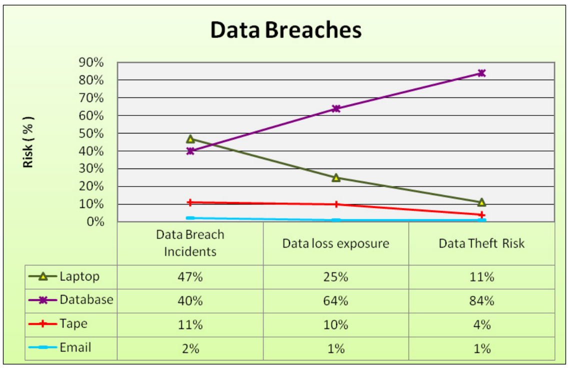
## 5. Data breaches analysis

We will show statistical analysis of data breaches that have been reported. Most breaches was tampering the personal information includes data elements useful to identity thieves, such as Social Security numbers, account numbers, and driver's license numbers. Some breaches, that do not expose such sensitive information, have been included in order to underscore the variety and frequency of data breaches. However, we have not included the number of records involved in such breaches in total because we want this compilation to reflect breaches that expose individuals to identity thefts as well as breaches that qualify for disclosure under state laws. The breaches posted below include only those reported in the United States. They do not include incidents in other countries.

The running total, we maintain at the end of the Chronology, represents the approximate number of records that have been compromised due to security breaches, not necessarily the number of individuals affected. Breaches are collected in period from 2005 to beginning of 2009. Some individuals may be the victims of more than one breach, which would affect the totals.  In reality, the number given below should be much larger. For many of the breaches listed, the number of records is unknown. All the information about specific data breaches can be found [11].

Open security foundation includes a search engine and news articles for the breaches listed below, and also provides database of its data breach records. It is a flat comma-separated value file or XML file that can be imported into a database or spreadsheet program for our own data analysis [12].

First we need to answer a simple question – when enterprises lose data, how do they lose it? To answer this, I took the raw statistics compiled by Privacy Rights Clearinghouse. We assume that data loss could fall into four categories: (Email - someone emailed lots of sensitive data in an unauthorized way from the inside of the enterprise, Tape – a tape with sensitive data got lost, Laptop – someone lost a laptop that had sensitive data, Database – someone accessed into a data server).



| | Data Breach Incidents | Data loss exposure | Data Theft Risk |
|---|---|---|---|
| Laptop | 47% | 25% | 11% |
| Database | 40% | 64% | 84% |
| Tape | 11% | 10% | 4% |
| Email | 2% | 1% | 1% |

Fig. 5: Statistical analysis of data breaches

## 6. Conclusion

Motivated by audit log requirements, we have presented an efficient approach to providing audit logs for transaction processing systems that can effectively and efficiently detect tampering. We based our approach on existing SQL coding and cryptographic techniques such as strong cryptographic hashing to provide authentification codes on collected data. Our contributions are as follows:

- **We proved** that standard digital forensic procedure are not efficient or are very hard to apply on database systems, depending on information systems and business environment.
- **We defined** three key elements. Without them access control will be not efficient enough.
- **We create** tamper detection model based on SQL triggers with purpose of collecting audit logs. It uses hashing algorithm to improve data integrity and prevent audit falsification.
- **We looked** at current researches in this area and found common elements.
- **We defined** future directions in this area. Every direction one can improve tamper detection model with the following elements: intention detection, elements of privacy over collected data and a performance issue of applying this model on big business environments. This can be measured on isolated environment applying statistical and comparation methods.

The final goal of this and future researches is to produce and evaluate enhanced tamper detection models with practical implantations that can be used in commercial database systems.

## 7. References

[1] Melinda JoyMalmgren, An Infrastructure fot database tamper detection and forensic analysis, The University of Arizona, 2007.

[2] Chisum J.W. Crime Reconstruction and Evidence Dynamics, Presented at the Academy of Behavioral Profiling Annual Meeting, Montrey, CA, 1999

[3] Anthony Reyes, Cyber Crime Investigation, Syngress Publishing, 2007

[4] Kyriacos Pavlou and Richard T. Snodgrass, Forensic Analysis of Database Tampering, from Department of Computer Science, University of Arizona, 2005

[5] Itzik Ben-Gan, Dejan Sarka and Roger Wolter, Inside Microsoft SQL Server 2005: T-SQL Programming, Microsoft Press 2006

[6] Amit Basu, Article on Forensic Tamper Detection is SQL Server Tables, http://www.sqlsecurity.com/

[7] T. Spyrou, J. Darzentas, Intention Modeling: pproximating Computer User Intentions for Detection and Prediction of Intrusions, 1996

[8] Shyam K. Gupta1, VikramGoyal, and Anand Gupta, Malafide Intension Based Detection of Privacy Violation in Information System, Bagchi and V. Atluri (Eds.): ICISS 2006, LNCS 4332, pp. 365–368, 2006.cSpringer-Verlag Berlin Heidelberg 2006

[9] Sabah S. Al-Fedaghi, Beyond Purpose-Based Privacy Access Control, Computer Engineering Department Kuwait University, 2007

[10] Kristen LeFevrey, Rakesh Agrawaly, Vuk Ercegovac, Raghu Ramakrishnan, Yirong Xuy David DeWitt, Limiting Disclosure in Hippocratic Databases, Proceedings of the 30th VLDB Conference, Toronto Canada, 2004

[11] Privacy Rights Clearinghouse, 3100 - 5th Ave., Suite B, San Diego, CA 92103. Web: www.privacyrights.org

[12] Open security foundation, http://datalossdb.org/download