Jurnal Teknologi

Full Paper

A GENERIC DATABASE FORENSIC INVESTIGATION PROCESS MODEL

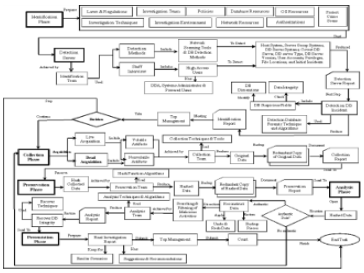
Arafat Al-Dhaqm*, Shukor Abd Razak, Siti Hajar Othman, Asri Nagdi, Abdulalem Ali

Department of Computer Science, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Malaysia

Article history
Received
1 February 2015
Received in revised form
24 March 2015
Accepted
1 August 2015

*Corresponding author
arafataldoqm@gmail.com

Graphical abstract



Abstract

Database Forensic investigation is a domain which deals with database contents and their metadata to reveal malicious activities on database systems. Even though it is still new, but due to the overwhelming challenges and issues in the domain, this makes database forensic become a fast growing and much sought after research area. Based on observations made, we found that database forensic suffers from having a common standard which could unify knowledge of the domain. Therefore, through this paper, we present the use of Design Science Research (DSR) as a research methodology to develop a Generic Database Forensic Investigation Process Model (DBFIPM). From the creation of DBFIPM, five common forensic investigation processes have been proposed namely, the i) identification, ii) collection, iii) preservation, iv) analysis and v) presentation process. From the DBFIPM, it allows the reconciliation of concepts and terminologies of all common databases forensic investigation processes. Thus, this will potentially facilitate the sharing of knowledge on database forensic investigation among domain stakeholders.

Keywords: Database forensic, investigation process, digital forensic

1.0 INTRODUCTION

Database threats and crimes are growing day by day and are harming confidentiality, integrity and availability of database systems. Traditional database security mechanisms such as authentication, authorization and control access cannot be persistent and defend alone the growth of database threats and cybercrimes by itself [1]. Thus a database forensic field is required to identify and collect evidences against those crimes for analyzing and documenting the events that caused those crimes and reveal databases that are tampered [2].

Database forensic is a branch of digital forensic that deals with database contents, metadata, log files, data files, and memory data in order to create a timeline, relationship or recover relevant data [3]. Database forensic (DBF) research still frequently reflects on the reasons why this is the case; in fact, only a few years ago hardly any scientific research existed about database forensics despite the realization that such work was urgently needed [4]. One possible reason for the lack of researches is the inherent complexity of a database management system (DBMS) when compared to, file systems [5]. While files are often abstracted as streams of bytes, a database is a collection of data where data elements are related to one another. The process or procedure that is used with digital investigation will directly affect the results of the examination. Selecting the unfitting investigative processes may lead to damage or lost evidences [6]. Avoiding one phase or changing any of the phases may lead to indecisive consequences; and give invalid conclusions. "Evidences captured in an ad hoc or unstructured manner may risk not being admissible in the court of law".

Several investigation models have been proposed by various authors over the years especially in digital forensic. Through our observation and analysis, many trends and features have been introduced by these models. However some of them focused on specific scenarios while the others focused on generic scenarios. Some of them have relative details while the others seem to be general. It could be a bit challenging or even unclear particularly to the newer forensic investigator to adopt the accurate or proper investigation model.

Due to the lack of generic database forensic investigation process model [4], the main objective of this study is to provide an obvious structure which is called Database Forensic Process Investigation Process Model (DBFIPM) to unify, facilitate, and share database forensic investigation process knowledge amongst database users and practitioners. This model provides a pure and specific database forensic concepts and terminologies which are used in the database forensic investigation. Unifying these concepts in one conceptual model will increase knowledge of users, newcomers and practitioners. Additionally, it will reduce the complexity and ambiguity of the investigation. The paper is organized as follows: Section II highlights the database forensic challenges and issues

In Section III we reviewed the existing investigation process models. Section IV displays a methodology which is used to propose a common phase of the investigation process. The results and discussion are displayed in Section V. The conclusion and future work are discussed in Section VI.

2.0 DATABASE FORENSIC CHALLENGES AND ISSUES

The Database Forensic field has been suffering from several challenges and issues, which make it heterogeneous, complicated and ambiguous amongst researchers, investigators and organizations. A variety of database infrastructure, multidimensional nature of database systems, several database forensic artifacts and lack of Database Forensic knowledge management are considered the main challenges and issues of Database Forensic field [2, 3, 7-9]. Additionally, Database Forensic knowledge is scattered everywhere on the internet, in books, dissertations, organizations, journals, chapters and online databases.

Database systems have a variety of infrastructures and services which totally dissimilar from one database system to another [1], Oracle database system has specific infrastructure and services which are dissimilar of other database systems such MS SQL Server, MySQL Server, and DB2. Physically, it consists of three kinds of physical operating files such as data files, log files, and control files. Thus each of the database systems has its own specific infrastructure and service. On the other hand, and based on the ANSI/SPARK model, the database system was divided into four layers known as the data model, data dictionary, and application schema and application data [9]. Accordingly, the specific database varies in forensic artifacts such as methods, models, frameworks, tools, activities, policies and so on. In the same context of Database Forensic issues, it has been suffering in terms of the multidimensional in nature of database systems.

A database system is multi-dimensional nature by default which consists of three levels such as the internal level, conceptual level, and external level from the bottom to top [4]. The internal level contains a physical operating file which considers the main dynamo of a database system. The conceptual level is the logical level which represents the logical infrastructure of database schema such as users, tables, indexes, procedures etc. the external level represents the GUI which deals with real users to facilitate manipulating data. Therefore, the various dimensions of database systems have been affected considerably on the Database Forensic. Consequently, it has been classified into three dimensions namely the destroyed dimension, compromised dimension and changed dimension [4]. Accordingly, investigators cannot determine which dimension has been tampered with and in which dimension the investigation will be conducted. Additionally, this variety of dimensions has produced several database forensic artifacts which are dealing with specific database systems incidents.

Subsequently, several of database forensic artifacts have been produced and spread everywhere. Examiners/investigators have utilized it to discover database incidents and reveal who is tampering with them? When did tampering happen? What did the data tamper with? And also why and how did the tampering happen. For example of Database Forensic artifacts, investigation process models, frameworks, algorithms, methods, tools, activities, techniques, policies, procedures, transactions and laws.

Specific Database Forensic investigation process models have been introduced to dealing with specific database systems. Oracle database system has its own process model which used to reveal database malicious activities [10]. However, it concentrated on reveal SQL injection attacks in database systems. Similarly, MS SQL Server has its own process methodology which consists of four process phases namely *investigation preparedness, incident verification, artifact collection and artifacts analysis* to discover database incidents [11].

Correspondingly, the MySQL database Server used a framework to reveal incidents, which developed by [12]. Nevertheless, the specific DBF investigation process models which are discussed have sharing processes, and terminologies. Arguably, digital forensic investigation process models may not be suitable for Database Forensic investigation due to several perspectives such as Database Forensic challenges and issues [13].

Several digital forensic investigation process models have been developed to discover digital crimes of computers, networks, mobiles and internet. However, the exiting models may not be suitable to deal with database incidents due to several perspectives as mentioned in the previous paragraph. Furthermore, digital forensic practices do not reflect characteristics of IT governance related to transactional databases [13]. Moreover, traditional digital forensic practices may not be appropriate to analyze a large volume of data stored in database servers [14]. Digital Forensics is not suitable for database systems mainly due to a concentration on identification, collection, handling, storage, incident response and training. Additionally, database events may be difficult to trace, unless mutual aid among digital investigations have to be added with database analysis [13]. Traditional digital investigations require shutting down of the target system down, and the physical removal of hard drive for its later imaging [15]. Assuming that target system is a database server, removing and imaging its hard drive is counterproductive because production servers are huge data storages which cannot be shut down, and their drives may be far larger than any external drive in the market, therefore it is impossible to image into a standard external drive [16]. Consequently, imaging just relevant evidence, like important transactional logs related to suspicious transactions, should be considered. In order to do this, "database log environment" has to be controlled for producing specific logs to record suspicious transactions, using simple SQL procedures to support investigation.

In summary of this Section, DBF domain has suffered several issues which make it complicated and heterogeneous amongst researchers and investigators. Nevertheless, and despite of heterogeneous and variety of forensic artifacts of Database Forensic, it has numerous sharing concepts and characteristics which have been highlighted in this study such as forensic dimensions, database layers, policies, procedures, investigators, identification, collection, log files, transactions, SQL statements, and incidents, which need to be harmonized and unified using metamodeling approach to facilitate investigation task. Therefore this study highlighted two main issues. The first issue, lacks of common investigation process model of Database Forensic which is the main purpose of this study, and the Second issue lack of abstract metamodel to structuring and managing whole Database Forensic Knowledge. The Second issue will be the future work of this study

3.0 INVESTIGATION PROCESS MODEL REVIEWED

The digital investigations models that are proposed and suggested in digital forensic domains are not small; therefore it would be quite difficult to review all of them. We reviewed the models based on chronological order, at least one model is proposed per year. Our reviewing and selection are not based on the features or properties of the models that determine that the model is better or greater than the other models. Our objective is to recognize and select the investigation phases from investigation models rather than choosing which model is the best.

Table 1 displays 21 digital forensic investigation process models which have been identified and collected.

Table 1 Digital Forensic Investigation Process Models

Model	Investigation Process Phases	References
M1	Acquisition, Identification, Evaluation, Admission.	][17
M2	Acquiring, Authenticating ,analysis	[18]
M3	Identification, Preservation, Collection, Examination, Analysis, Presentation, Decision	[19]
M4	Identification, Preparation, Approach strategy, Preservation, Collection, Examination, Analysis, Presentation, Returning evidences	[20]
M5	Readiness, Deployment, Physical investigation, Digital investigation, Review	[21]
M6	Readiness , Deployment , Traceback, Dynamite, Review	[22]
M7	Awareness, Authorization, Planning, Notification, Search, Collection, Transportation, Storage, Examination, Hypothesis, Presentation, Proof/ Defense, Dissemination.	[23]
M8	Readiness, Development, Physical crime scene investigation, Digital Crime Sense investigation, Presentation.	[24]
M9	Preparation, Investigation, Presentation	[25]
M10	Pre-Analysis, Analysis, Post-analysis	[26]
M11	Planning, Identification, Reconnaissance, Transport & Storage, Analysis, Proof & Defense, Archive Storage.	[27]
M12	Preparation, Incident, Incident response, Digital forensic investigation & Physical investigation, Presentation.	[28]
M13	Suspend database Operation, Collection Data, Preservation Data, Analysis, Reconstruct the database, Restore Database Integrity	[10]
M14	Setup the evidence collection server, Perform general steps to get basic information, Collection, analysis, notify	[29]
M15	Investigation preparation, Incident verification, artifact collection, preservation, artifact analysis, Report.	[11]
M16	Detection server, Data collection, Investigation of data collection	[30]
M17	Identification, Artifact collection, Preservation, Reconstruction, Artifact analysis, Report	[12]
M 18	Collection	[31]
M 19	Collection , preservation	[32]
M 20	Detection Covert System.	[33]
M 21	Incident reporting, Examination	[34]

Model	Investigation Process Phases	References
	Preparation, Physical Examination, Digital Examination, Documentation and Presentation, Post examination, Post Examination Analysis	

4.0 METHODOLOGY

This study utilizes Design Science Research (DSR) methodology towards propose process artifacts which is called DBFIPM [35, 36]. Design Science Research (DSR) is defined as research methodology that is used to create new and persistent artifacts for a special problem domain [35]. DSR concentrates on IT artifact with a high importance on significance in the application domain. According to [37] creation of DSR can be illustrated into four kinds of artifacts which include: constructs that organize the language to identify problems and solutions, models that use this language to describe problems and solutions, methods that define processes that offer assistance on how to answer problems and instantiations which are defined as combinations of constructs, models, and methods. DSR cycle includes a manner of assessment and repetition against produced artifacts. It obviously insists on the building and assessment of the artifact to be completely performed before the artifact is offered to users. According to [36] design science process includes six steps:

4,1 Identify and Collect Models

Twenty one generic and specific digital process models have been identified, reviewed and collected, towards propose DBFIPM. Table 1 displays these models.

4.2 Extract Investigation Process Phases and Candidate Common Process Phases

Fifty nine investigation process phases extracted from 21 digital process models. Table1 shows investigation process phases which are displayed in the investigation process phase's column.

Frequency Based Selection (FBS) is a feature selection technique that evaluates the importance of individual processes in the model developed [38]. It is based on the idea that the best model is formed using the most common processes by performing Frequency-based Selection, processes that do not have correlations (or a need) to the classification are removed from the developed model. Thus 31 process phases have selected over 59 processes towards developing DBFIPM. Table 2 displays the comprehensive analysis of digital forensic investigation process models.

Using the process frequency, an important value for each process in the developed model is estimated and expressed as the 'Degree of Confidence (DoC)'. This value designates the expected probability that the developed model process is used in a randomly chosen DBF and digital forensic process models. The DoC is derived by dividing 'the frequency of how many times a process appears in all the investigated models' with 'the total number of models'. For this purpose, DoC is based on the list of processes that appear in the proposed developed model and is defined as follows:

$$DoC = \frac{\text{Frequency of Process}}{\text{Total Models}} \times 100\% \quad (1)$$

The following five categories of processes based on their DoC are follows:

1. Very Strong (100 – 70 %),
2. Strong (69 – 50 %),
3. Moderate (49 – 30 %),
4. Mild (29 – 11 %), and
5. Very Mild (10 – 0 %).

Figure 1 illustrates the candidate common process phases which consist of the DBFIPM.

4.3 Allocate Synonyms Investigation Process with the Fitting Common Processes

After designating the common processes phases using FBS in Step 2, the synonyms investigation processes which have similar activities are allocated among common phases based on their functionalities and activities. Table 2 shows the common process phases and their synonyms.

4.4 Assign Obvious Definition of Each Which is of the Candidate Common Process

Several definitions have been assigned to each which is of the common process phases based on the aim and functionality of the process. However, in this study, the authors will choose and adapt the best definition which fits with the Database Forensic Investigation. Section V will explain this step.

4.5 Determine Specific Concepts Which are Related to Candidate Common Process Phases

Practically, the concepts and terminologies which form the common process phases are determined in this step to give more clarification of the proposed model nature. Figure 2 shows the conceptual generic database forensic investigation process model. The concepts which are related with common concepts are shown in Figure 2.

4.6 Identify Conceptual Relationships Among Common Process Phases and Their Concepts

The conceptual relationships among common process phases and their concepts allow investigators and users to determine the boundaries

of concepts and their dependencies, to develop their own models from the main conceptual model.

5.0 RESULTS AND DISCUSSION

Forensic investigation models that have a wider and specific coverage of database forensic domain are identified, collected, reviewed and listed in Table 1. Comprehensive analysis of these models and their process phases are shown in Table 2. Fifty nine investigation process phases have extracted, reviewed and compared towards candidate common investigation process phases. Nevertheless, 31 process phases have been selected to candidate the more frequent and suitable processes for the Database Forensic investigation domain. Thus, five common investigation process phases have been selected based on their frequency and repeating in process models using FBS method which is mentioned in Section IV.

In this study, the investigation process phases are divided into two parts: *Pure processes* and *Synonym processes*. Pure processes are the processes that have perfect and clear names such as *identification, collection, analysis, document, preparation, and presentation*, whereas the Synonyms processes are the processes that have alternative names of the pure names for example *acquisition, search and identify evidence, and reconnaissance* is the synonym names of the collection process phase.

Consequently, the five common investigation process phases which have been selected are: Identification, Collection, Preservation, Analysis and Presentation phase. Table 2 shows five colors which represent these phases. Hence, Red color represents the pure process phase which is called Identification Process Phase including its synonym processes such as incident verification, authentication, preparation, approach strategy, readiness, and so on, while the other colors like Green, Yellow, Blue, and Brown represent the Collection Process Phase, Preservation Process Phase, Analysis Process Phase, Presentation Process Phase as well as their synonym processes respectively. The common investigation process phases and their synonyms represent the most process phases which probably have covered the digital forensic discipline. Table 3 explains the DoC of value for each common process phase and its synonym process.

Practically, the authors reconcile, and improve the common investigation process phases by adding mandatory and optional forensic concepts and terminologies which distinguish the proposed model. For example, the mandatory forensic concepts and terminologies are law/regulation, database resources, investigation team, authorization, detection server, database incident, identification, preservation, and volatile and nonvolatile artifacts, whereas the optional concepts and terminologies are network resources, and OS resources which are displayed in Figure 2.

The advantages of this model are it reduces confusion and heterogeneous of the investigation task through providing an obvious structure which has pure database forensic investigation concepts such as forensic methods, algorithms, detection servers, volatile and nonvolatile artifacts, gathering evidences , database servers, database resources, guidelines, analysis, hashing, documentations, and so on. Additionally, offering customizing building models for example you can build your own model from Conceptual *Database Forensic Investigation Process Model* to solve your problem like: detect server model, detect database tampering model, detect database model, analysis volatile artifacts or nonvolatile and submit report models and so on. Thus the user can easily develop a customized model.

5.1 Identification Process Phase

Recognizing an incident from indicators and determining its type. This is not explicitly within the field of forensics, but significant because it impacts other steps. It prepares tools, techniques, search warrants, and monitoring authorizations and management support" [20]. The primary goal of identification process phase to identify investigation requirements concepts such as database resources, operating system resources, network resources, investigation teams, investigations techniques, investigation environment, policies, laws and regulations, authorizations and data associated with database incident.

Protect crime scene must be captured, protected and documented in details using proper procedures and experienced teams. The investigations teams should be skilled and experienced to avoid altered or damaged evidences during the investigation task. Detection server where database resides and database incident occurred is the first investigation check.

Detection methods (network scanning tools & DB detection methods) and staff interview (high access users) are using to provide specific information that used to assist investigators in detecting host machine and database server. Thus the detection sever report will be produced with detailed information about the detection server. Identification process phase is displayed in detail in Figure 2.

Investigation requirement concepts should be described in detail to give the Database Forensic investigation community unified access point of database forensic knowledge. *Database resources concept* is a collection of volatile and nonvolatile artifacts which are discussed in Section II. Thus, the database log files, history files, data files, authentication files, backup files, archive files, auditing files, alert files, trace files, password and parameter files, transaction logs, data cache, SQL cache, shared pool cache; data dictionary caches

should be identified, captured, documented and protected. An *operating systems resources concept* is a group of hardware and software concepts. Hardware concepts include hard disk drives, compact disks, flash disks, flash memory, main memory, PCs, laptops, and smartphones, whereas software concepts includes Microsoft Windows systems, Macintosh systems, and application programs. Another resource is *network resources concept* which includes TCP/IP, network traffic data sources firewalls, routers, packet sniffers, protocol analyzer, intrusion detection systems (IDS), remote access control, security event management software, network forensic analysis tools), dynamic host configuration protocol servers (DHCP), network monitoring software, internet server provider records(ISP), client server application, hosts' network configurations and connections, collecting network traffic data (legal considerations) [39]. *Investigation team concept* is a certified, skilled and experienced team that has enough training and previous experiences in that same field. It is classified into three investigation teams: The Identification team, the Collection and the Preservation team, and the Analysis and Presentation team. The identification team in collaboration with the organization management team is charge of preparing and identifying identification requirements resources and detecting the database server and database incident, and prepare the identification report, whereas the collection team is responsible of for collecting and preserving volatile and nonvolatile data taking into account the legal considerations in protecting information privacy, as well as generate collection and preservation reports to the analysis team. The analysis team has a powerful experience of analysis tools and algorithms to reveal database incidents through reconstructing database activities and discover: Who is a criminal? When did the crime happen? What data did it tamper? Why and how did the crime happen? Also restoring and recovering database continuity and integrity as soon as possible. Additionally, to produce the analysis and presentation report this is submitted to the top management and court. The most important issue that the investigation team must focus in is the trust of using investigation techniques to avoid damage or lost valuable evidences. *Investigation Techniques concept* is a collection of investigation tools, methods, and algorithms that are used to detect, gather, protect, analyze, reconstruct, recover and document database events. The investigation task should be achieved in a secure and trusted environment.

The *Investigation environment concept* considers the media that contains investigation procedures and functions. It includes the host server, alternative server, location, laboratory, and safety measures and investigation teams, and also must take into account fluctuations in the air in terms of humidity and temperature in order to examine and store procedures properly. Host server is the place where target database, OS, and applications reside. The alternative server is another server that maybe used during the investigation task to conduct main or additional examination. Location is the place where the investigation task is conducting; it may be conducted in the same place as the host server or moved to laboratory. The laboratory is full of the equipment's testing place which has whole opportunities and safety measures to conduct the investigation task. The safety measures are tools, policies, and awareness's that must follow, to protect valuable evidences and ensures the results. Examples of safety measures are fire extinguishers, power supply to avoid power outage, air filters, physical security (manual lock, auto lock, CCTV, alerts, smoking alerts, biometric devices), security guard, awareness posters, policies that define responsibilities and penalties in case of disasters that will happen. In fact investigation environment must be far-off flooding and earthquake regions which may cause damage and loss of data. Investigations teams must follow organization policies and take into account the laws and regulations of territory/state to avoid any prosecution in future.

Identifying *Policies concepts* of organization is giving investigators a general understanding of organization purposes, limitations, and procedures which must be followed and complied. Organization policies are procedures and rules that govern organization behaviors. Investigators teams should comply and understand organization policies before the start of the investigation steps. Reviewing these policies will give investigators knowledge about procedures and rules which may be conducted when incidents/disasters happened. For example if data tampered, intrusion, misused, lost, theft, compromised, damaged, deleted, changed, or fraud then what the offer procedures that should be complied with to mitigate or solve it. Strangeness or weaknesses of organization policies depend on orientations of organization and type of sponsored organization security. Therefore, the investigation team must assess and understand organization policies to avoid any prosecution in future when an unnatural action happens.

Laws or regulations are other concepts that should be identified and prepared for the investigation teams. Database forensic is a relatively new discipline to the courts and many of the existing laws used to prosecute computer related crimes, legal precedents and practices related to database forensics are in a state of flux, therefore it is very important for the forensic investigator teams to collect evidence in a way that is legally admissible in court. Forensic investigation team should also be aware of privacy laws and country specific laws that are imposed on data collection and retention for forensic purposes, violation of any one of these laws during practice of cyber forensics could constitute a federal felony. The existing laws/regulations are specifying investigations and responds to security breaches or policy violations. For example finalized

HIPPA (Health Insurance Portability and Accountability Act of 1996) rules include "information security" which encompasses incident response describing the attempted or successful unauthorized access, use, disclosure, modification or destruction of information or interference with system operations in an information system. HIPPA specifies that there should be thorough analysis and reporting of security incidents [12]. This gives rise to the need for database forensics which satisfies this demand. Organizations must, therefore, consider their incidence response policies carefully, which are part of their overall security policies.

Authentication, authorization, and accounting (AAA) concepts are a term for a framework for intelligently controlling access to computer resources (Network resources, OS resources, Database Resources), enforcing policies, auditing usage, and providing the information necessary to bill for services. These combined processes are considered important for effective network management and security. Forensic investigator teams are using these processes to get access to computer target to achieving their mission. As the first process, *authentication* is providing a way of identifying a user, typically by having the user enter a valid user name and valid password before access is granted. The process of authentication is based on each user having a unique set of criteria for gaining access. The AA server compares a user's authentication credentials with other user credentials stored in a database. If the credentials match, the user is granted access to the network. If the credentials are at variance, authentication fails and network access is denied. Following authentication, a user must gain *authorization* for doing certain tasks. After logging into a system, for instance, the user may try to issue commands. The authorization process determines whether the user has the authority to issue such commands. Simply put, authorization is the process of enforcing policies: determining what types or qualities of activities, resources, or services a user is permitted. Usually, authorization occurs within the context of authentication. Once you have authenticated a user, they may be authorized for different types of access or activity. Therefore, and in this case *forensic investigation teams* are using these processes to get accessing to network and database resources to achieving their mission. However, in some cases, especially when complaints have been raised against a malicious company. Malicious companies are doing illegal activities such as frauds or suspicious businesses. In this case the organization management is attempting to covert database server (permanent server) for a while to hide their activities. Thus the forensic investigation teams have two trends: protect crime scene and server detection, Figure 1 shows the processes that forensic investigators may follow.

Server detection deals with a method of detecting the server driving a database system. In case the internal system of a company is investigated, there are various systems to be examined as an investigation target and the investigation should be carried out within a time limit, it is difficult to judge what system should be first selected and checked into. In such a case, it is required to grasp the overall network circumstance in the company as soon as possible, so it is important to acquire the network topology inside the company. To acquire the network topology, detecting server group systems and the host system will be the main purpose and this study especially focuses on detecting the database server where data is stored [30, 33]. Thus, detection methods are divided into two parts: *Detection Methods* and *Staff Interview* as illustrated in Figure 1. Detection Methods include *Network Forensic Analysis Tools (NFAT)*, and *DB Detection Methods*. *Network Forensic Analysis Tools* allow administrators to monitor networks, gather all information about anomalous traffic, assist in network crime investigation and help in generating a suitable incident response [40]. NFATs also help in analyzing the insider theft and misuse of resources, predict attack targets in near future, perform risk assessment, evaluate network performance, and help to protect intellectual propriety. Many commands are available inbuilt in modern operating systems which can be used for assisting network forensics [40, 41].

DB Detection Methods are used to detect: database server either in normal or in covert case, database schema, user accounts, privileges, files locations, and target files [33], [30]. In the case when database hides, an investigator looks into the organization that is suspected of illegal acts, much human power and time will be required in order to detect the covert database system as evidence about the case, due to the reason that the structure of computers and networks is highly complicated. In addition, if the organization intentionally builds the covert database system for storing and managing data then investigators should be placed in a predicament to detect it. For the reasons given above, the investigator needs digital forensic technologies which effectively investigate the case and obtain evidence about its system structure of the computer and network [33]. The digital forensic techniques are used to detect covert database are DB traces detector, Net-BIOS, ping sweep, port scanning, and ActiveX Data Object technology [42, 43].

The second method which is used to detect database server is *Staff Interview*, it is very important to have *interviews* with staff member of the company in charge of the system. In the interview stage, it is possible to acquire information difficult to acquire through tools. As information is possible to acquire, we are able to know the location of servers and account of information in addition to basic information such as server IP and service port numbers. In fact, it is a crucial stage because we can possibly acquire more information through interviews than by scanning the database server with scanning tools.

Besides such a method, we also have a way of detecting servers by examining the contents of the host computer inside a company. To work on the host computer, *High Access Users* such as DBA, system administrator, and powered employees should access the database server and then trace remaining in the host computer. Basically, one of the typical traces remaining in the database server is that information about the database server remaining in the setting's file titled 'tnsnames.ora' in the case of Oracle [30].

Table 2 Comprehensive Analysis of Digital Forensic Investigation Process Models

Process	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	Type
Identification	√		√	√		√					√						√					Pure
Detection																				√		Synonym
Detection Server																√						synonym
Planning							√				√											synonym
Investigation Preparation															√							synonym
Setup The Evidence Collection Server														√								synonym
Suspend Database Operation													√									synonym
Readiness					√			√														synonym
Examination Preparation																					√	synonym
Authentication		√																				synonym
Preparation				√					√			√										Pure
Collection			√	√		√							√	√		√	√	√	√	√		Pure
Acquisition	√																					synonym
Acquiring		√																				synonym
Digital Crime Investigation					√			√				√										synonym
Search							√															synonym
Pre-Analysis										√												synonym
Collect Artifacts															√							synonym
Digital Examination																					√	synonym
Preservation			√	√		√							√		√		√		√			Pure
Defense							√															synonym
Transport& Storage											√											synonym
Examination																					√	synonym
Analysis		√	√	√		√				√	√		√	√			√					Pure
Reconstruction													√									synonym
Artifacts Analysis															√		√					synonym
Investigation of Data Collection																√						synonym
Presentation			√	√		√			√	√		√										Pure
Documentation																					√	synonym
Report																	√					synonym
Disseminating							√															synonym

Table 3 Degree of Confidence of Common Investigation Process Phases

Process	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	Frequency
Identification	√	√	√	√	√	√	√	√	√		√	√	√	√	√	√	√			√	√	85%
Collection	√	√	√	√	√	√	√	√		√		√	√	√	√	√	√	√	√		√	90%
Preservation			√	√		√	√				√		√		√		√		√		√	52%
Analysis		√	√	√		√				√	√		√	√	√	√	√					52%
Presentation			√	√		√	√		√	√		√					√				√	42%

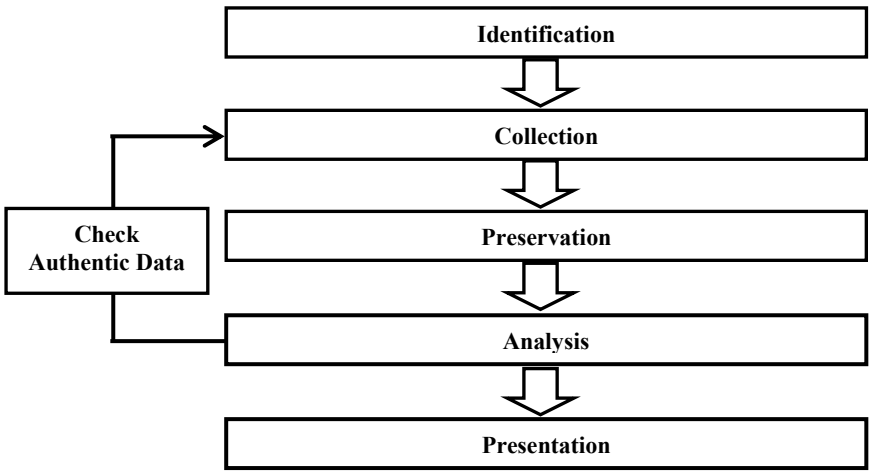


Figure 1 Database Forensic Investigation Process Model (DBFIPM)

Consequently, employees also use client programs to access the database server, especially SQL Server. For example, when SQL Server is installed, other programs are also basically installed, such as ' MS SQL Server Management Studio ' and PostgreSQL, followed by ' PostgreSQL pgAdmin III';' MySQL Query Browser' , a client used to access the MySQL server, and 'Oracle SQL Developer', a client used to access Oracle; and specific client programs to support various kinds of database products, such as 'SQLGate Series' and 'Toad Series', which manage the access breakdown as a variety of file formats. Inside the files exists information about the server access breakdown, such as account names, server addresses, service port numbers, schema names, and passwords. Even though they are encoded, it is possible to decode them, that is, with such information; it is possible to access the database server by acquiring information about the account, even without knowing information about the account [34].

Detection Server Report is generated to give clear ground for investigators, users, and management. It will be given a clear view for current and newcomer's investigators, to know how they should detect database servers in both cases, either in normal or in covert database corporations. Also, normal users and management will increase their knowledge about detecting database servers. Additionally, the next *Identification Report* will include this report and complete detail information about the identification phase. The purpose of this phase is to establish fundamentals for sharing investigation knowledge among database users and practitioners.

Identifying proper investigation requirements and detecting investigation targets which are documented in detail in the detection server report give investigators starting points towards moving to detect database incidents and submit the final identification report to the top management and court. *Detection database incident stage* includes checking malicious activities and identifies database dimensions which are achieved by the identification team. *Malicious activities* are authorized or unauthorized actions which destroy database dimension integrity or confidentiality such as SQL injections attacks, malware attacks, fraud credit card, and steal records. Database dimensions are database layers which mentioned in Section II. *Identification teams* are professional groups that have enough experts and training by using detection database forensic techniques and algorithms, moreover has full experiences in revealing database incidents.

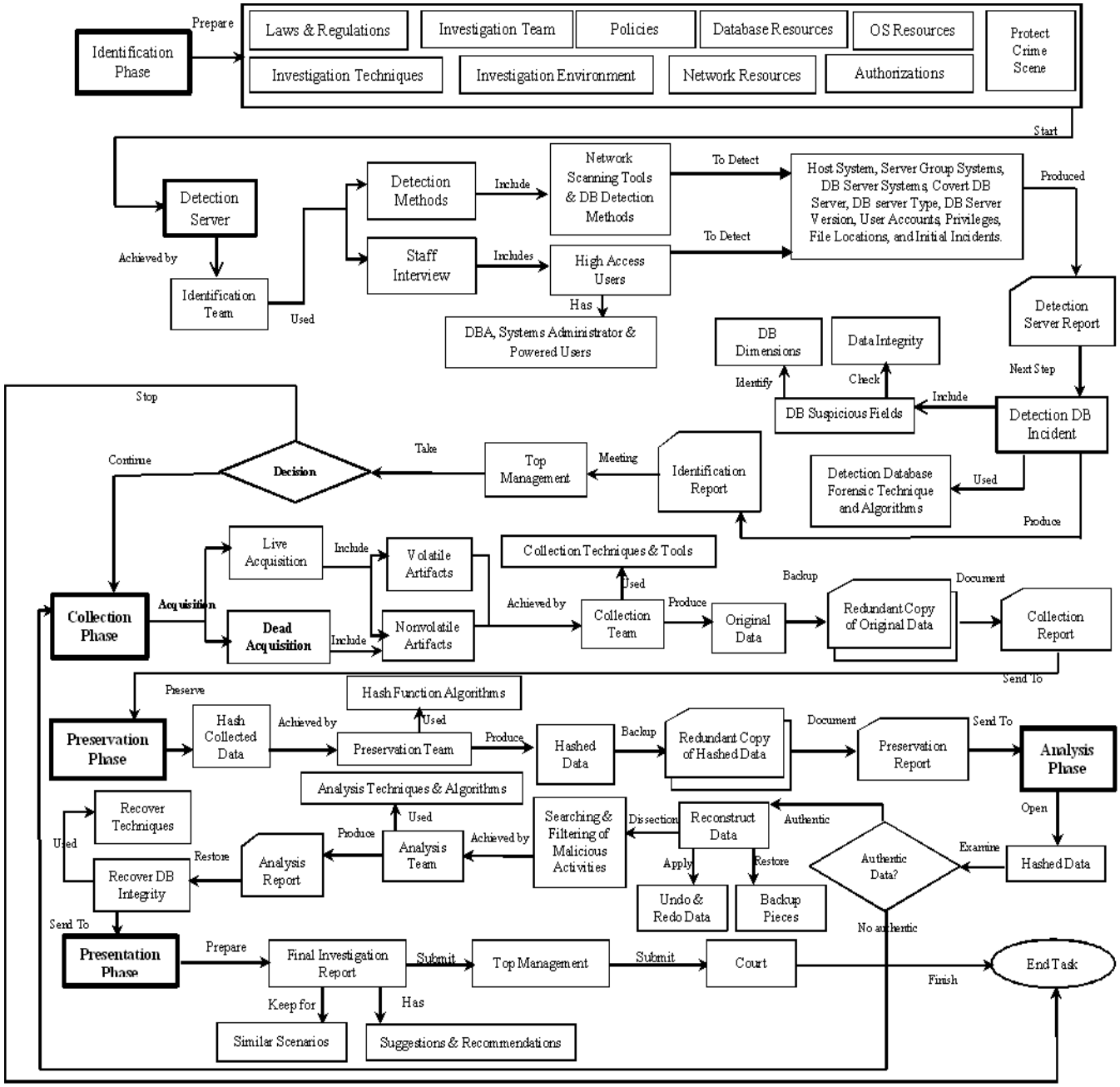


Figure 2 Conceptual Databases Forensic Investigation Process Model

Detection database forensic techniques and algorithms are special detection methods which are using by identification team to scanning database suspicious fields such as database detection algorithm [44], and database detection techniques [8, 45-47]. Mismanagement of detection techniques may lead to unexpected results or may be destroyed evidences. *Database suspicious fields* are fields that may be compromised, damaged or changed intentionally or unintentionally such as database files, data dictionary, logic schema and application schema. *Data Integrity* means checking whether the database system has detection tampering mechanism like strong cryptography one way hash function which applies in protecting the original data.

The Identification Report is to produce and submit to the top management to make a decision either stop or continue with the investigation task. *Identification report* is a detailed preparation and detection report which carries specific and particular information about the identification process phase such as identification team names, experiences, skills, certificates, resources, database incidents, type of incidents , incident time, type of attacks, attack resources, vulnerabilities, investigation

techniques, type of database server, host machine, users, privileges, laws/regulations, policies, authentication and authorization files, cost and time etc. It records most of the investigators skills which may be used the next time similar incident happen. It will become a reference for both users and managements. For users and newcomers it provides knowledge about which forensic investigation techniques and algorithms may be used during investigation and also what technical plans have the identification stage achieved? Furthermore, it offers a good background for management to know what investigation means. What are investigators responsibilities and roles? What are plans, strategies actions, costs, time of investigation? Top management is the a chief executive officer (CEO) who generally the most senior corporate officer (executive) or administrator in charge of managing a for-profit organizations. Thus, the decision will be taken by the CEO to continue or stop any investigation tasks, whether the decision is to continue investigation, an agreement must be written between organization and the investigation team. The CEO is therefore responsible for stopping or continuing the investigation task.

Finally, the identification report will be submitted to the top management. There are two options: stopping the investigation task or continuing. In case of continuous investigation, the agreement must be written among the top management and the investigation team to avoid any prosecutions in the future.

5.2 Collection and Preservation Phase

Collecting and preserving evidences without damaging or altering it is very challenging for the investigators, thus the data must be backed up and copies saved before starting the collection process. Procedurally, extracting and preserving evidences from database systems are somehow similar to extracting and preserving networks, computers and mobile evidences, nevertheless the concepts and terminologies are totally different for example the database artifacts include SQL cache, data cache, log transactions, data files, log files, control files, backup pieces, archive pieces, alert files, auditing files, trace files, configuration files, data dictionary views, authorization logs, authorization events, and so on. Furthermore, the collection and preservation team must have a good knowledge about database architecture and have licenses to allow him submitted evidences to court.

Moreover, the collection techniques and tools must be specific and reliable. The good feature of these phases which are illustrated in Figure 2 allows the user to build a customized model for example (dead acquisition, nonvolatile artifacts, and hash collected data concepts that consists dead acquisition model) which is illustrated in Figure 2. Indeed, this feature does not exist in other digital models because they were not concentrating on investigation concepts and their relationships.

5.3 Analysis Phases

Considers the core database forensic investigation phase which is used to analyze collected data and reveals: Who was tampering? When did tampering happen? What data was tampered with? Why and how did tampering happen? Figure 2 shows the concepts of the analysis and presentation phase. Therefore, collected and preserved data are presented in this phase and the authentic of the data checked. If the data is altered or damaged it will return to the collection phase to collect from the original data, nevertheless if the data is authentic, then the reconstruction process will begin. The recent backup set will restore and then recover database until failure happen by applying redoing or undoing log files [4, 10, 48, 49]. Forensic analysis algorithms [50] are used in revealing malicious activities and attackers. Database integrity is recovering using recovery techniques [45-47] to make consistent the database transactions and checkpoints. The results will be sent to the presentation phase.

5.4 Presentation Phase

The final stage is the presentation phase which is used to document the whole investigation task and then submit to the top management and court. Investigation documenting has many features which used to assist newcomers or investigators who face the same scenario in future and also useful for investigators to protect them from any future prosecution. It has many suggestions and recommendations which guide users and newcomers to increase their knowledge.

6.0 CONCLUSION

As a new and fast growing research field, knowledge about database forensic is important to be explored. In this paper, a generic model specific for database forensic investigation process, known as the DBFIPM is developed. To construct the DBFIPM, a number of existing investigation process models related to database is reviewed. From a thorough investigation against these models, the DBFIPM reveals that the database forensic investigation process has 5 common process phases which include the: i) identification, ii) collection, iii) preservation, iv) analysis phase and v) presentation phase. To validate the completeness of the DBFIPM model, the FBS technique is applied against the model. The future works of this research is to detail out all concepts and relationships in each of the identified phases (in DBFIPM) by adapting a software engineering approach known as a metamodel.

Acknowledgement

The authors would like to thank the Ministry of Higher Education Malaysia (MoHE) and UniversitiTeknologi Malaysia (UTM) for funding this study under the Fundamental Research Grant Scheme (FRGS) of GrantS No. (R.J130000.7828.4F498) and No. (R.J130000.7813.4F193).
.

References

[1] Bertino, E. and R. Sandhu. 2005. Database Security-Concepts, Approaches, And Challenges. *Dependable and Secure Computing, IEEE Transactions on*. 2(1): 2-19.
[2] Olivier, M. S. 2009. On Metadata Context In Database Forensics. *Digital Investigation*. 5(3): 115-123.
[3] Guimaraes, M. A., R. Austin and H. Said. 2010. Database Forensics. *2010 Information Security Curriculum Development Conference*: ACM. 62-65.
[4] Fasan, O. M. and M. Olivier. 2012. Reconstruction in Database Forensics. *Advances in Digital Forensics VIII*. Springer. 273-287;
[5] Hauger, W. K. and M. S. Olivier. 2014. The Role Of Triggers In Database Forensics. *Information Security for South Africa (ISSA), 2014*: *IEEE*. 1-7.
[6] Yusoff, Y., R. Ismail and Z. Hassan. 2011. Common Phases Of Computer Forensics Investigation Models. *International Journal of Computer Science & Information Technology (IJCSIT)*. 3(3): 17-31.
[7] Beyers, H., M. Olivier and G. Hancke. 2011. Assembling Metadata For Database Forensics. *Advances in Digital Forensics VII*. Springer. 89-99.
[8] Khanuja, H. K. and D. Adane. 2013. Forensic Analysis of Databases by Combining Multiple Evidences. *International Journal Of Computers & Technology*. 7(3): 654-663.
[9] Adedayo, O. M. and M. Olivier. 2014. Schema Reconstruction in Database Forensics. *Advances in Digital Forensics X*. Springer. 101-116.
[10] Wong, D. and K. Edwards. 2004. System And Method For Investigating A Data Operation Performed On A Database. Google Patents.
[11] Fowler, K. 2008. *SQL Server Forenisc Analysis*: Pearson Education.
[12] Khanuja, H. K. and D. D. Adane. 2012. A Framework For Database Forensic Analysis. *Published in Computer Science & Engineering: An International Journal (CSEIJ)*. 2(3).
[13] Grobler, C., C. Louwrens and S. H. Von Solms. 2010. A Framework To Guide The Implementation Of Proactive Digital Forensics In Organisations. *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*: IEEE. 677-682.
[14] Wright, P. M. and D. Burleson. 2008. *Oracle Forensics: Oracle Security Best Practices*. Rampant Techpress.
[15] Cohen, M., D. Bilby and G. Caronni. 2011. Distributed Forensics And Incident Response In The Enterprise. *Digital Investigation*. 8: S101-S110.
[16] Flores, D., O. Angelopoulou and R. J. 2012. Self. Combining Digital Forensic Practices and Database Analysis as an Anti-Money Laundering Strategy for Financial Institutions. *Emerging Intelligent Data and Web Technologies (EIDWT), 2012 Third International Conference on*: *IEEE*. 218-224.
[17] Pollitt, M. 1995. Computer Forensics: An Approach To Evidence In Cyberspace. *Proceedings of the National Information Systems Security Conference*. 487-491.
[18] Kruse II, W. G. and J. G. Heiser. 2001. *Computer Forensics: Incident Response Essentials*: Pearson Education.
[19] Palmer, G. 2001. A Road Map For Digital Forensic Research. *First Digital Forensic Research Workshop, Utica, New York*. 27-30.
[20] Reith, M., C. Carr and G. Gunsch. 2002. An Examination Of Digital Forensic Models. *International Journal of Digital Evidence*. 1(3): 1-12.
[21] Carrier, B. and E. H. Spafford. 2003. Getting Physical With The Digital Investigation Process. *International Journal of digital evidence*. 2(2): 1-20.
[22] Baryamureeba, V. and F. Tushabe. 2004. The Enhanced Digital Investigation Process Model. *Proceedings of the Fourth Digital Forensic Research Workshop*: Citeseer. 1-9.
[23] Ciardhuáin, S. Ó. 2004. An Extended Model Of Cybercrime Investigations. *International Journal of Digital Evidence*. 3(1): 1-22.
[24] Carrier, B. and E. H. Spafford. 2004. An Event-Based Digital Forensic Investigation Framework. *Digital Forensic Research Workshop*. 11-13.
[25] Köhn, M., M. S. Olivier and J. H. Eloff. 2006. Framework for a Digital Forensic Investigation. *ISSA*. 1-7.
[26] Freiling, F. C. and B. Schwittay. 2007. A Common Process Model for Incident Response and Computer Forensics. *IMF*. 7: 19-40.
[27] Perumal, S. 2009. Digital Forensic Model Based On Malaysian Investigation Process. *International Journal of Computer Science and Network Security*. 9(8): 38-44.
[28] Kohn, M. D., M. M. Eloff and J. H. Eloff. 2013. Integrated Digital Forensic Process Model. *Computers & Security*. 38: 103-115.
[29] Tripathi, S. and B. B. Meshram. 2012. Digital Evidence for Database Tamper Detection. *Journal of Information Security*. 3: 113.
[30] Son, N., K.-g. Lee, S. Jeon, H. Chung, *et al.* 2011. The Method of Database Server Detection and Investigation in the Enterprise Environment. *Secure and Trust Computing, Data Management and Applications*. Springer. 164-171.
[31] Azemović, J. and D. Mušić. 2009. Efficient Model For Detection Data And Data Scheme Tempering With Purpose Of Valid Forensic Analysis. *2009 International Conference on Computer Engineering and Applications (ICCEA 2009)*.
[32] Azemovic, J. and D. Music. 2010. Methods for Efficient Digital Evidences Collecting of Business Proceses and Users Activity in eLearning Enviroments. *e-Education, e-Business, e-Management, and e-Learning, 2010. IC4E'10. International Conference on*: IEEE. 126-130.
[33] Lee, G. T., S. Lee, E. Tsomko and S. Lee. 2007. Discovering Methodology and Scenario to Detect Covert Database System. *Future Generation Communication and Networking (FGCN 2007)*: IEEE. 130-135.
[34] Lee, K. and M. R. Boddington. 2012. A Workflow to Support Forensic Database Analysis.
[35] von Alan, R. H., S. T. March, J. Park and S. Ram. 2004. Design Science In Information Systems Research. *MIS quarterly*. 28(1): 75-105.
[36] Othman, S. H. and G. Beydoun. 2010. Metamodelling Approach To Support Disaster Management Knowledge Sharing.
[37] March, S. T. and G. F. Smith. 1995. Design And Natural Science Research On Information Technology. *Decision support systems*. 15(4): 251-266.
[38] De Kok, D. 2010. Feature Selection For Fluency Ranking. *Proceedings of the 6th International Natural Language Generation Conference*: Association for Computational Linguistics. 155-163.
[39] Kent, K., S. Chevalier, T. Grance and H. Dang. 2006. Guide To Integrating Forensic Techniques Into Incident Response. *NIST Special Publication*. 800-86.
[40] Pilli, E. S., R. C. Joshi and R. Niyogi. 2010. Network Forensic Frameworks: Survey And Research Challenges. *Digital Investigation*. 7(1): 14-27.
[41] Ali, A., A. Al-Dhaqm and S. A. Razak. 2014. Detecting Threats in Network Security by Analyzing Network Packets

using Wireshark. *Proceeding International Conference of Recent Trends in Information and Communication Technologies*. IRICT 2014.
[42] Bregu, J., D. Conklin, E. Coronado, M. Terrill, *et al.* 2013. Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis. *Journal of Forensic Sciences*. 58(1): 120-129.
[43] Lee, K.-g., A. Savoldi, P. Gubian, K. S. Lim, *et al.* 2008. Methodologies For Detecting Covert Database. *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on*: IEEE. 538-541.
[44] Snodgrass, R. T., S. S. Yao and C. Collberg. 2004. Tamper Detection In Audit Logs. *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*: VLDB Endowment. 504-515.
[45] Fruhwirt, P., M. Huber, M. Mulazzani and E. R. Weippl. 2010. Innodb Database Forensics. *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*: IEEE. 1028-1036.
[46] Basu, A. 2006. *Forensic Tamper Detection In SQL Server.*
[47] Khanuja, H. K. and D. S. Adane. 2014. Forensic Analysis for Monitoring Database Transactions. *Security in Computing and Communications*. Springer. 201-210.
[48] Frühwirt, P., P. Kieseberg, S. Schrittwieser, M. Huber, *et al.* 2013. InnoDB Database Forensics: Enhanced Reconstruction Of Data Manipulation Queries From Redo Logs. *Information Security Technical Report*. 17(4): 227-238.
[49] Xu, M., J. Yao, Y. Ren, J. Xu, *et al.* 2014. A Reconstructing Android User Behavior Approach based on YAFFS2 and SQLite. *Journal of Computers*. 9(10): 2294-2302.
[50] Pavlou, K. E. and R. T. Snodgrass. 2008. Forensic Analysis Of Database Tampering. *ACM Transactions on Database Systems (TODS)*. 33(4): 30.