A Systematic Literature Review on Necessity, Challenges, Applications and Attacks of Watermarking Relational Database

Abd. S. Alfagi[1], A. Abd. Manaf[1], B. A. Hamida[2], Mohd. Ghazli. Hamza[3]
[1]*Advanced Informatics School,*
Universiti Teknologi Malaysia Kuala Lumpur, Malaysia.
[2]*Department of Electrical and Computer Engineering,*
International Islamic University Malaysia, Kuala Lumpur, Malaysia.
[3]*Razak School of Engineering and Advanced Technology,*
Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.
alfagi2008@googlemail.com

***Abstract*—Systematic literature review (SLR) is a significant research methodology in software engineering and computer science. One precarious step in applying this methodology is to enterprise and perform appropriate and effective search approach. This is a time-consuming and error-prone step, which needs to be carefully planned and implemented. There is an apparent need for a systematic approach to designing and performing a suitable search strategy for optimally retrieving the target literature from digital libraries. To the best of our knowledge, five intensive review papers [1-5] have been found which are not SLR, instead they are a normal survey or literature review. In contrast to [1-9] this paper followed general guidelines for undertaking SLR in order to illustrate necessity, challenges, applications, and attacks of watermarking relational database. An advanced search has been performed in most relevant digital libraries to obtain potentially relevant articles published until the end of 2014. Forty-six primary studies (PSs) have been identified based on inclusion and exclusion criteria. The analytical study is mainly based on the PSs to achieve the objectives. The results illustrate the importance of digital watermark in protecting the relational database, the differences between watermarking relational database and multimedia objects and the demand to increase the level of attack resilience. In addition, the results indicate that watermarking relational database is an interested area for researchers.**

***Index Terms*—Systematic Literature Review; Database Watermarking; Watermarking Multimedia Objects; Database.**

I. INTRODUCTION

Digital watermarking is important in many real life applications. It has been considered as an advantageously complete useful solution for many problems encountered in the distribution of different multimedia objects such as image, text, and audio. Similarly, a digital watermark is effective in protecting relational databases since nowadays; sharing information online is an important activity for business and research, which also involves buying or selling of databases. Applying the digital watermark in protecting relational databases is a relatively new research area that deals with the legal issues such as copyright protection, ownership proof, and data integrity check [6][7]. Although Watermarking database relations is a highly an active area of research there is a lack of systematic literature reviews in this area to keep researchers up to date with the state of research in the area [7]. Systematic literature reviews aim to identify, assess and combine the evidence from primary research studies using an explicit and rigorous method. This method has been widely implemented in software engineering and computer science. Systematic literature reviews are aware of the importance of literature search, as well as the challenges involved in searching relevant studies when applying SLR methodology in different disciplines.

Therefore, this systematic literature review (SLR) aims at systematically illustrating the need of digital watermark in relational database protection. In addition, it aims at highlighting the challenges and issues that arise in the study of applying digital watermarking techniques on relational databases to help researchers in the field to have a wide prospector on most common issues and challenges. Further, this systematic literature review aims at surveying the most important applications of the digital watermark in securing relational databases to provide a statistical analysis of digital watermark applications as help protecting database relations. Besides that, it aims to illustrate the fundamental philosophical differences between information hiding, steganography, and watermarking as to reduce the overlapping of these concepts and hence, to depict the basic framework of watermarking relational database. Finally, this systematic review aims at addressing the differences between watermarking databases and multimedia objects as well as surveying the most possible attacks on watermarked databases in order to provide a brief description of most common attacks as well as illustrating their intend. In conducting this SLR, seven scientific online digital libraries were included in the search strategy, and potentially 608 relevant articles as a result of that search. After applying inclusion and exclusion criteria at the study selection stage as well as screening the selected articles at the assessing quality stage, identify 46 relevant primary studies (PSs).

II. RELATED WORK

There is a rich body of literature on watermarking multimedia objects [8] [9], starting from watermarking still images [10] and later extended to digital uncompressed along with compressed video [11], audio signal [12] and text

[13]. Besides that, digital watermarking has also been exploited in other digital media for tamper-proofing, and as obfuscation-tools for software protection [14]. Using digital watermarking in database protection started to have researchers' attention just recently, as such the first effort for protecting database relations using a watermark was in 2002 by [8]. Compared with watermarking multimedia objects, watermarking database relations is considered a relatively new field, and thus there is a lack of a rich body of literature survey in this field [7] [15]. Although there are some literature review papers (to the best of our knowledge only [1-5]) that provide a literature review on watermarking relational database, none is a systematic literature reviews. Unlike to [1-5], this systematic literature review followed general guidelines for undertaking systematic literature reviews as specified by [16] which detail a range of related work, provide a systematic and rigorous approach to illustrating necessity, challenges, applications, and attacks for watermarking relational database.

III. RESEARCH METHOD

This systematic literature review performs three main phases; planning, conducting, and reporting the result. In planning phase, initially we identify the need for a systematic literature review, then framing focused research questions using recent criteria called Population, Intervention, Comparison, Outcome, and Context (PICOC) used by [17]. While in the second phase, we initially searched the databases for primary studies, after that evaluated the articles for relevance and quality, then extracted data from the primary studies. In the reporting phase, the results are synthesized, analyzed and reported. This section explores the protocol that has been followed to review and reduce the chances of researcher bias. The protocol includes identifying the research questions, defining the search strategy, determining the study selection, and the study quality assessments.

A. Research Questions

Based on the primary studies that have been determined by the study selection process, this SLR attempts to answer the following research questions:

RQ1: What is watermarking, steganography and data hiding?
RQ2: Is there a necessity to watermark relational database?
RQ3: What are the applications of the digital watermark in protecting database relations?
RQ4: What are the differences between watermarking multimedia objects and relational database?
RQ5: What are the challenges in watermarking relational database?
RQ6: What are the attacks on watermarked database?

The first research question (RQ1) is motivated by the desire to define and differentiate between similar terms thus reduce the overlapping between these concepts and hence to depict the basic framework of watermarking. The second research question (RQ2) is motivated by the desire to highlight the necessity of digital watermark to protect relational databases. The third research question (RQ3) is motivated by the desire to know the most important applications of the digital watermark in protecting the relational database. The fourth and fifth research question are motivated by the desire to illustrate the main differences between watermarking multimedia objects and relational database (RQ4) as well as the challenges in watermarking relational databases. The six research question (RQ6) is motivated by the desire to describe the common attacks on watermarked databases.

B. Research Strategy

To obtain a comprehensive list of articles in the area we conducted an advanced search in most popular and relevant digital libraries that contain peer-reviewed journal articles, conference proceedings, and book chapters. The selected databases include PubMed, ACM, Springer, Scopus, IEEE, ISI, Google scholar, and Science Direct. The selection of these libraries increases confidence in the completeness of the review. In addition, to increase the comprehensiveness of this systematic literature review the searching years were specified from the year 2000 up to the end of 2014. Besides that, the search string w constructed based on the following factors:(1) The major terms extracted from the research questions are: (2) Alternative spellings and synonyms of the major terms; (3) Research keywords appeared in other relevant papers (e.g. [1-5]); (4) Boolean (AND) was used to connect the major research terms and Boolean (OR) used to connect alternative spellings and synonyms of the keywords.

The advanced search in each library was used in order to obtain the targeted search criteria. A general search string that have been used in all selected digital libraries are as follows: (database watermarking, watermarking-database, Relational Databases watermarking, Relational Databases copyright and ownership, database watermarking for integrity, Numerical database watermarking, Watermarking Relational Databases, Digital watermarking and Relational database).

C. Study Selection

This section explains the method of selecting PSs from relevant articles. An inclusion and exclusion criteria were applied in order to narrowly focus on reviewing digital watermark for relational database only. In inclusion criteria, the PSs have to be published in peer-reviewed journals or conference proceeding. In addition, the articles should related to watermarking relational database. Meanwhile, all articles that did not fulfill the inclusion criteria were excluded. Inclusion and exclusion criteria are important requirements and must be satisfied in all selected articles in order to ensure that the selected PSs were within the related and targeted area of research.

IV. RESULT

In this section, we provide an overview on the PSs then answering the research questions based on analyzing the 46 primary studies that we have identified in accordance with our review protocol.

A. Overview of the PSs

A total of 608 articles were identified at the end of searching strategy. After careful monitoring and applying study selection on the 608 articles, in 523 articles the inclusion criteria were not satisfied therefore they excluded and in 34 articles the full text was not available

consequently they also excluded. The remaining was 51 articles, 46 of them considered as relevant primary studies PSs and 5 articles are normal literature reviews or surveys, which are not systematically and do not follow [17]. The 46 articles were used in order to achieve the objectives of this systematic review. The distribution of the journal, conference and the total of all PSs over the years are shown in Figure 1(a), while (b) shows the percentage of PSs published in conferences and journals respectively.

The first PS that was considered in this SLR appeared in 2002, which is a conference paper and explains the first effort of watermarking relational databases, whereas the first journal was published in 2003 provides a watermarking relational database a framework, algorithms, and analysis. From Fig. 1 (a) can be clearly to noticed that, watermarking relational databases began to attract more researcher's attention after 2002. In addition, it indicates that researchers in watermarking relational database have shifted their publishing interests in recent years from conferences to journals. This indication is a positive, as journal articles are more prestigious and typically more complete as well report studies that are more extensive. Another noticeable point that, watermarking relational databases has seen a significant increase in the number of published articles (journals and conferences) during years especially in recent years. Figure 1 (b) shows that over the half (54%) of PSs articles were published in journals whereas the rest (46%) were published in conferences. The slight difference between percentages of journals and conferences indicates the importance of considering both journals and conferences in this SLR. The extracted observations from the Fig. 1. indicates that watermarking relational databases is an interesting area for research due to the noticeable increase in the number of publications.

B. *RQ1: What is Watermarking, Steganography and Information Hiding?*

Information hiding, steganography, and watermarking are three considerably similar terms and closely related to each other. They have a lot of intersections, many common characteristics and technical approaches. Therefore, it is important to illustrate the fundamental philosophical differences. This SLR answer this question to provide brief definitions for the terms to reduce the overlap and to depict the basic framework of a digital watermark. All PSs have defined neither information hiding nor steganography (0 %). To define those terms we have extracted the definitions from three different articles [18][19] and [20]. According to [18] "information hiding is a general term encompassing a wide range of problems beyond that of embedding messages in content. The term hiding can refer to either making the information imperceptible (as in watermarking) or keeping the existence of the information secret". Ali and Mahdi [19] defines steganography as a method that establishes a covered information channel in point-to-point connections only. [20] defined digital watermarking as a class of information hiding technique developed to ensure that the carrier signal quality is preserved, provides measures for copyright protection, broadcast monitoring, covert communication, copy control, tamper, and integrity proof of digital assets. In general, the digital watermark has two phases; embedding and extraction phases. These phases are almost similar in both watermarking multimedia objects and watermarking relational database as confirmed by 13 % of PSs. Figure 2. depicts the basic database watermark embedding and extraction phases.

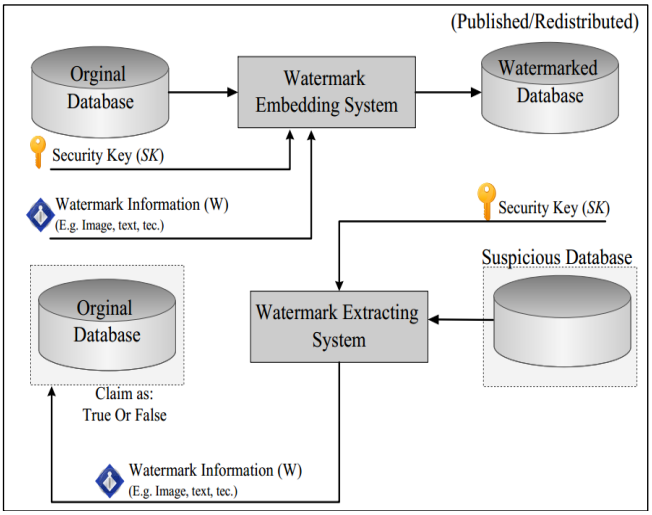

Figure 2: General Framework for watermarking relational database

A. *RQ2: Is There a Necessity to Watermark Relational Database?*

Databases are the repositories of the most important and expensive information in the companies. Most organizations share data electronically for different goals. Although, several security mechanisms have been deployed for databases protection such as access control and encryption. However, the number of reported data leaks and frauds each year is not negligible [21]. Access control and encryption are able to protect the exposure of sensitive information before granting the access once data is accessed or ancillary

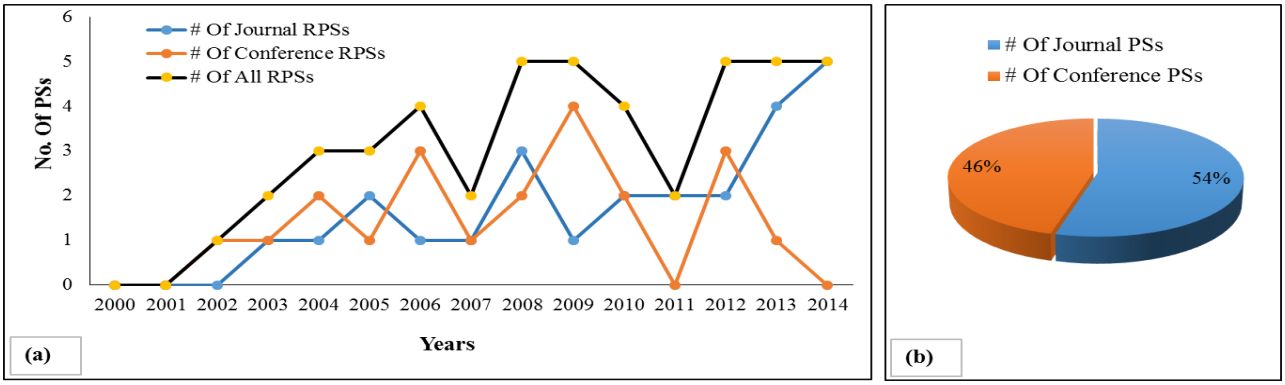

Figure 1: (a) Distribution of PSs over the year (b) number of PSs published in conference and journal

security attributes are removed; data are no longer protected [22]. Thus, it is important to highlight the need for the digital watermark in protecting relational databases. In addition, answering this question will reveal the importance of digital watermark applications as well as briefly explain how the watermark is used for ownership proof, copyright protection, fingerprinting and verify data integrityFor ownership proof, data owners can securely insert a watermark into a relational database using secret key before publishing or distributing their data. At the extraction phase, the data owners can demonstrate the presence of their watermark in order to verify and defeat third party ownership claiming. While for copyright protection, data owners securely embed owner's specific information (e.g. image, text, speech, …etc.) into the relational database to prevent others from claiming copyright [23]. Whereas in the fingerprinting data, the main goal of watermark is to identify the sources of data. In case, the owner of a digital content can embed distinct watermarks within the content supplied to different customers. Here, the watermark is helpful in identifying those customers who break the license agreements by supplying the content illegally to unauthorized parties[2]. Finally, for content authentication and verify data integrity data owners securely insert a fragile watermark into the relational database content using some secret parameters then for verify and check data integrity, the watermark has to be extracted watermark using same secret parameters. After that, the extracted watermark is used to check the integrity of data. Fragility is a desirable property when using digital watermark for authenticating and verifying data integrity [22] [24].

Digital watermark has been used as an advantageous and complete useful solution for many problems encountered in distributing different multimedia objects such as, image, text, and audio [2, 3, 24] [25, 26] Similarly, digital watermark is effective in protecting relational databases. In this SLR, all PSs (100%) considered digital watermark as a solution for copyright protection (50%), ownership proof (22%) and database integrity (28%). These percentages are strong evidence on the necessity of watermarking relational database for different purposes that other security mechanism such as access control or encryption could not provide. Besides that, there is a significant increase in the number of published articles during last decades that explained in section 4.1, another indication of the importance of watermarking relational databases. Thus, watermarking database relations is an important need.

B. RQ3: What are the Applications of Digital Watermarking in Protecting Relational Database?

Different digital watermarking techniques are designed to serve different purposes such as database integrity, copyright protection, and ownership proof. There have been many considerable applications of the digital watermark in protecting relational database's copyright, ownership and database integrity. Therefore, it is important to the readers to know numerically the application of digital watermarking in protecting the relational database. For this aim, a statistical analysis has been conducted based on the purpose of each PSs and a column chart has been drawn and presented in Figure 3 to illustrate in percentage, the applications of digital watermark in protecting relational database. From the chart can be clearly seen that digital watermark plays an important role in protecting database copyright, ownership, and integrity. Of all PSs, 50 % were intended to protect relational databases copyrights, which is the highest percent in the chart. Whereas, protecting database's ownership or integrity using digital watermark have almost similar percentage that is 21% and 28% respectively. Even though, applying a digital watermark for the copyright of relational database is important and noticeably recoded the highest percentage, the other applications as seen in Figure 3 are not ignored. Thus, the digital watermark has important applications in protecting copyrights, ownership, and integrity of the relational database.

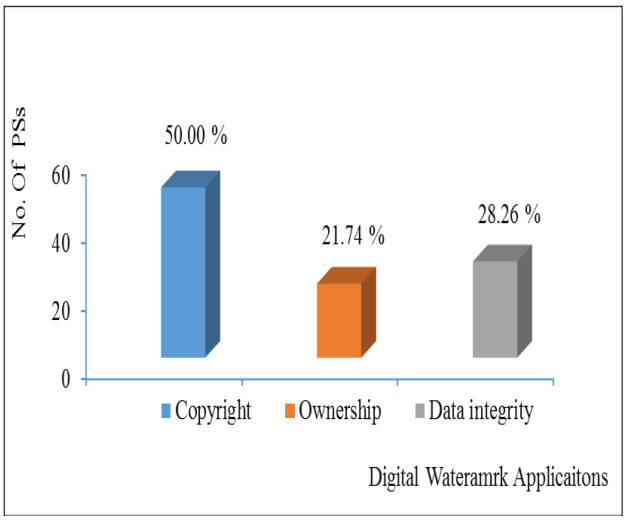

Figure 3: Applications of digital watermark in database relations

C. RQ4: What are the Differences between Watermarking Multimedia Objects and Relational Database?

The database is a collection of relations that integrate various kind of data presented in tabular format. In addition, each database has its structure, size and contents. While multimedia data usually have similar data contents that usually represented by numerical values (pixels). Since the relational data defers from multimedia data in many respects defiantly the embedding channels differ. Therefore, it is a reasonable to illustrate the basic differences between the embedding channels in multimedia objects and a relational database. About 80% of PSs have mentioned the differences between relational database and multimedia data objects such as data type (embedding channels), structure, size, changeability, and content sensitivity. Although (in section 4.2 13%) confirms that the basic concepts in watermarking multimedia is quite similar to watermarking database 70 % of PSs confirmed that existing digital watermarking techniques used for multimedia are not designed to handle relational databases due to the differences in embedding channels, requirements, available bandwidth, and content sensitivity.

Most PSs confirmed the embedding channels is a common problem in watermarking relational database, which is a very limited compared to multimedia objects. In addition, the embedding positions are not changeable in multimedia data whereas it is changeable in relational database. In the image, the embedding positions (pixels) are not changeable while in database embedding positions (attribute or tuples) may reorder or repositioned as explained by PS17. Furthermore, database normally conducts frequent update by adding, deleting or modifying its content whereas

any portion of multimedia objects is not dropped or replaced normally. Besides that, there are many psycho-physical phenomena based on human visual system and human auditory system which can be demoralized for watermark embedding. However, such phenomena are not exploited in the case of relational databases. These differences in characteristics of database and multimedia data lead to increasing the researchers' effort as well as give a rise to many technical challenges in developing digital watermarking techniques for watermarking relational database. In addition, these differences make watermark relational database much complex than watermarking multimedia data and made it an active area of researches.

D. RQ5: What are the Challenges in Watermarking Relational Database?

Digital watermarking technique ensured that a digital watermark is embedded into digital data and can only be detected or extracted by an authorized entity for many different purposes. However, watermarking database still faces many challenges [2]. Therefore, it is important to highlight the issues that arise in the study of applying digital watermarking techniques on relational databases in order to make easier for researchers in the field to have a wide overview of most common issues and challenges. For simplicity, we listed them as below with a brief description and statistically analyzed most common issues at the end of this section.

i. Iss1: Capacity: It is defined as the maximum amount of information or the number of bits that can be embedded into the original data without damage the usability of the data (carrier).
ii. Iss2: Usability: The desirable embedded amount of watermark information into the database should not degrade the usability of data.
iii. Iss3: Robustness: The embedded watermark should be robust to resist certain attacks whether they are malicious or accidental.
iv. Iss4: Reversibility: it is defined as the ability to obtain the original data from watermarked database after extracting the embedded watermark.
v. Iss5: Blindness: This point considers the knowledge needed to extract the watermark information from the watermarked database. The extraction should not require neither the original un-watermarked database nor the watermark itself.
vi. Iss6: Structure: Most databases are made of many relations and these relations between inter-related to each other. As a result, once should consider that the attributes that have inter-relation or joined before watermarking process should not be altered during watermarking.
vii. Iss7: Security: The security of watermarking system should rely on the private keys (e.g. security key or choice of attributes and tuples) not on the algorithm. Their keys should be kept secret and known to the database owner only. This point must achieve the requirements of public system where the security defense must lie only in the choice of the private parameters
viii. Iss8: Incremental watermarking: When a database has been watermarked, the algorithm must compute the watermark values for only the added or modified tuples. The tuples that have not altered during the watermarking process should not be re-watermarked.
ix. Iss9: Non-inference: When there is a need to embed more than one watermark in a single database relation the embedded watermarks should not conflict with each other.
x. Iss10: False Positiveness and false Negativeness: false positive or false hit is the probability of a valid extraction of the watermark from the un-watermarked database. While, false negative or false miss is the probability of not detecting a valid watermark from watermarked database.

According to four review papers [1] [2] [3] [27] these are the most common open issues. Besides that, two recent journal articles PS38 and PS40 confirmed that Iss1, Iss2, Iss3, and Iss4 are still open issues in developing digital watermark for relational database. Therefore, in this section we are focusing more on those issues.

Capacity, reversibility, robustness and minimize the distortion among other requirements need to be kept reasonably balanced [24]. In all PSs, there is always a trade-off between capacity and other two important properties of watermarking system such as distortion and robustness. A higher capacity is always obtained by sacrificing either robustness or distortion (or both). If too much data are hidden in the carrier (much more than the payload capacity), it will harm the usability of the data. The allowable level of degradation in database is different from database to another, which mostly based on the purpose of embedding and the data sensitivity. For instance, bank, health, and military database considered as a sensitive data [28]. Therefore, it is important to preserve the usability of such a sensitive data and thereby necessary that a good trade-off is needed between capacity, distortion, robustness and reversibility.

In this SLR, the majority of PSs focusing on providing a digital watermark technique that achieve high embedding capacity, high level of robustness, reversible and low distortion or zero distortion. About 72 % of PSs work to provide a robust watermark for copyright protection and ownership proof. Meanwhile less 28 % work on fragile and semi fragile watermark, which aim to preserve database integrity. Besides the Iss1, Iss2, Iss3 and Iss4, issue (Iss5) which related to the blindness has been surveyed in all PSs and we have found that about (48%) of PSs provide a blind watermark technique whereas (24%) were not blind and only 6% were semi-blind.

Finally, the current research on database watermarking has been primarily focused on how to overcome those challenges or to optimal/near optimal tradeoff between the basic properties of a watermarking system. As more techniques became available for watermarking relational database limitation of embedding capacity, no usability constrain, high false positive rate, frequent updates, and robustness are still open issues as confirmed by three review papers and PS38 and PS40.

E. RQ6: What are the Attacks on a Watermarked Database?

A watermarked data may undergo certain types of attacks before reach to the detector or extractor side. Regardless the intent of the attacks (intentional or unintentional) the result may lead to destroying the watermarked data, removal of the watermark or adding noise or extra information on the

watermarked data [2]. However, the type of attacks can be recognized based on the type of the watermark (e.g. fragile or robust) as well as the application being used (e.g. copyrights, ownership, tamper detection or fingerprinting) or the type of the data being carried (e.g. image, text, video) [9]. In addition, attacks are used to evaluate the ability of the watermark in surviving the maximum level of attacks and the ability to defend against several kinds of database attacks. Therefore, it is important to provide the reader with the most common and the possible attacks on watermarked database. This SLR answers this question in order to sort the most common attacks and labeled them (A1 to A8) with a brief description and illustrates the intent of some attacks. The most common attacks are described below:

i. A1. Benign Updates: it is the case when database contents are modified regardless whether the changes are intentional or unintentional. These changes include adding new, deleting some tuples or attributes or modifying values of tuples. However, these changes may involve the watermarked tuples and attributes resulting a damaging or removing the watermark. For instance, updating database contents may erroneously flipped the marked bits.
ii. A2. Value Modification Attack or Malicious Attacks: this category of attack has the following attacks:
    - Bit Attacks: In this kind of attack, attackers attempt to destroy the embedded watermark partially or totally by altering one or more bits in the watermarked tuples. The effectiveness of such attack depends on the amount of the information that obtained about the watermarked tuples. For example, if the attackers got more knowledge about the positions of watermarked bits most likely the attack will be more successful. In some cases, conducting Bit Attacks may damage or make database contents useless. Bit attack may be performed by:
        - Randomly assigning values to certain bit positions which known as Randomization Attack.
        - Setting the bit positions to zero which known as zero attacks.
        - Inverting some values of bits which known as bit flipping attack. For instance, attackers may randomly select some least significant bit LSB and toggles their values in order to destroy the watermark [9].
    - Rounding Attack: This kind of attack conducted by rounding all or major values of the numerical attribute. The success of rounding attack depends on the guesstimate of how many bits have been involved from that attribute in the watermarking. Underestimation of bits number may cause the attack unsuccessful whereas overestimation may cause the data useless.
    - Transformation: An attack related to the rounding attack is one in which the numeric values are linearly transformed. For example, attackers may convert the data to a different unit of measurement (e.g., Centimeter to Meter, feet or inch, or Fahrenheit to Celsius).
iii. A3. Subset Attack: This kind of attack just causes modification, deletion or addition on a subset of tuples $S \subseteq Rw$ where S is a subset and (Rw) watermarked relation. As result of such attack the watermark may lost or effected.
iv. A4. Superset Attack: To perform this attack attacker attempt to add new tuples or attributes to the watermarked database. Attackers aim to achieve the goal of miss leading the correct detection of the embedded watermark.
v. A5. Collusion Attack: In this attack, multiple fingerprinted copies of identical relation are required to be accessed by the attackers. Collusion attack may be performed by:
    - Mix and Match Attack: In this case, attackers may create their relation by taking disjoint tuples from numerous relations having the same information.
    - Majority Attack: a new relation has to be created in this attack with the same schema as the watermarked copies. However, each bit value in the new relation computed as the majority function of the corresponding bit values in all copies. The main goal of such attack is to prevent the owner from detecting the watermark.
vi. A6. False Claim of Ownership: In this type of attack attackers aim to insert another watermark in order to conflict the merchant's claim. It includes two kinds. First, one is Additive Attack; where a second watermark is added to the watermarked relational database to claim ownership this known as secondary watermark attack. The second one is Invertibility Attack; the attackers lunch an invertibility attack to claim his ownership if attackers can fictitious watermark which in fact, a random occurrence from a watermarked database.
vii. A7. Subset Reverse Order Attack: this attack is simple conducted by just exchanging the order or positions of the tuples or attributes in the relation to erase or disturb the embedded mark specially that depend on the order of tuples like fragile watermark.
viii. A8. Brute Force Attack: In this case, the attackers try to guess about the private parameters (e.g. secret key) by traversing the possible search.

According to four review papers [2] [3] [4] [5], these are the most common attacks that may watermark database undergo. Researchers are struggling to provide a digital watermark that able to resist as much kind of attacks as possible. However, the PS28 stated that "Till date only a few types of attacks are overcome. Besides that, about 10 % of PS18, 21, 27, 33, 42 recommended the need to strengthen watermark and to increase the level of attack resilience. Thus, a further work in this area is required to increase the level of attack resilience.

V. DISCUSSION

This section discusses and interprets the results reported in Section 4.

A. Differentiate Between Similar Terms (Related to RQ1)

The terms information hiding, steganography, and watermarking are closely related to each author. They have a great deal of overlapping because they have many common characteristics and technical approaches. The result presented in section 4.2 shows that five previous review papers have not explanation or definitions on those terms. In

addition, information hiding, steganography, and watermarking have not defined by any PS. This systematic literature review answers RQ1 in order to help the readers' gain better understanding of fundamental differences between information hiding, steganography, and watermarking as well as showing the main phases involved in database watermarking techniques.

B. The Necessity of Digital Watermark and its Applications in Protecting Database Relations (Related to RQ2 and RQ3)

The result reported in section 4.3 and 4.4 show that the digital watermark has important applications in multimedia data and database relations as well. Among many applications of digital watermark, this SLR has shown that digital watermark plays important roles in protecting relational database copyrights (50%), ownership proof (22%), and integrity (28%). Since the copyright protection recorded the highest percent among the other applications, we advise the researchers to focus on the applications of digital watermark that aim to protect database integrity and ownership proof.

C. Differences Between Watermarking Relational Database and Multimedia Objects (Related to RQ4)

The result reported in Section 4.4 show that, watermarking relational database and multimedia objects differ in many aspects. Embedding channels considered the most common difference which addressed by the majority of PSs. Besides that, frequent update that regularly happen to database relations hardens the watermark algorithm in finding embedding positions compared to multimedia data where the embedding positions allows have fixed places. These differences make watermark relational database much complex than watermarking multimedia data and made it an active area of researches.

D. Challenges and Open Issues in Watermarking Relational Database (Related to RQ5)

The result reported in Section 4.4 show that there are many issues remained open. A high embedding capacity is a major challenge in watermarking database relations. Since it is hard to hide much information such as private data, authentication data, tamper detection and localization data in very few locations. Many relational database watermark techniques presented in PSs have limited capacity especially for embedding a meaningful watermark. Because the embedded watermark bits length totally depend on the maximum number of least significant bits that can be altered [29]. However, issues such as getting the original data back (reversibility) and attack resilience (robustness) are two key challenges that reported by many PS and still open up to date. Reversibility is highly desired in many sensitive databases [30]. However, some database watermarking scheme are not reversible and provide permanent distortion which are not applicable to sensitive database such as medical or, military database. Therefore, database watermarking is an active area of research and there is a need to a watermark scheme with optimal trade-off between the embedding capacity, reversibility, and robustness to protect copyright, ownership, and integrity of such a sensitive database.

E. Attacks on Watermarked Database (Related to RQ6)

The result reported in Section 4.4 provides a list of most common attacks that may watermarked database undergo. The list was totally based on 5 review papers and the 46 PSs. However, many PS18, 21, 27, 33, 42 recommended the need to increase the level of attack resilience in order to make watermark resist as many attacks as possible. Therefore, we advise the researchers to focus on this need.

VI. CONCLUSION

Systematic literature reviews aim to identify, assess and combine the evidence from primary research studies using an explicit and rigorous method. This method has been widely implemented in software engineering and computer science. In this paper, a systematic literature review conducted to investigate the current state of knowledge about watermarking database relations. 46 primary studies have been identified in accordance with our review protocol and published between 2000 to the end of 2014. We described the related work that have been used to review watermarking database relations. In addition, we provided an overview of primary studies and analyzed primary studies articles to answer six research questions systematically. Unlike others (e.g. [1-5]) this systematic literature review followed general guidelines for undertaking systematic literature reviews. The major contributions of this paper can be concluded as:

i. Detailing an obvious range of related work, search strategy and study selection for relevant studies in the field of database watermarking.
ii. A systematic, evidence-based, and rigorous approach to conducting and reporting the result of the research question.
iii. Providing a systematic literature review instead of a normal review.

A lack of systematic literature reviews in this field to keep researchers up to date with the state of research in the area encourage authors to continue the evaluation and improvement of this approach by conducting SLR on the techniques that used in watermarking relational database.

ACKNOWLEDGMENT

The authors would like to express greatest appreciation to Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM) for financial support. Furthermore, Full thanks to support by Ministry of Defense, Libya under the Ph.D. scholarship programs.

REFERENCES

[1] Li, Y. "Database Watermarking: A Systematic View. Handbook of Database Security", pp. 329-355, 2008. Springer .
[2] Halder, R., Pal, S., Cortesi, A. "Watermarking Techniques for Relational Databases: Survey, Classification and Comparison". *Journal of Universal Computer Science*, vol. 16, pp. 3164-3190, 2010.
[3] Arathi, C. "Literature Survey on Distortion based Watermarking Techniques for Databases". *International Journal of Computer Science & Communication Networks* 2, 2012.
[4] Bilapatte, S., Bhattacharya, S., Sawarkar, S. "A review on watermarking relational databases". *International journal of applied engineering research and development* , 2014.
[5] Kshatriya, M.S., ,M.S., Prof.Dr.S.S.Sane. "A Study of Watermarking Relational Databases". *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 3, 2014.

[6] Al-Haj, A., Odeh, A. "Robust and blind watermarking of relational database systems". *Journal of Computer Science 4*, (2008),1024.

[7] Mehta, B.B., Rao, U.P. "A Novel approach as Multi-place Watermarking for Security in Database". arXiv preprint arXiv:1402.7341, 2014

[8] Agrawal, R., Kiernan, J. "Watermarking Relational Databases*". In: VLDB Conference.*

[9] Farfoura, M.E., Horng, S.-J., Lai, J.-L., Run, R.-S., Chen, R.-J., Khan, M.K. "A blind reversible method for watermarking relational databases based on a time-stamping protocol". *Expert Systems with Applications* vol. 39, 2012, pp. 3185-3196, 2012.

[10] Craver, S., Memon, N., Yeo, B.-L., Yeung, M.M. "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications". *Selected Areas in Communications, IEEE Journal* on, vol. 16, pp. 573-586, 1998.

[11] Hartung, F., Girod, B.: Watermarking of uncompressed and compressed video. *Signal processing*, vol. 66, pp. 283-301, 1998.

[12] Boney, L., Tewfik, A.H., Hamdy, K.N. "Digital watermarks for audio signals. In: Multimedia Computing and Systems", *Proceedings of the Third IEEE International Conference on*, pp. 473-480, 1996.

[13] Brassil, J.T., Low, S., Maxemchuk, N.F. "Copyright protection for the electronic distribution of text documents". *Proceedings of the IEEE, vol.* 87, pp. 1181-1196, 1999.

[14] Collberg, C.S., Thomborson, C. "Watermarking, tamper-proofing, and obfuscation-tools for software protection". *Software Engineering, IEEE Transactions* on, vol. 28, pp. 735-746, 2002.

[15] Guo, H., Li, Y., Liu, A., Jajodia, S. "A fragile watermarking scheme for detecting malicious modifications of database relations". *Information Sciences, vol.* 176, pp. 1350-1378, 2006.

[16] Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., Linkman, S. "Systematic literature reviews in software engineering – A systematic literature review". *Information and Software Technology*, vol. 51, pp. 7-15, 2009.

[17] Kitchenham, B., Charters, S. "Guidelines for performing systematic literature reviews in software engineering". *EBSE Technical Report EBSE-2007-01*, 2007.

[18] Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography. "The Morgan Kaufmann Series in Multimedia Information and Systems". 2007.

[19] Ali, Y.H., Mahdi, B.S. "Watermarking for relational database by using threshold generator". *Computer Sciences Department, University of Technology Baghdad. Eng. & Tech. Journal* 29, 2011.

[20] Khan, A., Husain, S.A. "A fragile zero watermarking scheme to detect and characterize malicious modifications in database relations". *TheScientificWorldJournal 2013*, 796726, 2013.

[21] Franco-Contreras, J., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., Roux, C. "Authenticity Control of Relational Databases by Means of Lossless Watermarking Based on Circular Histogram Modulation". *Security and Trust Management*, pp. 207-222, 2013.

[22] Bhattacharya, S., Cortesi, A. "A Distortion Free Watermark Framework for Relational Databases". *In: ICSOFT*, vol. 2, pp. 229-234.

[23] Manjula, R., Settipalli, N. "A new relational watermarking scheme resilient to additive attacks". *International Journal of Computer Applications*, vol.10, pp. 1-7, 2010.

[24] Abdullah, S.M., Manaf, A.A., Zamani, M. "Capacity and quality improvement in reversible image watermarking approach. Networked Computing and Advanced Information Management (NCM)", *2010 Sixth International Conference on*, pp. 81 – 85, 2010.

[25] Tsai, M.-H., Tseng, H.-Y., Lai, C.-Y. "A Database Watermarking Technique for Temper Detection". *In: JCIS*.

[26] Dadkhah, S., Abd Manaf, A., Hori, Y., Ella Hassanien, A., Sadeghi, S. "An effective SVD-based image tampering detection and self-recovery using active watermarking". *Signal Processing: Image Communication* vol. 29, pp. 1197-1210, 2014.

[27] Abokhdair, N.O., Manaf, A.B.A.: A Review Of Reversible Watermarking Properties, Applications And Techniques For Medical Images. (2013).

[28] Al-Sayid, N.A., Aldlaeen, D. "Database Security: Threats A Survey Study". *International Conference on Computer Science and Information Technology (CSIT)*, 2013.

[29] Lafaye, J., Gross-Amblard, D., Constantin, C., Guerrouani, M. "Watermill: An optimized fingerprinting system for databases under constraints". *Knowledge and Data Engineering, IEEE Transactions* on vol. 20, pp. 532-546, 2008.

[30] Mehta, Brijesh B., and Hardika D. Aswar."Watermarking for security in database: A review". *IT in Business, Industry and Government (CSIBIG), 2014 Conference on. IEEE*, 2014.