PAPER • **OPEN ACCESS**

Relational Database Watermarking Techniques: A Survey

To cite this article: Asmaa Alqassab and Mafaz Alanezi 2021 *J. Phys.: Conf. Ser.* **1818** 012185

Relational Database Watermarking Techniques: A Survey

Asmaa Alqassab[1]*, Mafaz Alanezi[2]

[1, 2] Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Iraq 1.2.

Emails: asmaa_mow@uomosul.edu.iq , mafazmhalanezi@uomosul.edu.iq

Abstract. While data is used in cooperative milieus for information extraction; Thus, it is vulnerable to security threats concerning ownership rights and data abusing. Due to unauthorized access to the data that may alter the originality, it results in significant losses of the organization. The relational databases which are free on-hand are used by research society for mining new information regarding their research works. These databases are vulnerable to security issues. The reliability of the data source must be authenticated before using it for any application purpose. Thus, to check the ownership and reliability of data, watermarking is applied to the data. Watermarking is used for the protection of the possession rights of shared Relational Data and for providing the solution for manipulating and tampering of data.

1. Introduction

The information travels through the web without any monitoring and could be crushed or tampered. Usually, the user of the information has no idea about the propriety of the information it uses, so when someone puts forged data in the information wittingly, the one who is going to analyze the data would make a misleading education and it might have a vastly effects on related research. Implementing the public authentication helps to protect the digital information on the internet, evidencing the copyright and integrity of digital information is most significant and thus digital watermarking is evolved.

Originally, digital watermarking is based on Information Hiding which is considered a popular technique for multimedia data such as images, videos, and audios. In case of people argued about the copyright of protected information, we can just extract the embedded watermarks to prove the ownership rights. Primely digital watermarking technique is applied to copyright protection and integrity of information content authentication. For copyright protection, robustness for the watermarks is needed. Attaining robustness means if someone modifies the data on purpose then it is still possible to extract the embedded watermarks with some tolerable distortion. For the integrity of information content authentication, the watermark's role is to inform whether the data is attacked. Anciently, the digital watermarking technique was used in the image process. Nowadays, it is used on databases due to the fact that the markets of databases have been widely increased. Large numbers of researches on watermarking multimedia data are available [1][2][3][4][5]. Watermarking relational databases have many technical challenges due to the differences in the characteristics of relational database and multimedia data, these differences involve [6][7][8][9]:

- The Database comprised of the tuples, in which each tuple is considered as a separated object. therefore, the watermark is spreading over these separated objects. While the multimedia

object is made up of a large number of bits so, there is a wide amount of space for hiding the watermark.

- In the case of the database, the tuples may change during the updates in the database. While any update doesn't affect the relative spatial/temporal positioning of multimedia object.
- In contrast to multimedia objects, Tuples may easily be dropped by applying delete operation in the database. Drop or Replace operation in a multimedia object result in noticed changes in the object.

2. Watermarking and Relational Databases

In general, watermarking approaches implement two operations: (A) watermark embedding and (B) watermark extraction. Embedding operation comprises two sub operations: watermark generation and watermark bits embedding. The watermark generation source might be the original database or any multimedia file. On the other hand, the extraction operation ordinarily implemented once it is needed as a proof in case of litigations and hence, this forms a significant challenge to the relational database in relation to rest data kinds due to the fact that in the database the data may need daily modifications like: (update, insertion, and deletion of the data). The extraction operation sub-operations include watermark bits detection, then extraction, and the reconstruction of the watermark, in addition to the watermark enhancement which is an optional sub-operation. Lastly, plainest watermarking operations require one value at least predefined as Secret Key (SK), solely revealed to the data possessor. The general structure of the relational database (RDB) watermarking is illustrated in figure 1[10][11][12][13].

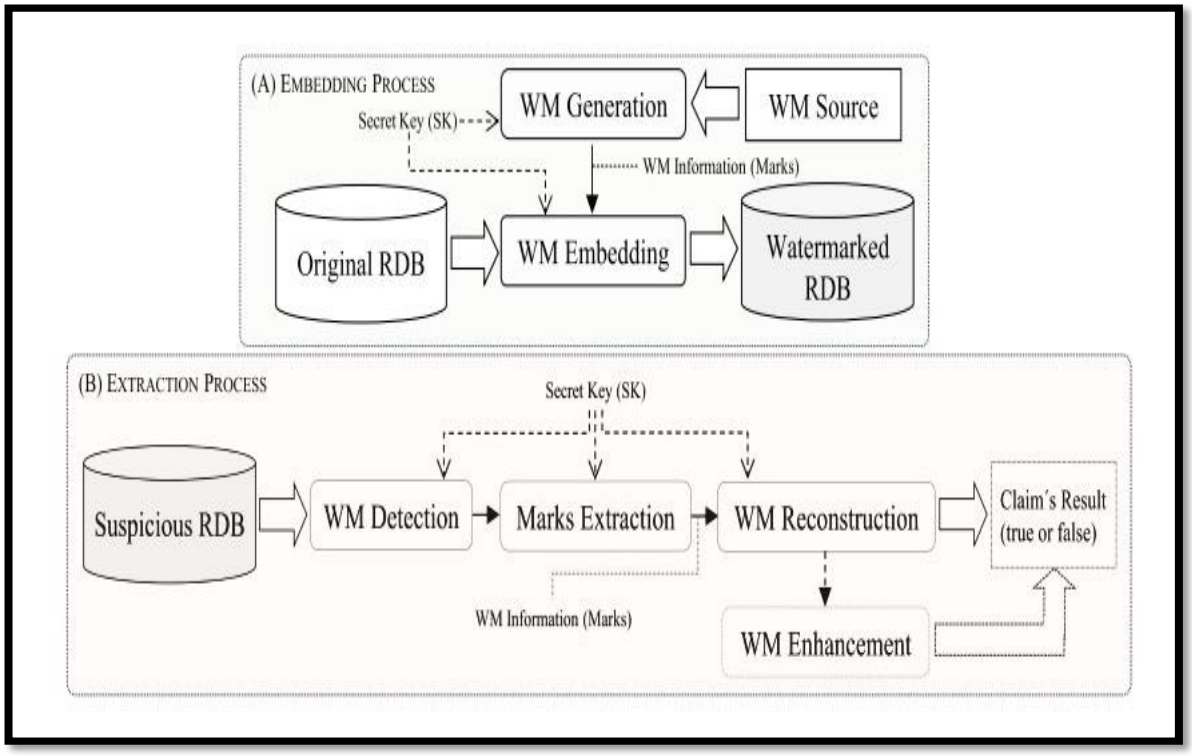

Figure 1. General Structure of RDB Watermarking

Relational databases watermarking can be classed into two types: Reversible Watermarking and Irreversible Watermarking Technique, based on their capability to regenerate original data (such that original data may or may not be retrieved after watermark decoding) [14][15]. The irreversible

Watermarking method probably causes modifying or altering the underlying original data to a certain extent. Reversible Watermarking is employed to get rid of such problem and that results in lossless and precise authentication of relational databases, in addition to that exact recovering of the original data attribute from the watermarked relational databases can be acquired from this watermarking method [16][17].

Any watermarking technique for databases has certain features that should be satisfied [18][19][20][21][22][23]:

- Robustness: This means the watermarking capability to resist manipulations caused by benign or malign attacks. Thus, the watermark has to firmly establish the possession of the database. Therefore, such watermarks should have the ability to be extracted even though its exposure to several indiscriminate modifications caused by different attacks.
- Imperceptibility: The viewer shouldn't be able to determine that there's a watermark inserted. The watermarked database has to look just like the same as the original cover database to the human eye.
- Security: A watermarking technique is considered secure when an unauthorized user is not able to erase, identify, modify, or recover the inserted watermark even if both of the embedding and extraction algorithms were known.
- Effectiveness: The watermarking technique must not be specified for certain databases, such that it may be swimmingly applied to embed a watermark in any selected database.
- False positive rate: Represents the number of the databases identified as it possesses an embedded watermark whilst it hasn't such. The false-positive rate must be low for any watermarking approach.
- Payload size: Or data payload represents the amount of the watermark bits that are embedded in the original database.
- Capacity: Watermark capacity indicates the maximal redundancy of payload size within the database.
- Complexity: For any watermark approach, the computational cost (embedding detection and extraction) correlates with the watermarking complexity. As the watermarking complexity increase, the cost will automatically increase too. Consequently, the algorithms employed must be efficient with low time complexity.
- Verifiability: This means that the watermark should have the capacity to present absolute and reliable proof to the liability for the secured data. It may be employed to manage illegal duplicating, observe the deployment of the data being secured, validness recognition, and decide whether the data need to be ensured.
- Fidelity: Represents the degree of resemblance of the image prior watermarking and that image posterior. This watermarking feature considered the most significant among other features.

3. Reversible Watermarking

The Reversible Watermarking may be also referred to as Lossless Watermarking, that makes it possible to retrieve the embedded data completely along with the entire recovering of the original (cover) data. Nowadays Reversible Watermarking has gained large attention due to its increased applications in law-enforcement, martial communication, and health-care.

Reversible Watermarking comprises the same watermark embedding and extraction stages as in irreversible watermarking in addition to a third data recovery stage [24][25][26].

The techniques of reversible watermarking for relational databases maybe apportioned into four stages: preprocessing, encoding, decoding, and data recovery.

Through the data preprocessing stage, two sub-stages implement various functions: selecting an appropriate feature for the embedding of the watermark bits, and watermark formulating.

Through the watermark encoding stage, watermark data bits are embedded in the chosen feature(s). Besides, some parameters are calculated during this stage to be used in both watermark encoding and decoding stages.

Through the watermark decoding stage, the embedded watermark bits are decoded in the suspicious positions of the data. As the watermark data bits are verifying, the originally inserted watermark and detected decoded watermark will be compared, and should be the same for the ownership evidencing.
In the data recovery stage, original data is recovered by executing post processing activities for recovery and error correction. then a comparison between the original and recovered data might also be done for the guarantee of the uncompromised data quality [27][28][29][30].

4. Digital Watermarking Applications for Relational Database

To conserve relational databases RDB, various watermarking technics were implemented for achieving various objectives like a database: (ownership probative, integrity of data, and copyright protection). So, to get a simple idea about these technics, statistical analyses have been computed depending on the aim of number of primary studies (PSs) that had been made as described in figure 2 [31]:

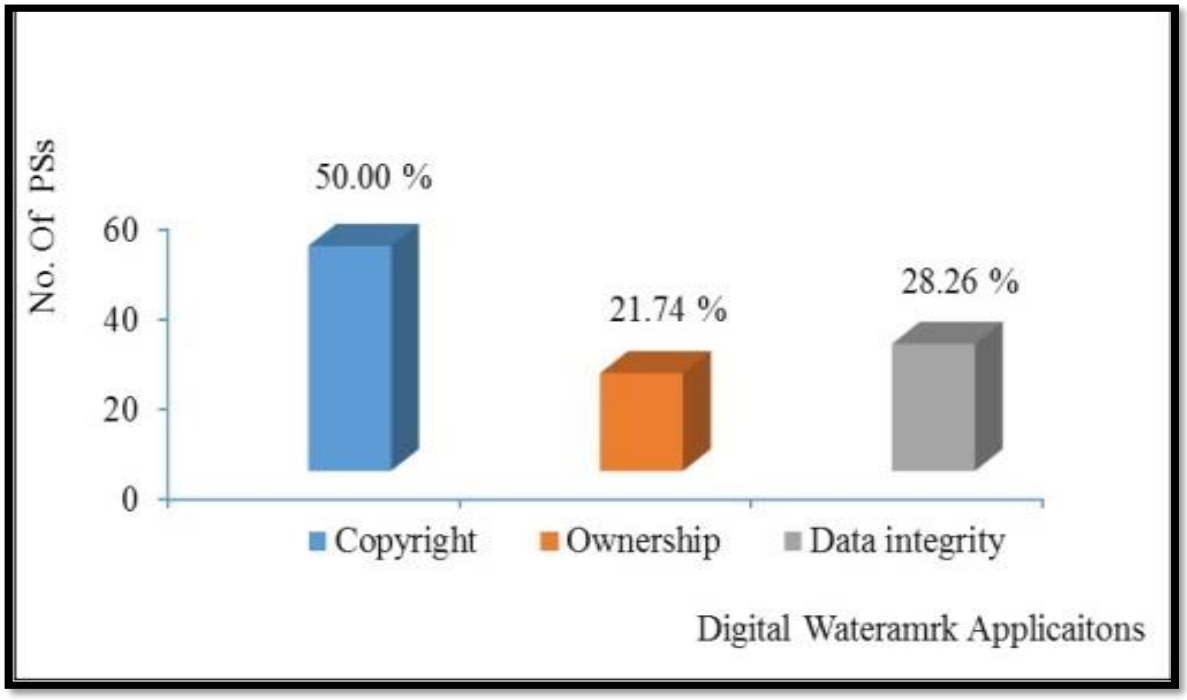

Figure 2. Applications of Digital Watermarking for RDB

Additionally, it can be still possible seeing a few other databases watermarking applications such as [32]:

- Impostor tracing and database trackability which aims to avert illegitimate database redeployment.
- Abuse disclosure or Integrity manage which handles data (uniformity and preciseness) stored in database.

A private-key based watermarking approach was suggested by Ali Y. H. and Mahdi B. S. for copyright protection of database numerical attributes. The approach employed the hash-based message authentication code in conjunction with a threshold generator which's used to specify whether the values of marking bit locations, will be "0" or "1" [33].
A zero-watermarking technique which employs an image as a private key for watermark generating process, was suggested by Murugan R. et al., for information copyright protection of relational

database. Here there isn't any inserted information into the cover data. Alternatively, watermarking key which has been constructed, will be registered with the Certification Authority. For this reason, both of data advantageousness and quality are maintained as nothing are added into the original cover data. Moreover, this technique can be applied for watermarking all kinds of database attributes [34].

A watermarking method for identifying database legitimate owner and showing elasticity to miscellaneous forms of attacks, suggested by Tzouramanis T., The presented method works on numeric tuples bits of relational database via rearranging every single tuple bits in an undercover manner then picking number of its data bits to direct the tuple to a particular watermark bit plus a single data bit set out by the value of the allotted watermark bit. [35].

To cope with ownership strifes of watermarked datasets in the case of additive attacks, Shyamala G. et al., suggested a technique for providing lesser distortion and utmost accuracy in decoding. Through analyzing the related decoding accuracy of the watermarking technique under different sorts of attacks, thus it demonstrates its robustness [36].

To accomplish database tamper detection and contents verification, Chang J. and Wu H., suggested a reversible fragile watermark approach on the basis of SVR (support vector regression) prediction, which embeds important properties of the originality of database. Frequent Pattern tree (FP-tree) data mining procedure is applied besides associated rules with the aim of identifying major characteristics for each selected data to be used for SVR prediction. This approach utilizes overhead data for data quality authentication but it is not robust enough against serious attacks [37].

An elaborated survey in database watermarking methods, submitted by Dwivedi A. K. et al., by which watermarking methods, power and weak points, along with their limitations, are analyzed. Furthermore, diverse sorts of databases watermarking issues and attacks investigated and discussed at length [38].

With the intention of right protection and authentication for relational databases, Murugan R. et al., suggested an invisible watermarking approach that can suit any of data attributes sort. The approach initializes by logical embedding of the database rights owner as a mark, and a watermarking key is created which is retained together with the watermark within the Certification Authority (CA), in which the true owner is registered. Afterwards the watermark extraction process can be carried out by applying watermark extraction procedure aided by the watermarking key. Ultimately, and as a means to evidence database integrity and true owner authenticity, both of the extracted and original watermarks are matched for the watermark accuracy judgement [39].

A lossless watermarking method for integrity manage of categorical data type in medical database, suggested by Coatrieux G. et al., here, tuples are arranged as secret sets by applying a one-way hash function, before mark embedding process. To confirm the associated set integrity, a digital signature is needed to be formed through encoding the mark in sets on the basis of histogram shifting modulation [40].

A reversible watermarking approach for reclaiming original data and detecting tamper with relational database, suggested by Unnikrishnan K. and Pramod K. V., In order to meet the database ideal areas for watermark embedding, both of optimization methods: firefly algorithm (FA) and orthogonal learning particle swarm optimization (OLPSO), are employed. This approach has the power to bring back the database to the initial status if any tampering occurs [41].

With the aim of database ownership verification, Ramani S. V., suggested such method through an unperceived watermark inserting so as to ensure security and robustness versus endeavors to eliminate the watermark. Image is used to show database watermarking ownership by converting the image into row bits which is then encrypted with MD5 security algorithm. Later these row bits ought to be embedded into the attribute of the database as a watermark [42].

A blind reversible databases watermarking technique that shows the rightful owners of the databases, suggested by Tale P. G. et al., within which original database relation is entirely retrieved whenever the watermark data is recovered and validated. Watermarks embedding operation is built upon the utilization of a secret embedding key for the non-numeric attributes through multi-word shifting of sub-sets of the tuples. The essential benefit of such technique is its capability to conceal watermarks of big sizes due to its wide bit capacity [43].

5. Watermarking Challenges and Attacks

The watermarked database may experience multifarious sorts of purposed or unpurposed attacks even before arriving its extractor or detector side and that often conduce to watermarked database ravaging, watermark elimination or noise insertion to the watermarked database. Essentially, databases attacks can be split into two main sorts: active attacks & passive attacks, for active attacks: the values of the original database are altered, this sort is more troubled due to the fact that it may cause user misleading, while for passive attacks: here attackers merely follow the user's communications beyond the network, without any database alterations [44][45][46].

Furthermore, attacks sorts, can be disaggregated according to the application or data kind besides the type of the watermark as well, as listed hereinafter [44][47][48][49][50]:

1. Benign updates: Watermarked database contents (attributes or tuples) are normally processed with an update, add, or delete operations, that can lead to erase the embedded watermark bits or can trigger the embedded watermark unrecoverable.
2. Malicious attacks: Attack sort that consists of the following varied attacks:
   - bit attacks: Try to devastate the embedded watermark through bits manipulations. The more bits are manipulated the more effectuality of this attack. This sort of attack can be executed in different forms: setting bits locations to zero, referred to as Zero Attack, inverting number of bits values, referred to as Bit Flipping Attack, and assigning certain bits locations with indiscriminate values, referred to as Randomization Attack.
   - rounding attack: Attempt to get rid of marks, through rounding the whole values existing in the numerical attributes. The activeness of this attack relies on precise estimation of the amount of bits locations shared in the watermarking process, which may be either above, lead to data distortion, or down, lead to failed attack.
   - transformation: This time the values existing in the numerical attributes are transformed linearly, means, data is converted to another measurement unit.
3. Subset attack: Applied to a subset of the attributes of a watermarked database, caused the watermark to be damaged or lost.
4. Superset attack: Try to insert additional attributes to the watermarked database to delude the embedded watermark detection.
5. Collusion attack: Attackers need to reach several fingerprinted replicates of analogous relation so that this attack can be accomplished by:
   - mix and match attack*:* Selecting separated attributes from several relations with identical data set to produce a new relation.
   - majority attack: Creating a new relation that has a similar schema just like the watermarked replicates. In this case, the owner won't be able to reveal the watermark.
6. False claim of ownership: Comprises inevitability and Additive attack. It aims to add additional watermark to cause a conflict and prevent the proprietor from right claiming.
7. Subset reverse order to attack: Merely includes interchanges of attributes locations of the relation leading to obliterate or affect the embedded watermark.
8. Brute force attack: Here, attackers have to figure out about the secret parameters (secret key for example).

6. Conclusion

Ensuring database security occupies the first place of institutions and organizations considerations as their data distributed over public and private networks, furthermore, is largely utilized for applications and research intent, thereby conserving the quality of data of the relational database is an essential move. Digital watermarking offers a new fashion for database security. The motivation behind this paper is to inspect existing studies thus far respecting the database watermarking approach and locate the points at which such approaches are potential to enhance, resulting in raising the security level that watermarking approaches offered. Numerous database security procedures were adapted for data protection. Database watermarking technology largely applied for rights protection applications, yet, various other applications of database watermarking were there, for example, integrity control of

databases and traitor tracing. Reversible watermarking approaches were adopted owing to their power in reprocessing watermarked data to recover original data with assured data quality. Ongoing efforts in this field of researches endeavor to meet robust along with a reversible outcome to ascertain that data subjects to watermarking shan't face integrity or quality degradation, and it should be attacked durable.

7. References

[1] Kumar M and Verma O 2018 Multiple watermarking of relational databases and ownership claim *JETIR* **5** 1006

[2] Tufail H, Zafar K and Baig A 2019 Relational database security using digital watermarking and evolutionary techniques *Computational Intelligence* (Wiley Periodicals Inc.) p 1

[3] Zhang Z, Zhang M and Wang L 2020 Reversible image watermarking algorithm based on quadratic difference expansion *Mathematical Problems in Engineering* (Hindawi) p 1

[4] Chang C, Nguyen T and Lin C 2013 A blind reversible robust watermarking scheme for relational databases *The Scientific World Journal* (Hindawi Publishing Corporation) p 1

[5] Hou R and Xian H 2019 A graded reversible watermarking scheme for relational data *Mobile Networks and Applications* (Springer Science+Business Media, LLC) p 1

[6] Mehta B and Rao U 2011 A novel approach as multi-place watermarking for security in database *Int'l Conf. Security and Management | SAM'11 |* p 703

[7] Panimalar S and Srinath D 2015 Reversible watermarking technique based on timestamping in relational data *IJAICT* **2** 961–962

[8] Shaikh U and Shedge K 2016 A secure watermarking technique for numeric and non-numeric relational data *IJIRCCE* **4** 18013

[9] Vashistha R and Vashistha D 2017 Watermarking with usability constraint of datasets *IJARSE* **6** 701–702

[10] Gort M, Uribe C, Cortesi A and Peña F 2019 HQR-Scheme: a high quality and resilient virtual primary key generation approach for watermarking relational data *Expert Systems with Applications* vol 138 (Elsevier Ltd.) p 2–3

[11] Gort M, Uribe C and Nummenmaa J 2016 A high capacity and robust image-based watermarking technique for relational databases *INAOE (Technical Report* No. CCC-16-010) pp 5–6

[12] Zhong J and Liao H 2015 The watermark technology application research guided by the idea of computer safe operation *2nd Int. Conf. on Electrical, Computer Engineering and Electronics* p 1708–1709

[13] Cao Z, Shi G and Wu Q 2019 Research on database watermarking based on independent component analysis and multiple rolling *IJDSN* **15** 2–3

[14] Kulkarni R and Patil D 2015 Watermarking of relational databases: survey *IRJET* **2** 786

[15] Kamran M and Farooq M 2018 A comprehensive survey of watermarking relational databases research (arXiv:1801.08271 [cs.CR]) p 4

[16] Gadiya P and Kale P 2016 Reversible watermarking for relational data: a brief review *IJCTA* **7** 25

[17] Aparna P and Kishore P 2020 A blind medical image watermarking for secure e-healthcare application using crypto-watermarking system *J. Intell. Syst*. **29** 1559

[18] Kumar S, Singh B and Yadav M 2020 A recent survey on multimedia and database watermarking *Multimedia Tools and Applications* (Springer Science+Business Media, LLC) p 6–7
[19] Khanduja V 2017 Database watermarking, a technological protective measure: perspective, security analysis and future directions *Journal of Information Security and Applications* vol 37 (Elsevier Ltd.) p 39
[20] Kiran and Garg K 2015 Digital watermarking: potential challenges and issues *IJCSET* **5** 48
[21] Qasim A, Meziane F and Aspin R 2018 Digital watermarking: applicability for developing trust in medical imaging workflows state of the art review *Computer Science Review* vol 27 (Elsevier Inc.) p 47
[22] Camara L, Li J, Li R and Xie W 2014 Distortion-free watermarking approach for relational database integrity checking *Mathematical Problems in Engineering* (Hindawi Publishing Corporation) p 1–2
[23] Allaf A and Kbir M 2019 A review of digital watermarking applications for medical image exchange security (Springer Nature Switzerland AG) p 473–474
[24] Thilagavathi N, Saravanan D, Kumarakrishnan S, Punniakodi S, Amudhavel J and Prabu U 2015 A survey of reversible watermarking techniques, application and attacks *Proc. Int. Conf. on Advanced Research in Computer Science Engineering & Technology (Unnao, India)* p 3
[25] Khan A, Siddiqa A, Munib S and Malik S 2014 A recent survey of reversible watermarking techniques *Information Sciences* vol 279 (Elsevier Inc.) p 252
[26] Iftikhar S, Kamran M and Anwar Z 2015 A survey on reversible watermarking techniques for relational databases *Security and Communication Networks* vol 8 (John Wiley & Sons Ltd.) p 2581
[27] Gishma.K and Vareed J 2016 A study of watermarking relational databases *IJARTET* **3** 59–60
[28] Xie M, Wu C, Shen J and Hwang M 2016 A survey of data distortion watermarking relational databases *IJNS* **18** 1023
[29] Shukla A, Bhambar S, Patil H, Kumavat K, Ghyar H and Desale S 2017 RRW: a novel watermarking technique for relational data *IJCA* **165** 16
[30] Iftikhar S, Kamran M and Anwar Z 2015 RRW - a robust and reversible watermarking technique for relational data vol 27 (*IEEE Transactions on Knowledge and Data Engineering)* p 2–3
[31] Alfagi A, Manaf A, Hamida B and Hamza M 2014 A systematic literature review on necessity, challenges, applications and attacks of watermarking relational database *JTECE* **9** 104
[32] Chathuranga K 2019 Watermarking technology for copyright protection of relational databases (*DOI:10.13140/RG.2.2.21623.06565)* p 6–7
[33] Ali Y and Mahdi B 2011 Watermarking for relational database by using threshold generator *Eng. & Tech. Journal* **29** 33–43
[34] Murugan R, Abraham J and Salim I 2019 A robust watermarking technique for copyright protection for relational databases *IJRTE* **8** 4040–4046
[35] Tzouramanis T 2011 A robust watermarking scheme for relational databases *6th Int. Conf. on Internet Technology and Secured Transactions (Abu Dhabi, United Arab Emirates)* p 783–790

[36] Shyamala G, Kanimozhi C and Kavya S 2015 An efficient distortion minimizing technique for watermarking relational databases *IJSETR* **4** 2050–2054

[37] Chang J and Wu H 2012 Reversible fragile database watermarking technology using difference expansion based on SVR prediction *Int. Symp. on Computer, Consumer and Control* (IEEE) p 690–693

[38] Dwivedi A, Sharma B and Vyas A 2014 Watermarking techniques for ownership protection of relational databases *IJETAE* **4** 368–375

[39] Murugan R, Jaseena K and Abraham J 2017 An invisible watermarking technique for integrity and right protection of relational databases *IJAER* **12** 15754–15758

[40] Coatrieux G, Chazard E, Beuscart R and Roux C 2011 Lossless watermarking of categorical attributes for verifying medical data base integrity *33rd Annual Int. Conf. of the IEEE EMBS (Boston, Massachusetts USA)* p 8195–8198

[41] Unnikrishnan K and Pramod K 2017 Robust optimal position detection scheme for relational database watermarking through HOLPSOFA algorithm *Journal of Information Security and Applications* vol 35 (Elsevier Ltd.) p 1–12

[42] Ramani S 2013 Watermark based copyright protection for relational database *IJCA* **78** 22–28

[43] Tale P, Dharaskar R and Thakare V 2017 Reversible relational database watermarking using levenshtein distance and prediction-error expansion *IJCMS* **6** 27–34

[44] Alfagi A, Manaf A, Hamida B, Khan S and Elrowayati A 2016 Survey on relational database watermarking techniques *ARPN-JEAS* **11** 422–423

[45] Sharma P 2016 Database security: attacks and techniques *IJSER* **7** 314

[46] Gahlot S, Verma B, Khandelwal A and Dayanand 2017 Database security: attacks, threats and control methods *IJERT* **5** 1

[47] Pande D, Upadhyay M, Pal S and Wankhade S 2014 Watermarking of relational databases using optimization technique *IJARCCE* **3** 8261

[48] Rathva M and Sahani G 2013 Watermarking relational databases *IJCSEA* **3** 72

[49] Zhu Q 2018 Digital watermarking technology based on relational database *JIM* **21** 1212

[50] Prajapati S and Tiwari N 2015 Image based relational database watermarking: a survey *IOSR-JCE* **17** 55–56

Acknowledgments
The researchers thank the Department of Computer Science, College of Computer Science and Mathematics, University of Mosul.