International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

An Efficient and Robust Approach for Relational Databases Watermarking

Ms. Snehal Kshatriya[1], Ms. Pushpalata Aher[2], Ms. Reena sahane[3], Ms. Neha kale[4]

[1,2,3]Assistant professor, Dept. of Computer Engineering, Sandip University, Nashik
[4]Assistant professor, Dept. of Computer Engineering, SIEM, Nashik

---

Abstract - *The sharing of relational database over internet create a need for security of these databases. Unauthorized access to this data may results in significant losses for the organization. Watermarking relational database is a solution to protect information from unauthorized duplication and to provide ownership protection. The aim of proposed work is to implement an efficient watermarking technique for relational databases. Proposed work deals with embedding the same watermark at different attributes at different places. Hence, it is difficult for an attacker to remove watermarks from the database. The proposed technique is work by inserting watermark bits in such a way that it should causes minimum distortion in data and data usability must remain intact after data has been watermarked. Most of the available techniques are depend on presence of primary key attribute. Proposed work provide solution if there is no primary key.*

Key Words: **Data usability; Knowledge preservation; ownership protection; watermarking**

1. INTRODUCTION

Relational Databases most often contain critical information which is used for decision support system. So there is need to secure this data. Hence data contributors need some technology that detects unauthorized access/changes to their databases. Security of relational databases through watermarking becomes an emerging research topic. Watermarking has been used to protects relational data as it contains crucial information. The aim of Watermarking is to secure a data from unauthorized changes and from being copied. Watermarking allows owner of the data to embed watermark bits into the data. The purpose of watermarking is that the watermark must not be easily forged or removed from the watermarked data [16].Watermark presence should be unnoticeable to naked-eye.

Most of the watermarking techniques are available for multimedia contents. But watermarking used for relational database is different from that of the used for multimedia. Watermarking relational data is somewhat complex task and hence watermarking relational databases literature is very limited and it focuses mainly on encoding binary bits in randomly chosen location in databases. Watermarking relational database is a complex task as it significantly changes data. Consideration of some constraints in the technique helps to optimize the amount of acceptable changes of numeric attribute while watermarking. As most of relational database watermarking techniques possess similar steps i.e. partitioning data using primary key attribute and then insert watermarking bits in selected attribute of chosen tuples by modifying its original value. These type of partitioning become a concern because watermark bits are embedded in multiple attributes at a same time. Most of these techniques are depend on presence of primary key. So, motivation behind this work is to design a system that embeds watermark bits at multiplace and is independent on presence of primary key attribute.

The remainder of paper is organized as follows: Section 2 discuss the related work done and its shortcoming. An overview of the proposed scheme is given in section 3. Section 4 discusses the expected results of the system. Finally, we conclude the paper

2. LITERATURE SURVEY

In this section, an overview of existing watermarking techniques for relational database are provided. The objective of this survey is clearly understand the limitations of existing schemes.

Table. 1 shows an overview of existing watermarking techniques. Existing methods that are available for Watermarking relational database contain similar steps. Firstly whole dataset is partitioned into non-overlapping partitions and then subset of dataset has been chosen in which watermark bits are encoded by slightly modifying the original value of that attribute. These type of partitioning become an issue because watermark bits are encoded in multiple attributes at a same time. But in proposed technique, the data is partition into non overlapping clusters of tuple. Watermark bits are then encoded into selected attribute of chosen tuples of each cluster. In this way, watermark bit sequence is embedded into multiplace i.e. in each cluster. And hence, it will difficult for an attacker to erase/destroy watermark bit sequence.

We will accurately decode watermark even if number of rows or even a whole cluster is destroyed in attack.

Table 2 shows the notation used in Paper.

Table-1: An overview of existing watermarking techniques.

Sr.No.	Paper Name	Techniques	Limitations
1	Agrawal and Kiernan,2002 [3], Agrawal et. al. 2003 [1],Y. Li et. al. 2006 [13],Zhou et. al. 2006 [14] Gupta et.al. 2009 [4]	Bit-level watermarking technique	Assumes unconstrained LSB manipulation during watermark embedding process. An attacker can easily destroy watermark. (for example shifting LSB)
2	Zhang, et.al. 2004. [7]	Image-based watermarking technique	Bits are not recovered correctly as it is used to reconstruct the embedded image.
3	Al-Haj and Odeh, 2008 [6]	Space-based watermarking technique	Suffer from watermark removal attack.
4	R. Sion et. al. 2008[8]	Statistical-property watermarking technique	An attacker can corrupt the watermark by launching large scale attacks on large number of rows. Decoding accuracy is degrade.
5	M. Kamran et.al.2013 [2]	Random bit level watermarking technique	Depend on presence of primary key attribute. Embedded Watermark is not imperceptible

Table -2: Notation Used

Notation	Description
Do	Original data
Dw	Watermarked data
Uc	Usability constraints
Ks	Secret key
W	Encoded watermark bit sequence
W'	Decoded watermark bit sequence
A	Selected attribute
m	Marker selection
b	Bit location used for inserting watermark
v	Number of attributes used in watermarking
W	Encoded watermark bit sequence

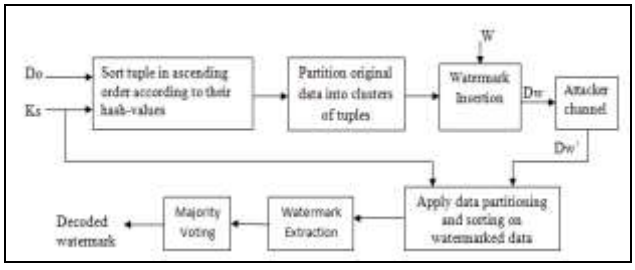

Fig 1:Block diagram of proposed system

3. PROPOSED WORK

A. Proposed System:

The overall Block diagram of the system is shown in the Fig.1. And Table I lists the symbols used in this paper.

Function of each block is discussed below: Overall process is divided into two parts.

1. Encoding stage
2. Decoding stage.

1) Watermark bits Generation: Watermark bit sequence W is generated by calculating Mac of secret key Ks. Number of tuples in cluster determines the length of watermark bit sequence.

2) Watermark Encoding Process: Watermark encoding process can be summarized as follows:

a) Sorting: Calculate MAC of each tuple and sort tuples in ascending order according to tuples MAC value.

b) Data Partitioning: The dataset Do is partitioned into non-overlapping clusters using secret key.

1. For each tuple r € Do do

2. r.MAC = H(Ks || MSB(r.A) || Ks

3. if(r.MAC mod m == 0)

4. then create cluster and store its start and end location.

c) Watermark Insertion: Watermark bits are encoded in selected attribute of each tuple using robust watermarking bits.

1. For all tuples, r € Do do

2. attr_sel = (r.MAC) mod v // Attribute selection

3. bit_sel = (r.MAC) mod b // bit location selection

4. Insert watermark bit at that location considering usability constraints.

3) Watermark Decoding Process: Watermark decoding process can be summarized as follows:

a) Sorting and Partitioning: The sorting and data partitioning are generated using procedure used in watermark encoding phase.

b) Watermark Extraction: For each cluster, first find the selected attribute and bit location. Then bit at that location take as decoded watermark bit.

c)Majority Voting : Majority Voting can be perform to correctly decode an encoded watermark bit. Table. 3 shows how does majority voting take place.

Table-3: Majority Voting for Proposed Work with "X" representing an error

Cluster1	1	0	X
Cluster2	1	0	0
Cluster3	1	0	0
Decoded Watermark	1	0	0

B. Attacker Model:

The attackers dilemma is to weaken or even destroy the inserted watermark and at the same time keep the data usable. We assume that, an attacker has no access to the original data and doesn't have any knowledge about any of the secret parameters used during watermark encoding stage. As an attacker doesn't have any idea about secret parameters, s/he cannot attack data as his aim is only to destroy the watermark and not the data.

4. EXPERIMENTAL SETUP

All Experimentation is performed using Pentium processor and 4 GB RAM. The operating system is windows 7(32 bit) with JDK1.6 and eclipse standard kepler.

A. Dataset:

Experiments are performed using Forest Covertype data [11] having 581,012 instances (observations) and 55 Attributes. First 10,000 tuples are used during experimentation and first ten integer-valued attributes for watermarking. Dataset characteristic is numeric.

A. Performance Measure:

The proposed system should preserve the knowledge in database. Watermark insertion changes the data but information loss should be zero. To prove this we will calculate information loss on database before and after watermark. This part of proposed system verifies whether the applied watermarking technique is lossless or not.

Where I is original database information and Iw is information obtained after watermark insertion. The knowledge is preserved if Iloss = 0. So this verification system checks whether the knowledge is preserved or not. We can also calculate percentage of loss in knowledge (if any) using above formula. We also check decoding accuracy of proposed system by comparing the encoded watermark bit sequence with decoded watermark bits. Decoding accuracy will be 100 % if W=W. Table 4 shows the information loss in terms of mean and standard deviation for proposed system.

B. Results:

1) Attack analysis: Fig.2, Fig. 3. and Fig.4. shows robustness against deletion, insertion and alteration attack respectively.

2) Information Loss: One of the aim of proposed technique is minimum data distortion. Table. III shows Information loss cause by proposed technique in terms of mean and variance and by technique given in [2]. One can easily noticed that loss in information is too small, hence requirement for minimum data distortion is satisfied.

Table IV. Shows Change in variance introduced by watermarking. System experimentally evaluate the impact of watermarking on the mean and variance of attributes values. For each attribute, this change in variance is very small. This is because technique used for watermark insertion and selection of subset of data for watermarking.

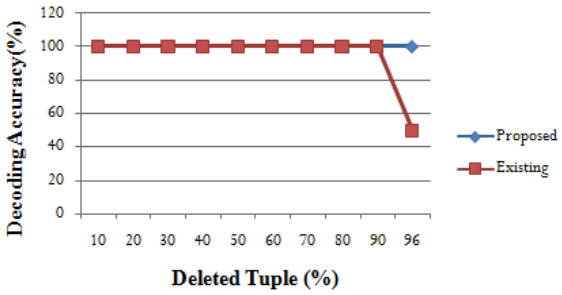

Fig. 2. Robustness of proposed system against deletion attack

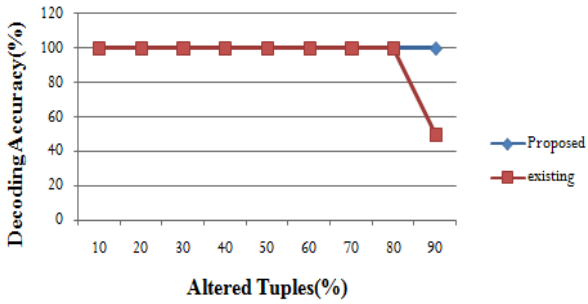

Fig. 3. Robustness of proposed system against alteration attack

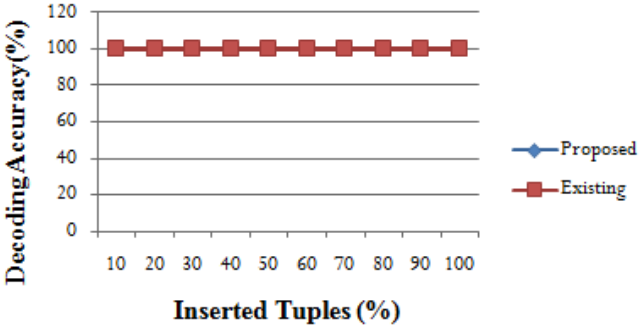

Fig. 4. Robustness of proposed system against insertion attack

Table-4: Information Loss

Information Loss	Existing	Proposed
Mean	0.0609%	0.0079%
Standard Deviation	0.339%	0.005%

5. CONCLUSION

Watermarking relational database is a solution to protect information from unauthorized duplication and to provide ownership protection. A robust and efficient watermarking

Table-5: Change in variance introduced by watermarking

Attributes	Mean	Variance	Approach given by[1]	Approach given by[2]	Proposed Approach
Elevation	2732.989	173984.1433	3	-0.2497	0.1379
Aspect	157.9701	12321.82137	1	2.6619	-0.00644
Slope	16.41234	73.2012282	1	-0.02373	0.001564
Horz-Dist-To-Hydrology	227.4865	44351.46041	1	2.99531	-0.058
Vert-Dist-To-Hydrology	52.07411	3593.19838	2	1.02952	0.4785
Horz-Dist-To-Roadways	1736.252	1879265.583	-9	-15.0338	-2.1388
Hillshade-9am	212.4666	952.9112703	1	0.3389	0.01955
Hillshade-Noon	218.7591	524.8878147	1	-0.6928	0
Hillshade-3pm	135.2411	2144.748181	1	0.846822	-0.01741
Horz-Dist-To-Fire-Points	1536.944	1276680.654	-3	0.867	0

scheme is introduced in this paper to handle vital information in relational data. Usability and robustness of system remains intact after data has been watermarked. As relational database watermarking significantly changes data.

This change in data may loss the knowledge in the dataset. Proposed system cause minimum distortion in data and hence preserves the knowledge in data. Also technique embeds each watermark bit in every selected attribute of chosen tuple of each cluster. Hence it will difficult for an attacker to destroy the watermark bit from database. Proposed scheme also provide solution if there is no primary key attribute.

In future, we will extend this research for non-numeric databases.

REFERENCES

[1] Agrawal, P. Hass and J. Kiernan, 2003.Watermarking relational data: Framework, algorithms and analysis. Int. J.Very Large Data Bases, 12:157-169.

[2] M. Kamran, Sabah Suhail, and Muddassar Farooq, "A Robust, Distortion Minimizing Technique for Watermarking Relational Databases Using Once-for-All Usability Constraints", IEEE TRANSACTIONS on knowledge and data engineering, vol. 25, no. 12, december 2013.

[3] Agrawal, R. and J. Kiernan, 2002. Watermarking relational databases. Proceeding of the 28th International Conference on Very Large Databases, Hong Kong, China, pp: 1-12.

[4] Gupta, G. and Pieprzyk, J. "Database relation watermarking resilient against secondary watermarking attacks" In Proceedings of the 5th International Conference on Information Systems Security, 2009 pages 222-236, Kolkata, India. Springer LNCS, Volume 5905. K. Elissa, "Title of paper if known," unpublished.

[5] Bhattacharya and Cortesi, "Database authentication by distortion- free watermarking" In Proceedings of the 5th International Conference on Software and Data Technologies, 2010.

[6] Wu, M., E. Tang and B. Liu, 2000. Data hiding in digital binary image. Proceeding of the IEEE International Conference on Multimedia and Expo, July 30-Aug. 02, New York, USA.

[7] Zhang, Z., X. Jin, J. Wang and D. Li, 2004. Watermarking relational database using image. Proceeding of the International Conference on Machine Learning and Cybernetics, Aug. 26-29, IEEE Explore Press, USA., 2004.

[8] R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," IEEE Trans. Knowledge and Data Eng., Vol. 16, no. 6, pp. 1509-1525, Dec. 2008.

[9] A. Deshpande and J. Gadge,"New Watermarking Technique for Relational Databases," Proc. Second Intl Conf. Emerging Trends in Eng. And Technology (ICETET), pp. 664-669, 2009

[10] Mohamed Shehab, Elisa Bertino, Arif Ghafoor, "Watermarking Relational Databases Using Optimization-Based Techniques," IEEE Transactions on Knowledge and Data Engineering,vol. 20,no. 1, 2008

[11] http://archive.ics.uci.edu/ml/datasets/covertype.

[12] Y. Li and R. Deng,"Publicly Verifiable Ownership Protection for Relational Databases," Proc. ACM Symp. Information, Computer and Comm. Security, pp. 78-89, 2006

[13] H. Guo, Y. Li, A. Liu, and S. Jajodia, "A Fragile Watermarking Scheme for Detecting Malicious Modifications of Database Relations," Information Sciences, vol. 176, no. 10, pp. 1350-1378, 2006.

[14] X. Zhou, M. Huang, and Z. Peng, "An Additive-Attack-Proof Watermarking Mechanism for Databases, Copyrights Protection Using Image," Proc. ACM Symp. Applied Computing, pp. 254-258, 2007.

[15] Y. Wang, Z. Zhu, F. Liang, and G. Jiang, "Watermarking Relational Data Based on Adaptive Mechanism," Proc. Intl Conf. Information and Automation, pp. 131-134, 2008.

[16] R. Halder, S. Pal, and A. Cortesi,"Watermarking Techniques for Relational Databases: Survey, Classification and Comparison," J. Universal Computer Science, vol. 16, no. 21, pp. 3164-3190, 2010.

[17] Qin Z., Ying Y., Jia-jin L. and Yishu L. (2006). "Watermark based copyright protection of outsourced database," In Proceedings of the 10th International Database Engineering and Applications Symposium (IDEAS06), pages 301308, Delhi, India. IEEE Computer Society