IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 8, 2017, accepted October 7, 2017, date of publication October 13, 2017,
date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2762693

CDBFIP: Common Database Forensic Investigation Processes for Internet of Things

ARAFAT AL-DHAQM[1,2], SHUKOR RAZAK[1], SITI HAJAR OTHMAN[1], KIM-KWANG RAYMOND CHOO[3], (Senior Member, IEEE), WILLIAM BRADLEY GLISSON[4], ABDULALEM ALI[1], AND MOHAMMAD ABRAR[1]

[1]Department of Computer Science, Faculty of Computing, Universiti Teknologi Malaysia, Johor 81310, Malaysia
[2]Department of Computer Science, Aden Community College, Aden 262, Yemen
[3]Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249-0631, USA
[4]School of Computing, University of South Alabama, Mobile, AL 36688, USA

Corresponding author: Arafat Al-Dhaqm (arafataldoqm@gmail.com)

This work was supported in part by the Universiti Teknologi Malaysia, in part by MOHE through FRGS under Grant R.J130000.7813.4F193, and in part by the Universiti Teknologi Malaysia (GUP) under Grant Q.J130000.2513.18H31.

ABSTRACT Database forensics is a domain that uses database content and metadata to reveal malicious activities on database systems in an Internet of Things environment. Although the concept of database forensics has been around for a while, the investigation of cybercrime activities and cyber breaches in an Internet of Things environment would benefit from the development of a common investigative standard that unifies the knowledge in the domain. Therefore, this paper proposes common database forensic investigation processes using a design science research approach. The proposed process comprises four phases, namely: 1) identification; 2) artefact collection; 3) artefact analysis; and 4) the documentation and presentation process. It allows the reconciliation of the concepts and terminologies of all common database forensic investigation processes; hence, it facilitates the sharing of knowledge on database forensic investigation among domain newcomers, users, and practitioners.

INDEX TERMS Forensics, database forensics, Internet of Things forensics.

I. INTRODUCTION

The integration of apps on mobile devices, software used on Internet of Things (IoT) devices, along with the need to improve functionality in numerous business applications, continue to motivate interest in the area of database forensics. Coupling this information with increasing interest in residual data in legal environments [1], [2], necessitates research in this relatively understudied area. Database forensics is a branch of digital forensics that uses database content, metadata, log files, data files, and memory data to create timelines, establish relationships and recover relevant data [3].

The numerous pieces of input that potentially impact a digital forensic investigation involving a Database Management System (DBMS) is inherently more complex than that of a traditional file system [4], particularly in an IoT environment that is likely to comprise a wide range of heterogeneous devices. For example, in an IoT deployment in battlefields (also referred to as Internet of Battlefield Things or Internet of Military Things), we would need to collect evidential data from a wide range of sensors installed on military vehicles, cameras (e.g. IP-based CCTV feeds), road side units, etc. These data are also likely to be in different formats (e.g. proprietary formats) and of different sizes. While files are often abstracted as streams of bytes, a database is a collection of data where data elements are related to one another. Hence, the process or procedures that are used in a digital investigation directly impact the results of the examination. Selecting investigative processes that do not fit, potentially, leads to incomplete and/or loss of evidence [5]. This incomplete or missing data may lead to indecisive consequences and give invalid conclusions. In other words, evidence that is not acquired in a forensically sound manner may risk being inadmissible in a court of law [6].

Thus, several database forensic investigation models have been proposed in the literature. The proposed models range in focus from specific scenarios to generic applications and, in doing so, provide a variety of details as a result. The diversity in scenarios and resulting details perceivably make it challenging, or even complicated, particularly for newer forensic investigators, to adopt accurate or proper

investigation models. Due to the lack of a generic/common database forensic investigation process model [7], [8], this study provides a structure, referred to as the Common Database Forensic Investigation Process (CDBFIP), which unifies, facilitates, and shares knowledge of database forensic investigation processes among database users and practitioners. Unifying these processes in a single abstract diagram increases the knowledge available to users, newcomers, and practitioners. Additionally, it reduces the complexity and ambiguity of the investigation.

This paper is organized as follows. Section II presents works related to this research. Section III describes the actual development process of our CDBFIP based on the Design Science Research (DSR) methodology. While it would be ideal to evaluate CDBFIP using real-world IoT cases, it is challenging to obtain access to such cases partly due to the sensitive nature of forensic investigations. Therefore, in Section IV, we use a set of nine database forensic models to validate CDBFIP. Finally, we conclude the paper in Section V with recommendations for possible future work.

II. BACKGROUND AND RELATED WORK

The database forensics domain uses database content and metadata to identify, collect, preserve, reconstruct, analyse and document evidence of crime. Owing to the complexity and multidimensional nature of DBMS, database forensics has received little attention from researchers [3]–[5], [7], [9], [10]. Existing works in database forensics can be viewed from three perspectives, namely: technology (tools, algorithms, and methods, etc), investigation process (e.g. identification, collection, analysis, and presentation) and the database forensic dimension (destroyed dimension, compromised dimension, changed dimension, etc) [11].

A number of studies have been conducted on Oracle databases; these works encompass specific Oracle investigation tasks, activities, processes, techniques, concepts and terminologies. For instance, a forensic investigation model was developed by Wong and Edwards [12] that consists of generic steps to discover information about an operation performed on the database [9]. A Log Miner tool was developed by Wright [13] that allows a Data Base Administrator (DBA) or forensic analyst to reconstruct actions taken on a database [11]. Seven practical investigation forensic models were developed by Litchfield [14]. The author addressed information available from redo logs, dropped objects, authentication, flashback, and a recycle bin. Wright and Burleson [15] published a forensic textbook on Oracle RDBMS; however, the book was written as a guide and intended for database administrators [9]. A block diagram to collect evidence was developed by Tripathi and Meshram [16] that is based on a series of practical methods developed to analyse database tampering in an Oracle database [14].

In addition, a number of practical studies have been carried out on Microsoft SQL Database Server (MSSQL Database Server). These studies covered database investigation tasks, activities, processes, techniques, concepts and terminologies. A methodology for an SQL Server Forensic Analysis was proposed by Fowler [10]; the proposed methodology involves four processes, namely, investigation preparation, incident verification, artefact collection, and artefact analysis, dealing specifically with an SQL server database [4], [11]. A Detection and investigation model was developed by Son *et al.* [17] to detect the database server and collect data. Fowler *et al.* [18] also proposed an approach to gather and analyze data in practical and real-world scenarios. Their approach covers technical concepts that investigators use to investigate compromised databases. An approach for detecting the tampering of forensic evidence in an SQL server was proposed by Basu [19]; the proposed technique shows how tampering can be detected and how to localize the affected data but does not provide tamper-prevention measures [20], [21]. A methodology to distinguish suspicious transactions from legitimate transactions among the SQL Server artefacts, in order to eliminate irrelevant data, was proposed by Khanuja and Adane [22]. This methodology specifically considered the investigation processes, technology, and database dimension in its solution.

Relatively few studies have been conducted that investigate a MySQL database in reference to database forensics. A study by Khanuja and Adane [3] did develop a framework for MySQL database forensic analysis. The proposed framework focuses on discovering malicious tampering in My SQL database. The authors further highlighted the database server artefacts that were employed for the investigation.

It is worth noting that a model for detecting database inconsistencies was proposed by Fruhwirt *et al.* [23]; nevertheless, there was no knowledge discovered for multiple log files or cache for further analysis [24]. To address the limitations of the model proposed by Frühwirt *et al.* [24], another model was proposed by Frühwirt *et al.* [25] to reconstruct the basic SQL statements from InnoDB's redo logs. However, this model focuses on the Data Manipulation Language (DML) statements and ignores the Data Definition Language (DDL) statements. To address situations where the user is unavailable or where the user is under investigation, Lawrence [26] proposed a technical investigation method to gain access to users' MySQL database without user consent. According to Adedayo [27], several works have been conducted in digital forensics. However, they largely focus on cloud forensics and smartphone forensics [28]–[45]. They do not directly mention ddatabase forensics. While certain aspects of database forensics have been investigated over the years [7], [11], [46], [47], there has been minimal research into the development of a standard/common approach for unifying investigation tasks and activities among investigators.

This paper aims to unify the investigation process of the database forensics in one platform called CDBFIP. It considers as abstract base for our previous paper [48] that proposed a metamodel (modelling language) for database forensics. Also, we developed a metamodel (modelling language for

mobile forensics to categorise and structure mobile forensics domain [49]. Therefore, this paper addresses the variety of the investigation process for whole database forensics domain.

III. METHODOLOGY

This study embraces a Design Science Research (DSR) methodology to design the proposed process artefact, hereafter referred to as the Common Database Forensic Investigation Processes (CDBFIP) [50], [51]. DSR is a research methodology that is used to create new and persistent artefacts for a special problem domain that enables analytics to be examined [52]. This particular implementation of DSR concentrates on Information Technology (IT) artefacts that significantly impact the application domain. According to March and Smith [52], the creation of DSR can be explained in terms of four kinds of artefacts which include the following: 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 the final artefact instantiations which are defined as combinations of constructs, models, and methods. The DSR life-cycle embraces repetitious assessment of produced artefacts. The execution of the DSR life-cycle necessitates that the building and assessment of the artefact are completed before the artefact is offered to users. The design science process includes six steps, which are illustrated in Fig. 1 [47], [53].

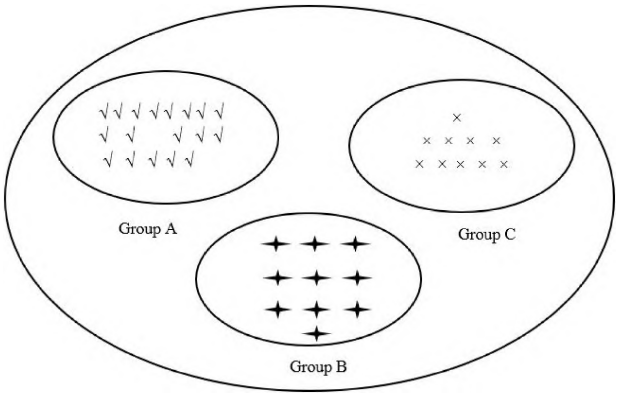

FIGURE 1. Database Forensic models covering different perspectives.

A. IDENTIFY AND SELECT DATABASE FORENSIC INVESTIGATION MODELS

Several database forensic models were discussed and analyzed in the literature review. Model selection for this research was based on coverage factors that were identified in previous research [47], [54]. A wide coverage of perspectives, concepts, and terminologies that are broadly applicable is required to fulfill the aim of proposing a common investigation process for database forensics. Using coverage metric quickly provides an indication of sourced model applicability. The model is said to have a high coverage value, if the model can cover all database forensic perspectives (i.e., a general model). The model has a reduced amount of coverage value,

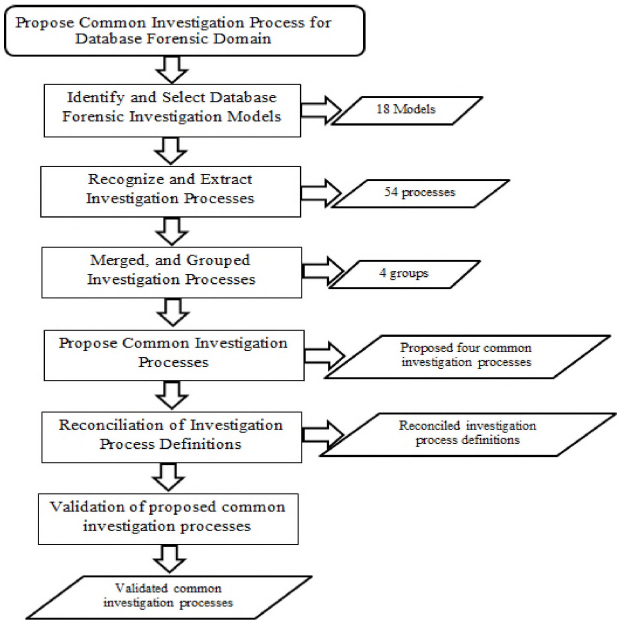

FIGURE 2. Method for proposing a common investigation process for database forensics, adapted from existing research [47], [53].

if the model only describes a specific database forensic perspective, such as the technology perspective (i.e., a specific model). Therefore, models that cover at least two database forensic dimensions, two databases forensic investigation processes, and at least one database forensic investigation technology (tool, method, or algorithm) were identified and selected for developing CDBFIP. However, models that covered only specific database forensics perspectives (i.e., one or two perspectives) were identified and selected for validation purposes. For example, a *System and Method for Investigating a Data Operation Performed on a Database Model*, which is proposed by Wong and Edwards [12], covered three perspectives of the database forensics domain. In addition, the *SQL Server Forensic Analysis Methodology approach* that is proposed by Fowler [10] covered the three database forensics perspectives. Table 1 displays thirty-eight (38) database forensic models that were identified and selected from forty (40) existing database forensic models. Fig 2 displays the models that cover both generic and specific database forensic perspectives. Group "A" refers to models that cover the all database forensic perspectives. Group "B" refers to models that cover only two database forensic perspectives. Group "C" refers to models that cover a single database forensic perspective. Step 3.2 concentrates on recognizing and extracting investigation processes from the eighteen (18) selected database forensics models.

A database forensic investigation process was derived from eighteen (18) selected models. During the course of the extraction, certain criteria [48], [55] were adhered to in order to identify a relevant and proper investigation process. The criteria used to identify the database forensic investigation processes are as follows:

TABLE 1. Database forensic investigation models.

ID	Year	Database Forensic Models "Group A"	Coverage
1	2004	System and method for investigating a data operation performed on a database [12]	√
2	2005	Forensic Analysis of a SQL Server 2005 Database Server	√
3	2007	Oracle Forensics Live Response [57]	√
4	2008	SQL Server Forensic Analysis Methodology [10]	√
5	2009	Database forensic investigation based on table relationship analysis techniques [58]	√
6	2009	Evidence Investigation Methodologies for Detecting Financial Fraud based on Forensic Accounting [59]	√
7	2009	On metadata context in Database Forensics [9]	√
8	2011	The Method of Database Server Detection and Investigation in the Enterprise Environment [17]	√
9	2012	Digital Evidence for Database Tamper Detection [60]	√
10	2012	Framework for Database Forensic Analysis [61]	√
11	2012	A Workflow to Support Forensic Database Analysis [62]	√
12	2012	On Dimensions of Reconstruction in database forensic [63]	√
13	2013	Forensic Analysis of Databases by Combining Multiple Evidences [22]	√
14	2014	Database Forensic :Investigating Compromised Database Management Systems [7]	√
15	2014	Role of metadata in forensic analysis of database attacks [64]	√
16	2014	Towards a forensic-aware database solution: Using a secured database replication protocol and transaction management for digital investigations [65]	√
17	2015	Ideal log setting for database forensics reconstruction [5]	√
18	2015	Database forensic analysis through internal structure carving [4]	√
		Database Forensic Models "Group B"	
1	2006	Forensic tamper detection in SQL server [66]	✦
2	2007	Finding Evidence of Data Theft in the Absence of Auditing [67]	✦
3	2007	Discovering Methodology and Scenario to Detect Covert Database System [68]	✦
4	2012	RECONSTRUCTION IN DATABASE FORENSICS [11]	✦
5	2012	Enriching Forensic Analysis process for Tampered Data in Database [69]	✦
6	2012	Arguments and Methods for Database Data Model Forensics [70]	✦
7	2013	*InnoDB database forensics: Enhanced reconstruction of data; manipulation of queries from redo logs [25]*	
8	2015	An Improved Framework for Tamper Detection in Databases [71]	✦
		Database Forensic Models "Group C"	
1	2004	Oracle database forensics using LogMiner [72]	×
2	2009	Efficient model for detection data and data scheme tempering with purpose of valid forensic analysis [73]	×
3	2010	Methods for Efficient Digital Evidence of Collection of Business Processes and Users' Activity in eLearning Environments[74]	×
4	2011	Detecting Database Attacks Using Computer Forensics Tools [75]	×
5	2011	An approach to examine the Metadata and Data of a Database Management System by making use of a forensic comparison tool [76]	×
6	2011	A framework for discovering internal financial fraud using analytics [77]	×
7	2012	Combining Digital Forensic Practices and Database Analysis as an Anti-Money-Laundering Strategy for Financial Institutions [78]	×
8	2012	Reconstructing Android User Behavior Approach Based on YAFFS2 and SQLite [79]	×
9	2014	Forensic Investigation of MySQL Database Management System [26]	×

C1: Titles, abstracts, related works and conclusions were excluded; the investigation process was either extracted from the diagram or from the main textual model.

C2: The investigation process must have a definition or activity; the investigation process must have a definition or activity to recognize the purpose, meaning, and functioning of the process.

C3: Irrelevant investigation processes not related to conducting database forensics were excluded.

C4: Include explicit and implicit investigation processes from models.

The above criteria were utilized in an attempt to avoid any missing or random process selections. For example, the investigation process model to reveal malicious database activities for an Oracle database was proposed by Wong and Edwards [12]. Their solution consists of four investigation processes, namely: *Suspending database operation, collecting data, reconstructing a database, and restoring database integrity*. Another model was developed by Wright [13] to discover and analyze intruder activities in an MSSQL server database. Wright [13] provided a solution that consists of four investigation processes, namely: *Verification, Evidence Collection, Timeline Creation, and Media Analysis*. In addition, two investigation processes were extracted from [56], which were *Identification and Collection,* to identify and collect volatile and non-volatile data.

Additionally, four investigation processes have been extracted from Fowler's [10] model to identify, verify, collect and analyze MSSQL server database incidents: *Investigation preparation, Incident verification, Artefact collection, and Artefact analysis*. In addition, two investigation processes were extracted from a model to detect and extract malicious relations among tables that were proposed by Lee and Choi [57]: *Preparation of database environment and data extraction*. The other three investigation processes, *Data acquisition, Financial data analysis, Final report and court submission*, were proposed by Choi *et al.* [58] to detect fraud statements. Furthermore, Oliver [9] proposed three investigation processes to extract, search and restore metadata that are used for investigative purposes: *Metadata extraction, Restoration, and Searchability.*

TABLE 2. Extracted database forensic investigation process from 18 models.

ID	Extracted Investigation Processes	Extracted process	C1	C2	C3	C4
1.	Suspend database operation, Collect data, Reconstruct database, Restore database integrity	4	√	√	√	√
2.	Evidence verification, Evidence Collection, Timeline Creation, Media Analysis	4	√	√	√	√
3.	Identification process, Collection process	2	√	√	√	√
4.	Investigation preparation, Incident verification, Artefact collection, Artefact analysis	4	√	√	√	√
5.	Prepare Database Environment, Extract Data process	2	√	√	√	√
6.	Data Acquisition, Financial Data Analysis, final report and court submission	3	√	√	√	√
7.	Metadata extraction, Restoration, Searchability	3	√	√	√	√
8.	Server detection, Data Collection, Investigation on Data Collected	3	√	√	√	√
9.	Setup evidence collection server, Collect files	2	√	√	√	√
10.	Identification, Artefact collection, Artefact analysis, Final Forensic Report	4	√	√	√	√
11.	Examination preparation, Collection phase, Reconstruction phase, Documentation & Presentation,	4	√	√	√	√
12.	Determine database dimension, determining acquisition method, collection of volatile artefacts, collection of non-volatile artefacts, preservation and authentication of collected data, and analysis of collected data	6	√	√	√	√
13.	Artefact collection, Forensic analysis	2	√	√	√	√
14.	Identification process, Collection process	2	√	√	√	√
15.	Collection metadata, Preservation metadata, Analysis database attacks	3	√	√	√	√
16.	Collection evidence, Reconstructing evidence	2	√	√	√	√
17.	Reconstruction process, Analysis process	2	√	√	√	√
18.	Reconstructing volatile artefacts, Recovering database schema	2	√	√	√	√

In addition, a process investigation model was introduced by Son *et al.* [17] for enterprise environments. The investigation process involves three steps: *server process detection* subsequent to an incident report, followed by *data collection* and an *investigation on the data collected.*

In addition, a tamper detection model was proposed by Tripathi and Meshram [16] to introduce the concept of tamper detection for digital evidence stored in a database. Thus, the two investigation processes highlighted in [16] are *Setup evidence collection server and file collection.*

A framework proposed by Khanuja and Adane [3] deals with the forensic analysis of a MySQL server database. It consists of four main investigation processes: *Identification, Artefact collection, Artefact analysis and Final forensic report.* Another framework was offered by Susaimanickam [59] to analyze database incidents. It consists of four investigation processes: *Examination preparation, Collection phase, Reconstruction phase, and Documentation and presentation.* In addition, six investigative processes were advocated by Fasan and Oliver [60] to investigate database incidents, namely: *Determine database dimension, Determination of acquisition method, Collection of volatile artefacts, Collection of non-volatile artefacts, Preservation and authentication of collected data, and Analysis of collected data.* Another forensic analysis model was proposed by Khanuja and Adane [22] for combining multiple pieces of evidences. Their solution offers two general processes for this purpose: *Artefact collection and Forensic analysis.* Additionally, a model was proposed by Beyers [7] to investigate a compromised database management system. Generally, it consists of two main investigation processes: the *Identification process and the Collection process.* In addition, another process model was proposed by Khanuja and Suratkar [61] to collect, preserve and analyze database metadata against database attacks. Three main investigation processes were proposed by Khanuja and Suratkar [61]: *Collection of metadata, Preservation of metadata, and Analysis of database attacks.*

The model proposed by Frühwirt *et al.* [62] is designed to reconstruct database events to detect intruder activities, via the following: a *Collection process* to gather evidence by replicating sources and a *Reconstructing evidence process* that is necessary to rebuild user activities and detect malicious activities. Similarly, an investigation process model was proposed by Adedayo and Oliver [5] to reconstruct and analyse database activity using log files via the *Reconstruction process* and *Analysis process.* Finally, Wagner *et al.*, [4] proposed a database forensic analysis model to reconstruct database activities through internal structure carving, via *Reconstructing volatile artefacts, and recovering database schema.*

Therefore, as an output of this section, fifty-four (54) investigation processes were identified and extracted from these eighteen (18) database forensic models. The proposed investigation processes are interlinked activities and tasks to achieve the goal of the Database Forensic activity. Thus, the next step merged and extracted investigation processes that were grouped based on their activities and meaning. Table 2 presents the extracted database forensic investigation process from the eighteen (18) models.

B. MERGING AND GROUPING OF THE EXTRACTED DATABASE FORENSIC INVESTIGATION PROCESSES

The fifty-four (54) extracted database forensic investigation processes are merged and grouped together based on similar activities and concepts [63], [64]. All investigation processes having similar activities and concepts are organised,

TABLE 3. The first group of organized, merged and grouped investigation processes.

Model	Similar Processes	Activity and Meaning
M1	Suspend Database Operations	Database operations are suspended, at least long enough to capture evidence of the intruder's actions. This may entail disabling new logins, terminating any or all existing sessions and disconnecting the database from users.
M2	Verification	Verifies and checks incident, isolates database server and confirms the incident.
M3	Identification	Identification process that deals with disconnecting database server from network to capture volatile data as well as prepare forensic environment and forensic techniques used to move captured data.
M10	Identification	Identification process is used to identify MySQL database files (text files, log files, binary files) and also identify MySQL utilities.
M14	Identification	The identification process is going to prepare database forensic layers and forensic methods, as well as prepare the forensic environment.
M4	Investigation Preparation	Investigation preparation identifies and prepares forensic workstations and forensic toolkits to respond to an incident and then disconnect from the database server.
M4	Incident Verification	Verifies the database incident through preliminary investigation.
M5	Prepare Database Environment	Prepares the investigation environment and obtains the right to access the database and execute the command.
M6	Data Acquisition	Secure the location of evidence and extract evidence that relates to the accounting fraud. The examiner should identify the target and verify where it is.
M8	Server Detection	Server detection is used to identify and detect the victim database server.
M9	Setup Evidence Collection Server	Preparing the investigation environment to reveal an incident.
M11	Examination Preparation	Examination preparation is used to detect a database incident, isolate a network, configure an investigation environment, identify policies, and prepare proper forensic tools **as well as making a decision** (TO DO WHAT).
M12	Determine Database Dimension	Identifies which dimension of the database has been attacked or hacked.
M12	Determining Acquisition Method	Identifies the proper acquisition methods for that dimension.

merged and grouped into separate collections. The first grouping scrutinized processes that broadly dealt with investigation preparation, incident identification, and verification. The initial merging and grouping of the extracted database forensic investigation processes is demonstrated as follows.

The *Suspension of Database Operation* in the model of Wong and Edwards [12] isolates the database server from the users in order to capture database activities, while the *Verification* process in the model of Fowler *et al.* [18] verifies and checks incidents, isolates the database server, and confirms the incident. In addition, the *Identification* process in [56] deals with disconnecting database servers from the network in order to capture volatile data. Likewise, the purpose of the *Investigation preparation* and *Incident verification* in [10] is to identify and verify database incidents through a preliminary investigation, prepare forensic workstations and forensic toolkits to respond to incidents and then disconnect the database server. Furthermore, the *Preparation of the Database Environment* proposed in Lee *et al's* [57] used to prepare the investigation environment and to obtain necessary permission to access the database and execute required commands. Additionally, the *Data Acquisition* process necessitates securing the location of evidence and extracting evidence that relates to a crime or an incident [58]. Another process of interest is *Server Detection* which is used to identify and detect the victim database server [17]. Furthermore, the *Setup Evidence Collection Server* process described in [16] is used to prepare the investigation environment to store incidents, while the *Identification* process described in [3] identifies relevant MySQL database files (text files, log files, binary files) and utilities.

Similarly, Susaimanickam [59] proposed a model for the *Examination Preparation* process, which is used to detect database incidents by cutting off the network, configuring the investigation environment, identifying policies, preparing the proper tools and also informing decision making.

In addition, Fasan and Oliver [60] proposed a model for *Determining Database Dimension* and *Acquisition Method*, which is used for identifying which dimension of the database has been attacked or hacked. Once this has been achieved, the proper acquisition methods for that dimension are then identified. An identification process model is proposed by Beyers [7] that is aimed at the database forensic layers, methods and environment. Therefore, fourteen (14) investigation processes have been organized, merged and grouped based on their activities and congruent concepts. Table 3 presents the first group of organized, merged and grouped investigation processes.

The second grouping inspected processes from the data collection perspective. Wong and Edwards [12] proposed a model for the data collection process aimed at assembling data metadata and intruder activity. Similar processes are introduced by Fowler *et al.* [18] that include *Evidence Collection* to collect evidence from the victim database server and an *Evidence Collection* process proposed by Litchfield [56] to collect volatile data from compromised database servers. Fowler [10] proposed an *Artefact Collection* model (Fowler, 2008) that is used to collect volatile and non-volatile MSSQL Server database artefacts such as log

TABLE 4. The second group of organized merged and grouped investigation processes.

Model	Process	Activity and Meaning
M1	Collecting Data	Collecting data is going to assemble data, metadata and intruder activities from the database server.
M2	Evidence Collection	Collects evidence from victim database server.
M3	Collection	Collection process uses collected volatile data from compromised database server.
M4	Artefact Collection	Artefact collection is used to collect volatile and nonvolatile MSSQL Server database artefacts such as log files, data files, a data cache, transaction logs, and log files.
M5	Extracting Data	Extraction data process is used to extract data about relationships that connect columns in database tables.
M6	Data Acquisition	Extract fraud data from the database server.
M7	Metadata Extraction	Metadata Extraction is used to extract the metadata of database dimensions that are used to determine who was authorized to perform a certain action.
M8	Data Collection	Data Collection is divided into a stage of selectively collecting files and a stage of collecting the entire files.
M9	Collecting Files	Collecting files process is used to collect Oracle files from specific locations and move them to the evidence collection server for further investigation.
M10	Artefact Collection	Artefact collection is used to collect and extract database files and metadata from compromised databases.
M11	Collection	Collect physical and digital data.
M12	Collection of Non-volatile Artefacts	Collection of nonvolatile artefacts such as database files, log files, and log transactions.
M12	Collection of Volatile Artefacts	Collect volatile artefacts such as data caches, redo log and undo log
M13	Artefact Collection	Artefact collection is used to collect volatile and nonvolatile MSSQL Server database artefacts such as log files, data files, data cache, transaction logs, log files and so on
M14	Collection	The collection process is used to collect and extract suspected database management system data and move them for further examination.
M15	Collection Metadata	This is used to collect detailed logs of SQL, MySQL, and the operating system.

files, data files, a data cache, transaction logs, and log files.

Additionally, a data Extraction process is proposed by Lee *et al.* [57] to extract data on relationships that connect columns in database tables. In addition, the *Data Acquisition* process, proposed by Choi *et al.* [58], has similar activities designed to extract fraud data from a database server. Additionally, the *Metadata Extraction* process offered by Oliver [9] to extract the metadata of the database dimension is used to determine who was authorized to perform a certain action. In addition, the *Data Collection* process presented by Son *et al.* [17] is subdivided into two stages that consist of a stage dedicated to selectively files and another stage that focuses on collecting entire files. In addition, a file collection model was introduced by Tripathi and Meshram [16] to collect Oracle files from specific locations and move them to the evidence collection server for further investigation. Furthermore, the *Artefact Collection* process was proposed by Khanuja and Adane [3] to collect and extract database files and metadata from compromised MySQL Server databases. Similarly, Susaimanickam [59] proposed a *Collection* process as a sub-process of physical and digital examination to collect physical and digital data. Moreover, the *Collection of Volatile Artefacts* and*Non-Volatile Artefacts* processes were proposed by Fasan and Oliver [60] to collect database files, log files, log transactions and also volatile artefacts such as data caches, redo log, and undo log.

Similar to the *Artefact Collection* process presented by Fowler [10], an *Artefact Collection* process was proposed by Khanuja and Adane [22]. The *Collection* process introduced by Beyers [7] allows one to collect and extract suspected database management system data and move it to a secure area for further forensic investigation, and the *Metadata Collection* process proposed by Khanuja and Suratkar [61] allows one to collect detailed multiple logs of SQL, MySQL and operating systems.

Implicitly, the preservation process is mentioned by Wong and Edwards [12] under the *Collecting Data* process to protect the integrity of data by hashing collected data. It is also mentioned in Fowler's [10] model under the *Artefact Collection* process to prevent any modification of collected data. In addition, Choi *et al.* [58] mentioned implicitly as "Secure data sources" under the *Data Acquisition* process. In [9], it was mentioned as an integrity measure to provide an exact copy of the original data. In the model proposed by Susaimanickam [59], it was described as the main preservation process in the *Physical & Digital Examination* process. Explicitly, it was mentioned in [60] along with the *Collecting Metadata* process as the preservation of metadata. Thus, the preservation process was grouped with the collection process. Therefore, sixteen (16) investigation processes have been organized, merged and grouped based on their activities and meaning. Table 4 presents the second grouping of organized, merged and

TABLE 5. The third group of organized, merged and grouped investigation processes.

Model	Process	Activity and Meaning
M1	Reconstructing Database	Reconstructing database is used to rebuild intruder activities and reveal malicious actions.
M1	Restoring Database Integrity	Restoring database integrity is used to restore database consistency.
M2	Media Analysis	Focuses on analyzing activities and revealing malicious intruders.
M4	Artefact Analysis	Artefact analysis focuses on analyzing authentication and authorization artefacts as well as configuring and versioning artefacts. Furthermore, it analyzes activity reconstruction and data recovery artefacts, which makes up the largest grouping of artefacts.
M6	Financial Data Analysis	Deep analysis of the account data and other related business data should be executed. It is used to reveal fraudulent transactions.
M7	Restoration and Searchability	Recreation of data that have been (partially) destroyed or only partially recovered.
M8	Investigation on Data Collected	Methods of investigating data can be largely divided into three types, such as investigating data collected by using an agent remotely, investigating data collected by using backup commands and investigating data collected by using the entire files.
M10	Artefact Analysis	All data acquired through incident verification and collection phases are consolidated and analyzed.
M11	Reconstruction	Rebuild database events.
M12	Analysis of Collected Data	Analyse collected data using forensic analysis tools.
M13	Forensic Analysis	Forensic analysis involves temporal detection, the determination of the time. It also involves spatial detection, the determination of where the location of the data in the database was altered.
M15	Analysis Database Attacks	Reconstruct and analyse database attacks to reveal who the attacker is., when the attack happened, where it happened and how it happened.
M16	Reconstructing Evidence	Reconstructing evidence is used to rebuild user activities and detect malicious activities.
M17	Reconstruction	Identifying changes made to a database, identifying who may be responsible for the changes, confirming what we expect to see in the database, and determining the timeline of events in the database.
M17	Analysis	Use of log analysis and/or log management tools to enhance the analysis of the volume of information that may be retrieved from log files during database forensics; the analysis process should be automated.
M18	Reconstructing Volatile Artefacts	Recover the newly introduced data from inserts and updates. Also, recover recently performed user actions (i.e., reconstructing the fact that data were inserted, deleted or updated). Additionally, discover information about changes that were canceled and undone (i.e., aborted transactions).
M18	Recovering Database Schema	Discovering the original schema, structure identifiers, identify pages from the same structure, and discover other components of the schema.

grouped investigation processes having similar activities and concepts.

The third grouping broadly focused on database reconstruction, artefact analysis and overall forensic analysis. For example, an *Analysis* process has been mentioned in several models. In the model of Wong and Edwards [12], it was used to *Reconstruct a Database* and *Restore Database Integrity* after collecting data to rebuild intruder activities along with revealing malicious actions and restoring database consistency. In addition, Fowler [18] mentioned it as different names in two models. For example, in the model of Folwer *et al.* [18], it was mentioned as part of the *Timeline Creation* and *Media Analysis processes*, while Fowler [10] described it as part of the *Artefact Analysis* process to reconstruct timeline events and analyse malicious activity. In addition, Choi *et al's.* [58] model referred to the analysis process as *Financial Data Analysis* and used it to reveal fraudulent transactions. Other models referred to the analysis process as *Restoration and Searchability* [9], where Son *et al's.*, [17] model referred to it as the *Investigation on Data Collected* process. Furthermore, Khanuja and Adane's [3] model mentioned it explicitly as *Artefact Analysis*. It is mentioned implicitly as part of the *Reconstruction* phase along with the physical and digital examination process in [59]. Other models referred to the Analysis process as *Forensic Analysis* [22], *Analysing Database Attacks* [61], *Reconstructing Evidence* [62], *Reconstruction* [5], and *Reconstructing Volatile Artefacts* [4].

Some models highlighted certain investigation processes explicitly or implicitly. For example, the *Reconstruction* process was mentioned explicitly in several models [4], [5], [12], [62], while it was mentioned implicitly under the *Artefact analysis* process in [10] and also mentioned implicitly in Choi *et al.'s* [58] model under the *Financial Data Analysis* process. In addition, it is mentioned as the *Restoration* process in [9], which used it to refer to the recreation of data that have been (partially) destroyed or only partially recovered by a forensic capture process. Moreover, it is mentioned implicitly under the *Physical & digital examination* process in [59]. Therefore, it will be merged and grouped along with the *Analysis* process because it is part of the process that is used to rebuild timeline events during the analysis process. Therefore, seventeen (17) investigation processes have been organized, merged and grouped based on their activities and concepts. Table 5 presents the third grouping of organized, merged and grouped investigation processes with similar concepts and activities.

An additional refinement of the identified processes examined the similarities from the broad perspective of documentation which would include presentation and reporting. The *Documentation and Presentation* process was mentioned by Susaimanickam [59] to document and present whole

TABLE 6. The fourth group of organized merged and grouped investigation processes.

Model	Process	Activity and Meaning
M11	Documentation and Presentation	Document and present whole investigation stage and submit the results to the court.
M6	Final Report and Court Submission	The investigation results should be reconstructed and written up as a report to submit to the court.
M10	Final Forensic Report	Includes the investigation results and investigation steps.

TABLE 7. Mapping process of forensic investigation group 1

Models\Process	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Suspend database operation	√																	
Verification		√																
Identification			√							√				√				
Investigation preparation				√														
Incident verification				√														
Prepare Database Environment					√													
Data Acquisition						√												
Server detection								√										
Setup evidence collection server									√									
Examination preparation											√							
Determine database dimension												√						
Determining acquisition method												√						

investigation stages and submit the results to the court. In addition, Choi *et al.* [58] offered it as a final report and court submission, whereas Khanuja and Adane [3] mentioned it as a final forensic report. Table 6 displays the fourth group of organized and merged investigation processes.

In the summary of this section, fifty-four (54) extracted database forensic investigation processes have been organized, merged, and grouped based on their activities and concepts. Four forensic investigation groups have been highlighted. Every group has similar processes in meaning and activities, regardless of actual naming. Every group has a common investigation process amongst other processes. Therefore, the common investigation process for every group will be proposed in the next section.

C. PROPOSED COMMON INVESTIGATION PROCESS FOR DATABASE FORENSIC EXAMINATION

Four forensic investigation groups have been highlighted in Section 3.3 based on similarities in concept and activities. This section aims to propose a common investigation process for every group. The mapping process has been adopted to propose common investigation processes. The investigative process, which has a higher frequency in the group, will be a candidate and be referred to as a common investigation process. From these phases, four common database forensic investigation processes were derived: *Identification process, Artefact Collection & Preservation, Artefact Analysis, and Documentation and presentation process*. Since some of these processes overlap, it was necessary to take into account the activities performed in each of the investigative processes and to not rely solely on naming conventions [63], [64]. The mapping process is adopted to select the more frequent investigation process for every forensic investigation group. *Identification process* was identified as an investigation process by fifteen (15) investigation processes in Group 1. It appears three times in three models [56], [3], [7]. Table 7 shows the mapping process of forensic investigation in Group 1.

In addition, *Artefact Collection* was identified as an investigative process by sixteen (16) investigation processes in Group 2. It appears three times in three models [10], [22], [3]. Table 8 shows the mapping process of forensic investigation in Group 2.

The *Artefact analysis* process was identified as a common investigation process among seventeen (17) investigation processes in forensic investigation Group 2. It appears two times in two different models [10], [3]. Table 9 shows the mapping process of forensic investigation in Group 3.

In addition, the *Documentation and Presentation* process was identified as a common investigation process from forensic investigation Group 4. This process has rarely been mentioned by researchers. Therefore, the author suggests using this process due to its generality. Fig 3 displays the proposed common investigation process for a database forensic investigation.

In summary, four (4) common investigation processes have been proposed based on their frequency or generality. Three common investigation processes have been proposed based on their frequency (*Identification, Artefact Collection, and Artefact Analysis*). However, the *Documentation and Presentation process* has been proposed based on its generality. It is more general than other investigation processes in

TABLE 8. Mapping process of forensic investigation group 2.

Models\Process	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Collecting data	√																	
Evidence Collection		√																
Collection process			√								√							
Artefact collection				√						√			√					
Extraction data process					√													
Data Acquisition						√												
Metadata Extraction							√											
Data Collection								√										
Collect files									√									
Collection phase											√							
a collection of non-volatile artefacts												√						
a collection of volatile artefacts												√						
Collection metadata															√			

TABLE 9. Mapping process of forensic investigation group 3.

Models\Process	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
Reconstructing d0atabase	√																	
Restoring database integrity	√																	
Media Analysis		√																
Artefact analysis				√						√								
Financial Data Analysis						√												
Restoration and Searchability							√											
Investigation on Data Collected								√										
Reconstruction phase											√							
Analysis of collected data												√						
Forensic analysis													√					
Analysis of database attacks															√			
Reconstructing evidence																√		
Reconstruction process																	√	
Analysis process																	√	
Reconstructing volatile artefacts																		√
Recovering database schema																		√

the forensic investigation. The next section will combine the various definitions of database forensic investigation processes and reconcile them to the abstract definition.

D. RECONCILIATION OF INVESTIGATION PROCESS DEFINITIONS

Explicit definitions are important in science. Precise definitions help to give a clear meaning of concepts. Differences between definitions are reconciled in this phase.

In choosing or synthesizing, the common investigation process definitions to be used are shortlisted in this phase. If there is a conflict in the definition between two or more sources, then the definitions are reconciled based on common definitions and environment applicability. Some models explicitly omit defining their concept; in such cases, they do not provide any input to the reconciliation process. Therefore, this section aims to harmonize and reconcile the investigation process shortlisted definition. The same reconciliation approach process has been implemented in previous research [47], [53]. Similar definitions are reconciled and harmonized into one abstract definition.

For example, the proposed Identification process defined by Litchfield [56] "deals with disconnecting the database server from the network to capture volatile data as well as prepare forensic environment and forensic techniques to moved

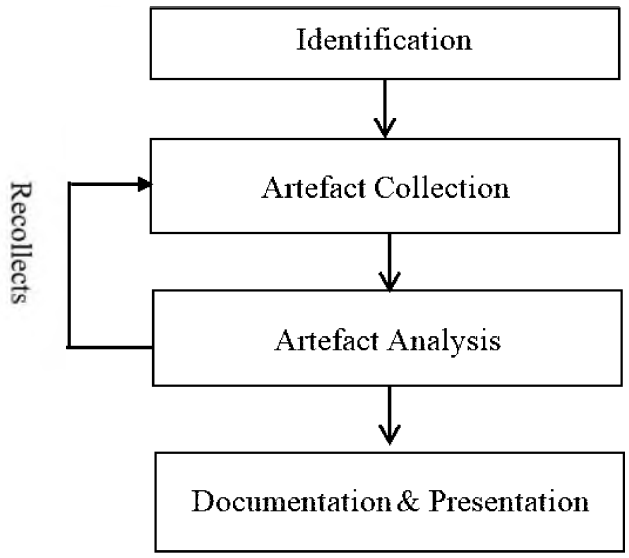

FIGURE 3. Proposed common database forensic investigation processes (CDBFIP).

captured data in", whereas [3] defined it as "Identification process used to identify MySQL database files (text files, log files, binary files) and also identify MySQL utilities". On the other hand, Beyers [7] defined it as "Identification process for preparing database forensic layers and forensic methods, as well as preparing forensic environment (found environment, clean environment)". Therefore, the definition of Litchfield's [56] is specific to isolating the database server along with preparing the forensic environment and forensic techniques, whereas the definition by Khanuja and Adane [3] is too specific to deal with preparing MySQL database files and utilities. However, the definition of Beyers [7] is similar to [56]. Thus, Byers's definition and Litchfield's definition will be reconciled and harmonized in one abstract definition:

"Identification process is the first Database Forensic investigation process that is used to prepare a clean database forensic investigation environment and trust forensic techniques, as well as allow the investigation team to isolate the database server enough from the network to prevent users from tampering and capturing volatile and non-volatile data".

In addition, "Artefact Collection" was defined by three different researchers. Fowler [10] defined it as "collection of volatile and non-volatile MSSQL Server database artefacts such as log files, data files, data cache, transaction logs, and log files". Whereas, Khanuja and Adane [3] defined it as "collection and extraction of database files and metadata from a compromised database". In addition, Khanuja and Adane [22] defined it as a "collection of volatile and nonvolatile MSSQL Server database artefacts such as log files, data files, a data cache, transaction logs, and log files". From an analysis viewpoint, [10] definition is too specific to MSSQL server database, whereas [22] defined the Artefact Collection process from two perspectives. The first definition focused on non-volatile artefacts, whereas the second definition was borrowed from Fowler [10]. Consequently, a harmonized and reconciled version of this definition is as follows: "Artefact Collection process is the second Database Forensic investigation process that is used to collect and preserve volatile and non-volatile artefacts from the suspect database using trust forensic techniques".

Additionally, the Artefact Analysis process has two definitions. Fowler [10] defined it as "*analysing authentication and authorization events, as well as configuring and versioning artefacts. Furthermore, analysing activity reconstruction and data recovery artefacts, which make up the largest grouping of artefacts*", whereas Khanujha and Adane [3] defined it as "*analysing all data that are acquired through the incident verification and collection phases*". Clearly, [10] definition is brother than [3]. Thus, we harmonize and reconcile these definitions as "Artefact Analysis process is the third Database Forensic investigation process that is used to analyze acquired data, activity reconstruction and data recovery using special forensic techniques to reveal who is tampering, when and where the tampering happened and how the tampering happened."

Finally, the Documentation and Presentation process was defined by Susaimanickam [59] as "Documents and presents whole investigation stages, and submit the results to the court". However, the authors reconciled it as "Documentation and Presentation investigation process is the fourth Database Forensic investigation process that is used to document and present the investigation stages and submit the results to the court".

In summary, the proposed common investigation process definitions have been reconciled and harmonized into abstract definitions. The next section concentrates on comparing the proposed common investigation processes against other existing database forensic models to validate and verify the completeness and generality of proposed common investigation processes.

IV. VALIDATION OF PROPOSED COMMON INVESTIGATION PROCESSES

The version of the proposed common investigation process will now be validated and improved to make it complete and coherent.

The derived processes, of the proposed common investigation process, are validated and compared to processes from other similar existing domain models [79]–[81]. For this purpose, we used a set of nine popular and widely used database forensic models in a validation set. Every investigation process, in each of the models, can be appropriately derived from an investigation process. Where required, we modified the proposed investigation process to ensure that it can represent all models in the validation set. This validation ensures that the database forensics domain can be represented in each of the models in the validation set. Where applicable, proposed investigation processes were modified to ensure that every model can be represented. No processes were changed, e.g., Identification-process, Artefact Collection-process,

Artefact Analysis process, and Documentation & Presentation process in the proposed common investigation processes.

A. COMPARISON AGAINST EFFICIENT MODEL FOR DETECTING DATA AND DATA SCHEMES TAMPERING FOR THE PURPOSE OF VALID FORENSIC ANALYSIS

A detection model, proposed by Azemović and Mušić [72], aimed at detecting database system tampering. It offered an efficient approach to providing audit logs for transaction processing systems that can effectively and efficiently detect tampering. SQL coding and cryptographic techniques such as strong cryptographic hashing were used to provide authentication codes on collected data. Thus, this model implicitly provides collection process. the collection process is used to extract data from log files and secure the validity of the collected data. Therefore, the proposed common investigation process "Artefact Collection" covered the existing investigation process in the model "Collection process".

B. COMPARISON AGAINST METHODS FOR EFFICIENT DIGITAL EVIDENCE COLLECTION OF BUSINESS PROCESSES

Three techniques were proposed by Azemović and Mušić [73] to collect digital evidence from database systems: Triggers, Log file backup, and Replications. Triggers are very powerful elements in a database system that, if used properly, can be very helpful in detecting data modifications. In addition, the log file backup concept can be used on a regular basis for collecting and keeping digital evidence of users' activity. Additionally, the replication method is a set of technologies for copying and distributing data and database objects from one environment to another; it then synchronizes between databases to maintain consistency. Therefore, this model offered a Collection process to collect evidence using three methods against user behavior. Thus, the proposed Artefact collection process has covered the Collection process.

C. COMPARISON WITH ASSEMBLING METADATA FOR DATABASE FORENSICS

A collection process model has been offered by Beyers *et al.* [82] to locate the key evidence and maintain the integrity and reliability of the evidence. This model describes a database forensic method that transforms a DBMS into the required state for a database forensic investigation. The method segments a DBMS into four abstract layers that separate the various levels of DBMS metadata and data. Therefore, the proposed common investigation process "Artefact Collection" covered the existing investigation process "Collection process" in this model.

D. COMPARISON AGAINST ORACLE FORENSICS PART 2: LOCATING DROPPED OBJECTS

The model offered by Litchfield [83] allows the investigator to recover evidence directly from the data files of the compromised server, although an attacker may drop objects. Several Oracle views and tables assist the investigator in locating dropped objects such as OBJ$, SOURCE$, IDL_UB1$, IDL_CHAR$, and RECYCLEBIN$ tables. Therefore, the proposed investigation process "Artefact analysis" has been covered by the Recover evidence process that is offered in this model. In addition, the activities of this process, such as reconstructing events and timelines, are covered as well.

E. COMPARISON AGAINST "ON THE COMPLETENESS OF RECONSTRUCTED DATA FOR DATABASE FORENSICS"

A specific reconstruction process was offered by [84] to rebuild deleted relations or deleted records from the database to introduce evidence against database incidents. A reconstruction algorithm is used through a reconstruction process to allow the investigators to provide evidence against database modifications. Thus, this process and its activities were enclosed by the proposed investigation process "Artefact Analysis".

F. COMPARISON AGAINST DATABASE TAMPERING AND DETECTION OF DATA FRAUD

Analysis process was proposed by Gawali and Gupta [85] to detect and analyze malicious activity. Therefore, the forensic detection algorithm and forensic analysis algorithm have been offered in this model to detect corruption events and determine when and where database tampering occurred. The outcome of this model has covered the proposed investigation process "Artefact Analysis".

G. COMPARISON AGAINST ENRICHING FORENSIC ANALYSIS PROCESS FOR TAMPERED DATA IN DATABASE

A forensic analysis process for tampered data in the database was offered by Abhonkar and Kanthe [68] to detect and analyze database tampering. The Tiled Bitmap Algorithm was used to reveal when the tampering occurred and what data were altered. The outcome of this model covered the proposed investigation process "Artefact Analysis" with respect to activities and tasks.

H. COMPARISON AGAINST AN APPROACH TO EXAMINING THE METADATA AND DATA OF A DATABASE MANAGEMENT SYSTEM

This process model was proposed by Beyers *et al.* [75] to conduct a forensic examination on a DBMS. A Preparation process was offered in this model. The investigator prepares the forensic examination environment, for example, setting up the Ubuntu operating system and PostgreSQL DBMS, and also prepares the forensic comparison tool, which can be used by the investigator to reveal database tampering. Therefore, the preparation process was covered by the proposed common investigation process Identification process.

I. COMPARISON AGAINST FORENSIC ANALYSIS OF DATABASE TAMPERING

The forensic analysis process model of database tampering was proposed by Pavlou and Snodgrass [86] to analyze corruption events in database systems. Several forensic

TABLE 10. A comparative summary: proposed common investigation process and existing database forensic models.

ID	Processes in Compared Model [68-76]	Processes in Proposed Model Process	Status
M1	Identification process Collection process	Identification process Artefact Collection	Supported
M2	Collection process	Artefact Collection	Supported
M3	Collection process	Artefact Collection	Supported
M4	Recover evidences process	Artefact Analysis	Supported
M5	Reconstruction process	Artefact Analysis	Supported
M6	Analysis process	Artefact Analysis	Supported
M7	Forensic analysis process	Artefact Analysis	Supported
M8	Preparation process Collection process	Identification process Artefact Collection	Supported
M9	Forensic analysis process	Artefact Analysis	Supported

analysis algorithms were introduced in this model and are to be applied after detecting a database intrusion in order to answer questions regarding when and where tampering occurred. Therefore, the forensic analysis process that was proposed in this model, along with the activities, has been covered by the proposed common investigation process "Artefact Analysis".

In summary, the proposed common investigation processes for database forensics was validated against nine existing database forensic models. The proposed common investigation process covers the investigation processes that are derived from the compared models. Table 10 displays the proposed common investigation process in comparison to the database forensic models.

V. CONCLUSION

The escalating dependency of IoT devices to retain and manipulate data stored in databases emphasizes the need to develop an in-depth understanding of the nuances associated with conducting a database forensic investigation (Kebande and Ray, 2016; Teing et al., 2017). This need will be more pronounced as IoT becomes the norm in both civilian and military settings, such as Internet of Battlefield/Military Things.

In this paper, we identified fifty-four (54) investigation processes that were extracted from eighteen (18) database forensic models. These processes were then grouped and refined based on common objectives to identify mutual investigation processes for the database forensic domain. This analysis resulted in the Common Database Forensic Investigation Processes (CDBFIP). Based on a thorough investigation of identified processes and models, the CDBFIP reveals that the database forensic investigation process has four common processes: identification, artefact collection and preservation process, artefact analysis process, and documentation and presentation process. We then used nine existing database forensic investigation models to validate the completeness of the CDBFIP model.

Future work will include examining the concepts and relationships in each of the identified processes in the CDBFIP by applying a software engineering approach known as a metamodel, as well as evaluating the CDBFIP using real-world IoT datasets. Future work will also investigate the application of probabilistic and measurement science techniques into specific IoT database extraction solutions. The idea is to start to be able to quantify confidence in terms of data relevance, origination and tool extraction performance. We will also be seeking out opportunities to implement and evaluate the proposed CDBFIP in a real-world IoT infrastructures, which will allow us to validate and refine the model.

ACKNOWLEDGEMENTS

The authors would like to thank the experts at Cybersecurity Malaysia for their assistance and evaluation of this work. We also acknowledge Professor John Walker for his assistance and evaluation of this work.

REFERENCES

[1] K. J. Berman, W. B. Glisson, and L. M. Glisson, "Investigating the impact of global positioning system (GPS) evidence in court cases," presented at the Hawaii Int. Conf. Syst. Sci. (HICSS), Kauai, HI, USA, Jan. 2015.
[2] J. McMillan, W. B. Glisson, and M. Bromby, "Investigating the increase in mobile phone evidence in criminal activities," presented at the Hawaii Int. Conf. Syst. Sci. (HICSS), Wailea, Hawaii, 2013.
[3] H. K. Khanuja and D. D. Adane, "A framework for database forensic analysis," *Comput. Sci. Eng., Int. J.*, vol. 2, no. 3, p. 27, 2012.
[4] J. Wagner, A. Rasin, and J. Grier, "Database forensic analysis through internal structure carving," *Digit. Invest.*, vol. 14, pp. S106–S115, Aug. 2015.
[5] O. M. Adedayo and M. S. Olivier, "Ideal log setting for database forensics reconstruction," *Digit. Invest.*, vol. 12, pp. 27–40, Mar. 2015.
[6] C. Hooper, B. Martini, and K.-K. R. Choo, "Cloud computing and its implications for cybercrime investigations in Australia," *Comput. Law Secur. Rev.*, vol. 29, no. 2, pp. 152–163, 2013.
[7] H. Q. Beyers, "Database forensics: Investigating compromised database management systems," Univ. Pretoria, Pretoria, South Africa, Tech. Rep., 2014.
[8] S. A. Razak, S. H. Othman, A. A. Aldolah, and M. A. Ngadi, "Conceptual investigation process model for managing database forensic investigation knowledge," *Res. J. Appl. Sci., Eng. Technol.*, vol. 12, no. 4, pp. 386–394, 2016.
[9] M. S. Olivier, "On metadata context in database forensics," *Digit. Invest.*, vol. 5, pp. 115–123, Mar. 2009.
[10] K. Fowler, *SQL Server Forensic Analysis*. London, U.K.: Pearson Education, 2008.
[11] O. Fasan and M. Olivier, "Reconstruction in database forensics," in *Advances in Digital Forensics VIII*. Pretoria, South Africa: Springer, 2012, pp. 273–287.
[12] D. Wong and K. Edwards, "System and method for investigating a data operation performed on a database," U.S. Patent 0 289 187 A1, Dec. 29, 2005.
[13] P. M. Wright-GSEC and G. GCFW, "Oracle database forensics using LogMiner," Tech. Rep., 2005. [Online]. Available: https://www.giac.org/paper/gcfa/159/oracle-database-forensics-logminer/105140
[14] D. Litchfield, "Oracle forensics part 2: Locating dropped objects," NGSSoftw. Insight Secur. Res. (NISR) Pub., Next Generat. Secur. Softw., Manchester, U.K., Tech. Rep., 2007.
[15] P. M. Wright and D. Burleson, *Oracle Forensics: Oracle Security Best Practices*. New York, NY, USA: Rampant TechPress, 2008.
[16] S. Tripathi and B. B. Meshram, "Digital evidence for database tamper detection," *J. Inf. Secur.*, vol. 3, pp. 113–121, Apr. 2012.
[17] N. Son, K.-G. Lee, S. Jeon, H. Chung, S. Lee, and C. Lee, "The method of database server detection and investigation in the enterprise environment," in *Secure and Trust Computing, Data Management and Applications*. Loutraki, Greece: Springer, 2011, pp. 164–171.

[18] K. Fowler, G. Gold, and M. MCSD, "A real world scenario of a SQL server 2005 database forensics investigation," Inf. Secur. Reading Room Paper. SANS Institute. Tech. Rep., 2007.
[19] A. Basu. (2006). *Forensic Tamper Detection in SQL Server*. [Online]. Available: http://www. sqlsecurity.com/chipsblog/archivedposts
[20] K. E. Pavlou and R. T. Snodgrass, "Generalizing database forensics," *ACM Trans. Database Syst.*, vol. 38, p. 12, Jun. 2013.
[21] K. E. Pavlou and R. T. Snodgrass, "Achieving database information accountability in the cloud," in *Proc. IEEE 28th Int. Conf. Data Eng. Workshops (ICDEW)*, Apr. 2012, pp. 147–150.
[22] H. K. Khanuja and D. S. Adane, "Forensic analysis of databases by combining multiple evidences," *Int. J. Comput. Technol.*, vol. 7, no. 3, pp. 654–663, 2013.
[23] P. Frühwirt, M. Huber, M. Mulazzani, and E. R. Weippl, "InnoDB database forensics," in *Proc. 24th IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Apr. 2010, pp. 1028–1036.
[24] P. Frühwirt, P. Kieseberg, S. Schrittwieser, M. Huber, and E. Weippl, "InnoDB database forensics: Reconstructing data manipulation queries from redo logs," in *Proc. 7th Int. Conf. Availability. Rel. Secur. (ARES)*, Aug. 2012, pp. 625–633.
[25] P. Frühwirt, P. Kieseberg, S. Schrittwieser, M. Huber, and E. Weippl, "InnoDB database forensics: Enhanced reconstruction of data manipulation queries from redo logs," *Inf. Secur. Tech. Rep.*, vol. 17, no. 4, pp. 227–238, 2013.
[26] A. C. Lawrence, "Forensic investigation of MySQL database management system," Univ. Cork, Cork, U.K., Tech. Rep., 2014.
[27] O. M. Adedayo, "Reconstruction in database forensics." Dept. Comput. Sci., Univ. Pretoria, Hatfield, U.K., Tech. Rep., 2015.
[28] D. Quick, B. Martini, and R. Choo, *Cloud Storage Forensics*. Amsterdam, The Netherlands: Syngress, 2013.
[29] A. Azfar, K.-K. R. Choo, and L. Liu, "Forensic taxonomy of android productivity apps," *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 3313–3341, 2016.
[30] N. D. W. Cahyani, B. Martini, K.-K. R. Choo, and A. Al-Azhar, "Forensic data acquisition from cloud-of-things devices: Windows Smartphones as a case study," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 14, p. e3855, 2017.
[31] F. Daryabar, A. Dehghantanha, and K.-K. R. Choo, "Cloud storage forensics: MEGA as a case study," *Austral. J. Forensic Sci.*, vol. 49, no. 3, pp. 344–357 , 2016.
[32] Y.-Y. Teing, A. Dehghantanha, K.-K. R. Choo, and L. T. Yang, "Forensic investigation of P2P cloud storage services and backbone for IoT networks: BitTorrent Sync as a case study," *Comput. Elect. Eng.*, vol. 58, pp. 350–363, Feb. 2016.
[33] N. H. Ab Rahman, W. B. Glisson, Y. Yang, and K.-K. R. Choo, "Forensic-by design framework for cyber physical cloud systems," *IEEE Cloud Comput.*, vol. 3, no. 1, pp. 50–59, Jan./Feb. 2016.
[34] A. Azfar, K. K. R. Choo, and L. Liu, "An android communication app forensic taxonomy," *J. Forensic Sci.*, vol. 61, no. 5, pp. 1337–1350, 2016.
[35] Q. Do, B. Martini, and K. K. R. Choo, "Is the data on your wearable device secure? An Android Wear smartwatch case study," *Softw., Pract. Exper.*, vol. 47, no. 3, pp. 391–403, 2016.
[36] D. Quick and K.-K. R. Choo, "Big forensic data reduction: Digital forensic images and electronic evidence," *Cluster Comput.*, vol. 19, no. 2, pp. 723–740, 2016.
[37] T. Y. Yang, A. Dehghantanha, K.-K. R. Choo, and Z. Muda, "Windows instant messaging app forensics: Facebook and Skype as case studies," *PLoS ONE*, vol. 11, no. 3, p. e0150300, 2016.
[38] Q. Do, B. Martini, and K.-K. R. Choo, "A forensically sound adversary model for mobile devices," *PLoS ONE*, vol. 10, no. 9, p. e0138449, 2015.
[39] D. Quick and K.-K. R. Choo, "Impacts of increasing volume of digital forensic data: A survey and future research challenges," *Digit. Invest.*, vol. 11, no. 4, pp. 273–294, 2014.
[40] D. Quick and K.-K. R. Choo, "Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata?" *Digit. Invest.*, vol. 10, no. 3, pp. 266–277, 2013.
[41] Q. Do, B. Martini, and K.-K. R. Choo, "A cloud-focused mobile forensics methodology," *IEEE Cloud Comput.*, vol. 2, no. 4, pp. 60–65, Jul./Aug. 2015.
[42] F. Immanuel, B. Martini, and K. K. R. Choo, "Android cache taxonomy and forensic process," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Aug. 2015, pp. 1094–1101.
[43] B. Martini and K.-K. R. Choo, "Remote programmatic vCloud forensics: A six-step collection process and a proof of concept," in *Proc. IEEE 13th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Sep. 2014, pp. 935–942.
[44] B. Martini and K.-K. R. Choo, "Distributed filesystem forensics: XtreemFS as a case study," *Digit. Invest.*, vol. 11, no. 4, pp. 295–313, 2014.
[45] N. H. A. Rahman, N. D. W. Cahyani, and K.-K. R. Choo, "Cloud incident handling and forensic-by-design: Cloud storage as a case study," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 14, p. e3868, 2017.
[46] W. K. Hauger and M. S. Olivier, "The state of database forensic research," in *Proc. Inf. Secur. South Africa (ISSA)*, Aug. 2015, pp. 1–8.
[47] A. M. R. Al-Dhaqm, S. H. Othman, S. A. Razak, and A. Ngadi, "Towards adapting metamodelling technique for database forensics investigation domain," in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, Aug. 2014, pp. 322–327.
[48] A. Al-Dhaqm, S. Razak, S. H. Othman, A. Ngadi, M. N. Ahmed, and A. A. Mohammed, "Development and validation of a database forensic metamodel (DBFM)," *PLoS ONE*, vol. 12, no. 2, p. e0170793, 2017.
[49] A. Ali, S. A. Razak, S. H. Othman, A. Mohammed, and F. Saeed, "A metamodel for mobile forensics investigation domain." *PLoS ONE*, vol. 12, no. 4, p. e0176223, 2017.
[50] R. H. Von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research." *MIS Quart.*, vol. 28, no. 1, pp. 75–105, 2004.
[51] S. H. Othman and G. Beydoun, "Metamodelling approach to support disaster management knowledge sharing," in *Proc. 21st Australasian Conf. Inf. Syst. (ACIS)*, 2010, pp. 1–10.
[52] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decision Support Syst.*, vol. 15, no. 4, pp. 251–266, 1995.
[53] G. Beydoun *et al.*, "FAML: A generic metamodel for MAS development," *IEEE Trans. Softw. Eng.*, vol. 35, no. 6, pp. 841–863, Nov. 2009.
[54] S. Kelly and R. Pohjonen, "Worst practices for domain-specific modeling," *IEEE Softw.*, vol. 26, no. 4, pp. 22–29, Jul. 2009.
[55] A. Ali, S. A. Razak, S. H. Othman, and A. Mohammed, "Extraction of common concepts for the mobile forensics domain," in *Proc. Int. Conf. Rel. Inf. Commun. Technol.*, 2017, pp. 141–154.
[56] D. Litchfield, "Oracle forensics part 4: Live response," Tech. Rep., 2007. [Online]. Available: https://www.nccgroup.trust/uk/our-research/oracle-forensics-part-4-live-response/
[57] D. Lee, J. Choi, and S. Lee, "Database forensic investigation based on table relationship analysis techniques," in *Proc. 2nd Int. Conf. Comput. Sci. Appl. (CSA)*, Dec. 2009, 2009.
[58] J. Choi, K. Choi, and S. Lee, "Evidence investigation methodologies for detecting financial fraud based on forensic accounting," in *Proc. 2nd Int. Conf. Comput. Sci. Appl. (CSA)*, Dec. 2009, pp. 1–6.
[59] R. Susaimanickam, "A workflow to support forensic database analysis," Dept. Inf. Technol., Murdoch Univ., Perth, WA, USA, Tech. Rep., 2012.
[60] O. M. Fasan and M. S. Olivier, "On dimensions of reconstruction in database forensics," in *Proc. WDFIA*, 2012, pp. 97–106.
[61] H. Khanuja and S. S. Suratkar, "'Role of metadata in forensic analysis of database attacks,'" in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, Feb. 2014, pp. 457–462.
[62] P. Frühwirt, P. Kieseberg, K. Krombholz, and E. Weippl, "Towards a forensic-aware database solution: Using a secured database replication protocol and transaction management for digital investigations," *Digit. Invest.*, vol. 11, no. 4, pp. 336–348, 2014.
[63] A. Basu, "Forensic tamper detection in SQL server," Tech. Rep., 2006. [Online]. Available: http://amitfrombangalore.blogspot.com/2015/08/forensic-tamper-detection-in-sql-server.html
[64] D. Litchfield, "Oracle forensics part 5: Finding evidence of data theft in the absence of auditing," NGSSoftw. Insight Secur. Res. (NISR), Next Generat. Secur. Softw. Ltd., Sutton, U.K., Tech. Rep., 2007.
[65] G. T. Lee, S. Lee, E. Tsomko, and S. Lee, "Discovering methodology and scenario to detect covert database system," in *Proc. Future Generat. Commun. Netw. (FGCN)*, Dec. 2007, pp. 130–135.
[66] P. D. Abhonkar and A. Kanthe, "Enriching forensic analysis process for tampered data in database," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 5, pp. 5078–5085, 2012.
[67] H. Q. Beyers, M. S. Olivier, and G. P. Hancke, "Arguments and methods for database data model forensics," in *Proc. 7th Int. Workshop Digit. Forensics Incident Anal.*, 2012, p. 1.
[68] M. K. Kambire, P. H. Gaikwad, S. Y. Gadilkar, and Y. A. Funde, "An improved framework for tamper detection in databases," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 1, pp. 57–60, 2015.
[69] P. M. Wright, "Oracle database forensic using log-miner," in *Proc. Jan. 10th 2005 London Jun. 2004 Conf.*, Jan. 2005, pp. 1–40.
[70] J. Azemović and D. Mušić, "Efficient model for detection data and data scheme tempering with purpose of valid forensic analysis," in *Proc. Int. Conf. Comput. Eng. Appl. (ICCEA)*, 2009, pp. 83–89.

[71] J. Azemovic and D. Music, "Methods for efficient digital evidences collecting of business proceses and users activity in eLearning enviroments," in *Proc. Int. Conf. e-Edu., e-Bus., e-Manag., e-Learn.*, Jan. 2010, pp. 126–130.
[72] F. Fatima, "Detecting database attacks using computer forensics tools," Dept. Comput. Sci., Texas A&M Univ. Corpus Christi, Corpus Christi, TX, USA, Tech. Rep., 2011.
[73] H. Beyers, M. S. Olivier, and G. P. Hancke, "An approach to examine the metadata and data of a database management system by making use of a forensic comparison tool," in *Proc. ISSA*, 2011, pp. 1–6.
[74] P. K. Panigrahi, "A framework for discovering internal financial fraud using analytics," in *Proc. Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Jun. 2011, pp. 323–327.
[75] D. A. Flores, O. Angelopoulou, and R. J. Self, "Combining digital forensic practices and database analysis as an anti-money laundering strategy for financial institutions," in *Proc. 3rd Int. Conf. Emerg. Intell. Data Web Technol. (EIDWT)*, Sep. 2012, pp. 218–224.
[76] M. Xu *et al.*, "A reconstructing android user behavior approach based on YAFFS2 and SQLite," *J. Comput.*, vol. 9, no. 10, pp. 2294–2302, 2014.
[77] S. R. Selamat, R. Yusof, and S. Sahib, "Mapping process of digital forensic investigation framework," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 10, pp. 163–169, 2008.
[78] Y. Yusoff, R. Ismail, and Z. Hassan, "Common phases of computer forensics investigation models," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 3, pp. 17–31, 2011.
[79] R. G. Sargent, "Verification and validation of simulation models," presented at the 37th Conf. Winter Simulation, Orlando, FL, USA, 2005.
[80] R. G. Sargent, "Verification and validation of simulation models," *J. Simul.*, vol. 7, pp. 12–24, Feb. 2013.
[81] R. G. Sargent, "Model verification and validation," in *Modeling and Simulation in the Systems Engineering Life Cycle*. London, U.K.: Springer, 2015, pp. 57–65.
[82] H. Beyers, M. Olivier, and G. Hancke, "Assembling metadata for database forensics," in *Advances in Digital Forensics VII*. Orlando, FL, USA: Springer, 2011, pp. 89–99.
[83] D. Litchfield, "Oracle forensics part 2: Locating dropped objects," NGSSoftware Insight Security Research (NISR), Tech. Rep., 2007.
[84] O. M. Adedayo and M. S. Olivier, "On the completeness of reconstructed data for database forensics," in *Digital Forensics and Cyber Crime*. Pretoria, South Africa: Univ. Pretoria, 2012, pp. 220–238.
[85] P. P. Gawali and D. S. R. Gupta, "Database tampering and detection of data fraud by using the forensic scrutiny technique," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 2, pp. 439–446, 2013.
[86] K. E. Pavlou and R. T. Snodgrass, "Forensic analysis of database tampering," *ACM Trans. Database Syst.*, vol. 33, no. 4, p. 30, 2008.

ARAFAT AL-DHAQM received the B.S. degree in computer science and information system from the Technology University of Iraq in 2002, and the M.S. degree in information security from Universiti Teknologi Malaysia, Malaysia, in 2013, where he is currently pursuing the Ph.D. degree in computer science. He was a Lecturer with the Aden Community College, Yemen. His current research interest includes propose high level model (metamodel) for database forensic investigation field. He is under the supervision of Prof. M. Dr. S. A. Razak.

SHUKOR RAZAK is currently an Associate Professor at Universiti Teknologi Malaysia. He has authored and co-authored many journals and conference proceedings at national and international levels. His research interests are on the security issues for mobile ad hoc networks, mobile IPv6, vehicular ad hoc network, and network security. He also actively conducts several researches in digital forensic investigation, wireless sensor networks, and cloud computing.

SITI HAJAR OTHMAN received the Ph.D. degree from the University of Wollongong, Australia. She is currently a Senior Lecturer with the Department of Computer Science, Universiti Teknologi Malaysia, Malaysia. Her current research interests include conceptual modeling, metamodeling, disaster management, information security, computer forensic, knowledge retrieval, disaster recovery, and business continuity planning.

KIM-KWANG RAYMOND CHOO (SM'15) holds the Cloud Technology Endowed Professorship with the Department of Information Systems and Cyber Security, The University of Texas at San Antonio, and has a courtesy appointment at the Department of Electrical and Computer Engineering. He is a fellow of the Australian Computer Society. He was named the Cybersecurity Educator of the Year - APAC in 2016 (the Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and he led his team to win the Digital Forensics Research Challenge 2015 organized by the Germany's University of Erlangen-Nuremberg. He was a recipient of various awards, including the ESORICS 2015 Best Paper Award, the 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, the Fulbright Scholarship in 2009, the 2008 Australia Day Achievement Medallion, and the British Computer Society's Wilkes Award in 2008. He serves on the Editorial Board of *Cluster Computing*, *Digital Investigation*, the IEEE Access, the IEEE Cloud Computing, the *IEEE Communications Magazine*, *Future Generation Computer Systems*, the *Journal of Network and Computer Applications*, and *PLoS ONE*. He also serves as the Special Issue Guest Editor of *ACM Transactions on Embedded Computing Systems* (2017), *ACM Transactions on Internet Technology* (2016), *Digital Investigation* (2016), *Future Generation Computer Systems* (2016, 2018), the IEEE Cloud Computing (2015), the *IEEE Network* (2016), the IEEE Transactions on Cloud Computing (2017), the IEEE Transactions on Dependable and Secure Computing (2017), the *Journal of Computer and System Sciences* (2017), *Multimedia Tools and Applications* (2017), *Personal and Ubiquitous Computing* (2017), *Pervasive and Mobile Computing* (2016), and *Wireless Personal Communications* (2017).

WILLIAM BRADLEY GLISSON received the B.Sc. degree in management and the B.Sc. degree in information systems and operations management from The University of North Carolina at Greensboro, in 1993 and 1999, respectively, the M.Sc. degree in information management from the University of Strathclyde, Scotland, in 2001, and the Ph.D. degree in computing science from the University of Glasgow, Scotland, in 2008. He has ten years of industrial experience, which includes working for U.S. and U.K. Global Fortune 500 financial institutions. He has been the primary investigator on residual data research projects funded predominately by industry. He is currently an Associate Professor with the University of South Alabama. Previous to this appointment, he was the Director of the Computer Forensics M.Sc. Program with the University of Glasgow for five years. He builds on previous administrative and teaching experiences to teach and improve digital forensic courses while researching relevant real-world digital forensic issues. His area of research focuses on digital forensics, information assurance, software engineering, and applied computing science with specific interest in the security, business and health care implications associated with residual data.

ABDULALEM ALI received the B.S. degree in computer science from Thamar University, Yemen, in 2004, and the M.S. degree in information security from Universiti Teknologi Malaysia, Malaysia, in 2013, where he is currently pursuing the Ph.D. degree in computer science. His current research interest includes propose high level model (metamodel) for mobile forensics.

MOHAMMAD ABRAR received the M.S. degree from The University of Agriculture, Peshawar, Pakistan, in 2008, and the Ph.D. degree from Universiti Teknologi Malaysia in 2017. He is currently serving as an Assistant Professor with Preston University Kohat, Pakistan. His research interests include data mining, classification, algorithm analysis, and database system.