Reconstruction Investigation Model for Database Management Systems

Abdulaziz Saleh Alraddadi

College of Computer Science and Engineering, Taibah University Yanbu, Saudi Arabia

Article history
Received: 31-07-2023
Revised: 24-10-2023
Accepted: 25-10-2023

Email: alraddadi1@yahoo.com

Abstract: There have been increased levels of cybercrime in the database industry, which has hurt the confidentiality, integrity, and availability of these systems. Most organizations apply several security layers to detect and prevent database crimes. For this reason, Database Forensics (DBF) plays a very important role in capturing and discovering, who the criminal is, when the crime was committed, and which part of the database the crime occurred. Several forensic models have been proposed for the DBF field, which can be used to identify, collect, preserve, examine, analyze, and document database crimes. However, most of these models focused on specific database systems due to the variety of the database infrastructure and the multidimensional nature of the database systems. The most important part of the DBF field is the analysis process used to investigate the captured data and discover the attack. Thus, this study proposes an Integrated Reconstruction Investigation Model (IRIM) for database forensics using a metamodeling method. It consists of two main processes: The examining process and the discovering and reporting process. A real scenario has been used to validate the effectiveness of the proposed model. According to the results, the proposed model could detect database cybercrimes and allow domain forensic practitioners to capture and analyze database crimes efficiently.

Keywords: Design Science Research Database Forensics, Database Systems, Reconstruction Process, Investigation Model

Introduction

The field of digital forensics has evolved over the past few years to address the challenges posed by cybercrimes, cybersecurity threats, and the protection of digital data in our increasingly digital age. As a result of using advanced techniques, tools, and methods, digital forensics experts can solve crimes, strengthen cybersecurity, and provide valuable evidence during legal proceedings. There are several challenges that the field of digital forensics faces, including data overload, encryption, the rapid development of technologies, and legal considerations, which necessitate continued innovation and expertise in this field.

On the other hand, the field of cyber security is one of the most crucial areas for protecting sensitive information (Al-Dhaqm et al., 2023a; Alotaibi et al., 2023a). Security techniques, models, frameworks, procedures, policies, and processes have been proposed in several fields. These include database systems (Al-Dhaqm et al., 2023b), networks, robotics (Mohammed et al., 2021), prevent cyberbullying (Yafooz et al., 2023), wireless networks (Onwuegbuzie et al., 2022; Al-dhaqm et al., 2013) cloud security (Al-Mugern et al., 2023; Zubair et al., 2022), the internet of vehicles, IoT (Saleh et al., 2023), UAV field and mobile field (Yahya et al., 2023), medical field (Qureshi et al., 2022; Ngadi et al., 2012; Onwuegbuzie et al., 2020; Abd Razak et al., 2020; Altowayti et al., 2022; Rasool et al., 2022; Mohammed et al., 2022; Onwuegbuzie et al., 2021; Bakhtiar and Al-dhaqm, 2012). Database crimes increase daily, threatening privacy, integrity, and accessibility (Al-Dhaqm et al., 2023b). It is therefore essential to have a database forensics field that investigates who is the criminal when the crimes occur, and how they occurred (Salem et al., 2023). The purpose of database forensic investigations is to obtain digital evidence primarily from the target database for analysis, preservation, reconstruction, and presentation as evidence during subsequent litigation proceedings (Alfadli et al., 2021; Zawali et al., 2021; Al-Dhaqm et al., 2015; Alhussan et al., 2022a). In most cases, the evidence can be used to develop a hypothesis which can then be presented as evidence that can be used as a basis for litigation (Alotaibi et al., 2022b). As a specialty field of...
digital forensics, database forensics offers the opportunity to employ scientifically prescribed methods for the identification, collection, preservation, reconstruction, analysis, and documentation of database incidents (Al-Dhaqm et al., 2020a; Alotaibi et al., 2022a). In spite of this, database forensics is still seen as a complex, ambiguous, and heterogeneous area due to the fact that database systems are considered to be multifaceted (Al-Dhaqm et al., 2020b).

Thus, database forensic investigation is the process of investigating and analyzing databases as part of forensic science principles and techniques to uncover evidence pertaining to cybercrime or other illegal activities by utilizing forensic science principles and techniques (Al-Dhaqm et al., 2020a). An analysis of a database involves the examination and retrieval of data from a database, as well as the identification and analysis of any potential security breaches and unauthorized access to that database (Al-Dhaqm et al., 2017a). Due to the increasing reliance on databases to store and manage sensitive information in our society, the field of database forensics has gained significant importance in recent years. Many industries, across a variety of sectors, use databases to store data, including customer records, financial transactions, and personal information, and thus database services are very common across them (Frühwirt et al., 2013; Alotaibi et al., 2023b). As a result of database forensics, the objective is to identify and collect evidence that will be useful to the legal process or internal investigations that may require evidence. An example of this type of evidence could be deleted or altered records, log files, access logs, and other digital artifacts that may provide insight into the actions of an individual or group involved in illegal activities that can provide evidence of their involvement.

Due to the complexity and heterogeneity of database systems, specific and redundant collection and analysis models in the database forensic field are considered redundant. This study explores the need for identifying an integrated model that focuses on examining and analyzing. This is a step toward reducing redundancy. The authors have proposed an Integrated Reconstruction Investigation Model for the database forensic field, herein called IRIM. This model is evaluated based on a real scenario. The metamodeling method was adapted for development and validation (Al-Dhaqm et al., 2020c). Using metamodels or response surface models, one can create surrogate models in modeling and simulation. Models such as these approximate the behavior of complex systems or processes. As part of this study, the FTK Imager and Hash my files forensic tools have been applied to solve the problem related to a database attack scenario. FTK Imager is open-source software developed to access data that can be used to perform many criminal analysis functions (Utomo et al., 2023). Hash my files is a small utility that allows you to calculate the MD5 and SHA1 hashes of a file or files with the click of a button (Aderibigbe and Chi, 2017).

The proposed IRIM is intended to be used to reconstruct database events or incidents based on the available evidence in a database. It is a process where different techniques, methodologies, and expertise are combined, to construct a comprehensive picture of what occurred.

The purpose of this section is to discuss existing digital forensic works and to focus on the models that have been proposed in the literature for the examination and reconstruction of database process models. Several forensic examinations and reconstruction models have been proposed for the database systems. For example, the authors Wong and Edwards (2005) proposed a reconstruction model to rebuild intruder activities by showing malicious actions to reconstruct a database. In this way, the database was able to be restored. The research conducted by Fowler et al. (2007) has also suggested that part of the suggested model includes media analysis, data recovery, timeline creation, and string search processes as part of its implementation. Additionally, Fowler (2008) stated that this is part of the artifact analysis process that is utilized when analyzing malicious activity and retracing events that occurred in the past. However, (Choi et al., 2009) referred to that analysis process as a business and financial data analysis, using it to uncover fraudulent transactions that had taken place. There are other models that describe the analysis process as one of restoration and searchability (Onwuegbuzie et al., 2022), where research by Al-Dhaqm et al. (2013) essentially, this is the process of investigating how the data collection process was conducted. Furthermore, research by Khanuja and Adane (2012); implicitly describes this as a part of the reconstruction process and the physical and digital examination process. Also, Adedayo and Olivier (2015) proposed the forensic analysis process, which uses log management or log analysis tools to enhance the information volume analysis retrieved from log files in database forensics. In other models, the reconstruction and analysis process was described as an analysis of database attacks and analysis of anti-forensic attacks (Khanuja and Suratkar, 2014), reconstructing evidence (Frühwirt et al., 2014), forensic analysis (Khanuja and Adane, 2013) and rebuilding volatile artifacts (Wagner et al., 2015).

Additionally, Chopade and Pachghare (2019) proposed a survey study, which concentrated on the latest research on forensic examination of RDBMS and NoSQL databases and the survey of artifacts to be studied for database forensics. A prototype developed by Orosco et al. (2020) focuses on analyzing the possibility of rebuilding database contents from the Redo logs of a MySQL DBMS and gathering related data from the redo log files. Another research by Adamu et al. (2020) provided a study on the database anti-forensics agents and the bad effects at numerous phases of the database forensics processes. In
addition, Marsh et al. (2019) proposed a model to perform a deep forensic examination of HarperDB using a grouping of two methods: Database Forensics (DBF) which presents related stages to perform database forensic investigation, and common database forensic investigation process specifying appropriate stages to examine IoT environments.

Also, research by Al-Dhaqm et al. (2021a) also offers a face validation approach for the Database Forensics Metamodel (Bakhtiar and Al-Dhaqm, 2012), used to evaluate the completeness, logicalness, and usefulness of the database forensics domain. Other research includes (Choi et al., 2021) which offered a recovery model to retrieve deleted information from MSSQL.

The authors of the article (Al-Dhaqm et al., 2017a) have proposed a common process for database forensic investigation based on design science research. The proposed process is divided into four phases: Identification; collecting the artefacts; analysing the artefacts; and documenting and presenting the results (Abd Razak et al., 2016). Using a metamodeling approach, the researchers in Al-Dhaqm et al. (2017b; 2014) attempted to integrate existing attempts to represent database forensic knowledge in a reusable form while providing a single viewpoint from which to look at database access. According to the authors in Al-Dhaqm et al. (2020d), a harmonized mobile forensic investigation process model for the mobile forensic field goes a long way towards incorporating and streamlining the whole redundant investigation process within the mobile forensic sector. A metamodel for the mobile forensic domain has been proposed by Ali et al. (2015; 2017; 2018). According to the authors in Ali et al. (2017); Kebande et al. (2020), a metamodeling approach has been developed for mobile forensics, wherein common concepts that pertain to mobile forensics were identified.

A generic process model for database forensic investigation was presented by Al-Dhaqm et al. (2016). The process involves identifying, collecting, preserving, analyzing, and presenting the evidence collected during a forensic investigation. Using the developed model, the concepts and terminologies of all common database forensic investigations can be reconciled (Al-Dhaqm et al., 2021b). In order to facilitate the management of domain knowledge among practitioners, (Al-Dhaqm et al., 2018; Alhussan et al., 2022b) proposed a model-driven database forensic investigation system called model-driven database forensic investigation system. Table 1 displays the advantages and disadvantages of the existing database forensics models.

<table>
<thead>
<tr>
<th>No.</th>
<th>Ref.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Wong and Edwards (2005)</td>
<td>As presented in this study, the author proposed a reconstruction mechanism to reconstruct an intruder's activities by demonstrating malicious actions that can be used to reconstruct a database. As a result, it would be possible to restore the database in this manner.</td>
</tr>
<tr>
<td>2.</td>
<td>Fowler et al. (2007)</td>
<td>In accordance with the researchers' findings, the implemented model would be accompanied by processes such as media analysis, data recovery, timeline creation, and string search process to accomplish its goals.</td>
</tr>
<tr>
<td>3.</td>
<td>Fowler (2008)</td>
<td>As it appears, this procedure is a part of the artifact analysis process that is used to analyze malicious activity and retrace previously occurred events in order to identify perpetrators.</td>
</tr>
<tr>
<td>4.</td>
<td>Choi et al. (2009)</td>
<td>To uncover fraudulent transactions that had taken place, the company engaged in what it called a business and financial data analysis in order to uncover potential frauds.</td>
</tr>
<tr>
<td>5.</td>
<td>Khanuja and Adane (2012)</td>
<td>There is a framework proposed in this research paper that enables the analysis and reconstruction of the activity of any unsuspected behavior within a database to be performed. There is a purpose for identifying, collecting, analyzing, validating, interpreting, generating forensic reports, and preserving the evidence for digital investigations in order to conduct a forensic investigation.</td>
</tr>
<tr>
<td>6.</td>
<td>Adedayo and Olivier (2015)</td>
<td>The proposed forensic analysis process, which uses log management or log analysis tools, enhances the information volume analysis retrieved from log files in database forensics.</td>
</tr>
<tr>
<td>7.</td>
<td>Adedayo and Olivier (2015)</td>
<td>In this study, the authors present the notion of an ideal log setting to facilitate the successful reconstruction of databases for the purpose of forensic analysis. Several of the most popular database management systems have been compared with their default preferences for database management systems in this study. Furthermore, for a database log to be useful during the reconstruction process, it is necessary to identify what information must be logged in that log.</td>
</tr>
<tr>
<td>8.</td>
<td>Khanuja and Suratkar (2014)</td>
<td>This study focuses on the importance of metadata for database forensics and its significance in this study. In accordance with their proposal, they developed a system that enables forensic analysis of databases to be performed independent of the database management system used in order to generate metadata files.</td>
</tr>
<tr>
<td>9.</td>
<td>Frühwirt et al. (2014)</td>
<td>In this study, a novel approach was presented that utilizes transaction and replication sources to build a forensic-aware database management system. In order to reconstruct evidence during a forensic investigation, the team relied on internal data structures as a baseline for the evidence.</td>
</tr>
<tr>
<td>10.</td>
<td>Khanuja and Adane (2013)</td>
<td>An extensive study of database forensics has been carried out by the authors of this study and they have proposed a methodology for collecting and analyzing evidence and artifacts (volatile and non-volatile) such as data caches, log files, and so on.</td>
</tr>
<tr>
<td>11.</td>
<td>Wagner et al. (2015)</td>
<td>As part of their research, the authors have developed an academic tool that supports a wide range of databases seamlessly, rebuilding data content such as tables from any leftover fragment of storage on disk or in memory that remains. This study proposes an automatic method of reverse engineering storage in a new database (with a minimum amount of user intervention) in order to detect volatile data changes and identify user action artifacts that might exist.</td>
</tr>
</tbody>
</table>
As a result of this comprehensive review of all database forensic models, it becomes evident that the database forensic domain lacks an integrated reconstruction model that can be used to reconstruct and investigate database crimes using a database forensic investigation.

Materials and Methods

The purpose of this study is to develop and validate the IRIM by using a metamodeling approach (Al-Dhaqm et al., 2017a). By using metamodeling, you have the ability to integrate and define models from a variety of domains (Geisler et al., 1998). As a result, it is possible to identify and share common processes between these different viewpoints. Therefore, metamodeling can be applied successfully in a very wide range of different application domains, especially for standardization purposes. Basically, metamodeling describes the process of identifying the general processes that exist within a given problem domain and the relationships between them. In the current domain, it is used for the solving of complex, interoperable, and heterogeneous issues (Whittle, 2002) as a result, metamodels should be formally defined thoroughly in addition to being well-structured. Therefore, the methodology consists of five stages as shown in Fig. 1. (1) Collecting DBF examination and reconstruction models, (2) Extracting common processes (3) Combining common processes, (4) Proposing an IRIM model and (5) Evaluating the proposed IRIM model.

DBF examination and analysis models are collected in the first stage and common examination and analysis processes are extracted in the second stage. In stage three, processes that have similar names and meanings or different names with similar meanings are combined into one process. As a result of stage 3, we are able to propose the IRIM model based on the output of the third stage. The effectiveness and capabilities of the proposed IRIM will be evaluated in the final stage. Table 2 displays extracted processes from the collected models. Figure 2 displays the proposed IFIM. The proposed IRIM consists of two main processes: The examination process and, the discovering and reporting process:

1. Examination process: The purpose of this process is to check the authenticity of the data captured during the capture process. This process is used after the acquisition and preservation processes have been completed. By using the proper forensic tools, the investigation team acquires and preserves the volatile and non-volatile data from the relevant resources to solve the case. As a result, the investigation team will rehash the hashed values of the captured data and verify the authenticity of the captured data by rehashing the hashed values. If the data is authentic, the investigation team will move on to the next step in the investigation process, otherwise, it will make another copy of the original data.

2. Discovering and reporting process: The purpose of this process is to find and report any database crimes that may have occurred. It is important for the investigation team to assign suitable keywords that can be used to help find matching patterns of database crimes that may assist with the investigation. An investigation team should attempt this process several times until they find the evidence of the database crime, which is an iterative process that involves trying a variety of keywords until they find the evidence of the database crime. If the investigation team finds evidence, they will compile and organize the evidence into a report that will be documented as effectively as possible.

<table>
<thead>
<tr>
<th>No.</th>
<th>Similar processes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Restoring database integrity</td>
<td>Wong and Edwards (2005)</td>
</tr>
<tr>
<td>14</td>
<td>Media analysis</td>
<td>Fowler et al. (2007)</td>
</tr>
<tr>
<td>15</td>
<td>Timeline creation</td>
<td>Fowler et al. (2007)</td>
</tr>
<tr>
<td>16</td>
<td>Data recovery</td>
<td>Fowler et al. (2007)</td>
</tr>
<tr>
<td>17</td>
<td>Search string</td>
<td>Fowler et al. (2007)</td>
</tr>
<tr>
<td>18</td>
<td>Artifact analysis</td>
<td>Fowler (2008)</td>
</tr>
<tr>
<td>19</td>
<td>Economic and commercial data analysis</td>
<td>Choi et al. (2009)</td>
</tr>
<tr>
<td>20</td>
<td>Rebuilding and search ability</td>
<td>Olivier (2009)</td>
</tr>
<tr>
<td>21</td>
<td>Examination of data composed</td>
<td>Son et al. (2011)</td>
</tr>
<tr>
<td>22</td>
<td>Artifact investigation</td>
<td>Khanuja and Adane (2012)</td>
</tr>
<tr>
<td>23</td>
<td>Rebuilding</td>
<td>Susainamickam (2012)</td>
</tr>
<tr>
<td>24</td>
<td>Reconstruction of the database</td>
<td>Agededavy and Olivier (2015)</td>
</tr>
<tr>
<td>25</td>
<td>Forensic analysis</td>
<td>Khanuja and Adane (2013)</td>
</tr>
<tr>
<td>26</td>
<td>Study anti-forensic crimes, examination database crimes</td>
<td>Khanuja and Suratkar (2014)</td>
</tr>
<tr>
<td>27</td>
<td>Rebuilding evidence</td>
<td>Frühwirt et al. (2014)</td>
</tr>
<tr>
<td>28</td>
<td>Rebuilding process</td>
<td>Khanuja and Adane (2013)</td>
</tr>
<tr>
<td>29</td>
<td>Recovering volatile artifacts</td>
<td>Wagner et al. (2015)</td>
</tr>
<tr>
<td>30</td>
<td>Recovering database schema</td>
<td>Orosco et al. (2022)</td>
</tr>
<tr>
<td>31</td>
<td>Examination stage</td>
<td>Marsh et al. (2019)</td>
</tr>
</tbody>
</table>
The purpose of this section is to demonstrate the suitability of the proposed IRIM to verify and analyze crimes committed against the database. For this purpose, the authors prepared a scenario as shown in Fig. 3. The authors assumed that the attacker compromised the account of a customer, inserted SQL injection into the customer table, updated the secret key column, and prevented the customer from making an online purchase through the website. In the next few paragraphs, we shall describe in detail the capabilities of the proposed IRIM as it relates to examining and verifying database incidents:

a) Examination process: A preliminary investigation of this case showed that, in accordance with the above scenario, there had been a compromise in the customer table. It appears that some customers' records were duplicated by an anonymous attacker, as can be seen in Fig. 4. There has been a hacking attempt on the account of the customer named "Fahad", who has the customer_ID 121356 and a duplicate record has been created for his account by the attacker.

According to the initial analysis in Fig. 5, a comprehensive investigation was conducted by the investigation team using FTK images to capture the entire volatile data from the RAM. Figure 6 displays the captured data gathered from the RAM. Sequentially, the gathered data has been preserved using the hash my file tool as shown in Fig. 7. The hash my file tool produced a unique hash value for the captured data. Meanwhile, to verify the authenticity of the captured data for analysis, the FTK Imager was used to verify the hashed image that had previously been hashed. According to Fig. 8, the results of the verification showed that the captured data had the same value as the hash value of Fig. 7, which indicates that the data is original and has not been altered. As a result, the main objective of the examination process of the IRIM has been achieved successfully:

![Fig. 1: Metamodeling approach (Al-Dhaqm et al., 2017a)](image1)

![Fig. 2: Integrated reconstruction investigation model for database systems](image2)

![Fig. 3: Scenario of database attack](image3)
Results and Discussion

In this section, we discuss the results of the study. The literature has proposed several models and frameworks for the study of database forensics, each of which discusses DBFs from a different perspective. This study discusses the DBFs from the reconstruction perspective and proposes an integrated reconstruction and investigation model based on a metamodeling approach. Basically, it consists of two processes: Examination and discovery, as well as reporting processes. We have validated the proposed model using a real scenario based on the data that we have collected. Based on the results of the study, the proposed model could be used to detect database cybercrimes and permit domain forensic practitioners to capture and analyze database crimes more efficiently.

In this study, FTK Imager was used as a forensic tool in order to investigate and analyse database crimes as part of an investigation and investigation process.
In order to unravel the complexity of crimes that were committed during this time period, the use of this tool, which can perform forensic image analysis, extract data, and perform in-depth analyses, has proven to be invaluable in unravelling the complexity of the crimes that were committed during that time period. A high level of effectiveness and reliability have been shown to be associated with the FTK Imager tool and proposed model in terms of detecting database-related criminal activities based on the findings of this research.

Comparing the proposed IRIM with the existing DBF models shown in Table 3, the proposed IRIM is compatible with the existing models and covers the majority of them as well. For example, Wong and Edwards (2005) proposed two processes that can be used to restore the integrity of the database, namely reconstructing the database, and restoring its integrity. It was proposed by Fowler et al. (2007) that four processes could be employed: Media analysis, timeline creation, data recovery, and search string extraction. A summary of the proposed processes by different authors can be found in Table 3. There is no doubt that all these models are covered by the proposed IRIM.

**Conclusion**

It is widely believed that database forensics is one of the most important fields in the use of capturing and analyzing data that comes from databases and this is the purpose of database forensics. As a result, this field is vital when it comes to investigating, discovering, and reconstructing events related to databases. There are several forensic techniques and methodologies that are applied to the investigation, detection, and reconstruction of events related to databases, through the application of forensic techniques and methodologies. Several reconstruction models and frameworks have been proposed in the literature for database systems. Nevertheless, most of these models focused on a specific database system because of the variety of database infrastructures available and the multidimensional nature of the database systems. In this study, an integrated reconstruction investigation model for the database system called the Integrated Reconstruction Investigation Model, is proposed. This model consists of two main processes: The examination process and the recovery and reporting process. The proposed model has been evaluated using real scenarios and the results showed that the IRIM is a powerful tool that can be used to verify and analyze database crimes effectively in an accurate and efficient manner. The future work of this study will be to validate the completeness of the proposed IRIM by comparing it with other models.

**Acknowledgment**

I would like to express my deepest gratitude to the researchers who supported this study with all the necessary references.

**Funding Information**

For this research, there have been no funds available.

**Ethics**

By signing this document, I certify that this article has not been published anywhere else in the world.


