International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 7, October 2014, PP 248-252
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Relational Database Watermarking

Akshay Mayekar[1], Megha Jha[1], Sheetal Mule[1], Chidvilasinee Shridattopasak[1]

[1]Computer Engineering, Dnyanganga College of Engineering and Research, Pune, India

Abstract: *Proving ownership rights on outsourced relational databases has become a important issue in today's internet-based environments and in many content distribution applications. In this paper, we present a mechanism for proof of ownership based on the secure embedding of a robust imperceptible watermark in relational data. We formulate the watermarking of relational databases as an optimization problem and discuss efficient techniques to solve the optimization problem and to handle the constraints. Our watermarking technique is resilient to watermark synchronization errors because it uses a partitioning approach that does not require marker tuples. Our approach overcomes a major weakness in previously proposed watermarking techniques. Watermark decoding is based on a threshold-based technique characterized by an optimal threshold that minimizes the probability of decoding errors. We implemented a proof of concept implementation of our watermarking technique and showed by experimental results that our technique is resilient to tuple deletion, alteration, and insertion attacks.*

Keywords: *Relational Database, Database Watermarking, Robust Watermarking Techniques, Digital Assets, Watermark Encoding*

1. INTRODUCTION

The rapid growth of Internet and related technologies have opened new horizons in the social and business domain and has re-defined traditional perceptions of fields such as trade, banking and social welfare[1].

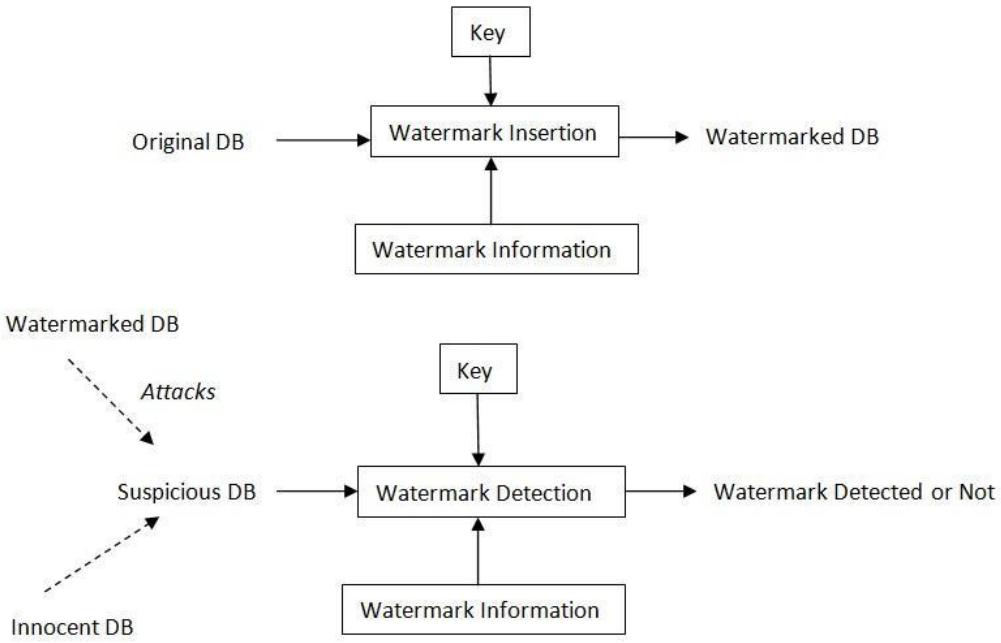

Fig1. *Watermark embedding and decoding*

These technologies have offered an uncontrollable access and distribution of digital contents. In such a condition the problem of ownership becomes a crucial topic. As these contents are freely distributed over internet the security measures fail to provide sufficient security to these contents[1][7]. In the past few years the term watermarking has emerged in the field of ownership protection for digital contents. The watermarking techniques allows the watermark to be securely embedded into the data. The watermark describes the information about the owner, origin or recipient of the data. The watermark is chosen in such a way that it remains unnoticed even after embedding and it is difficult to separate such a watermark from the data. Watermark adds a level of copyright protection to the digital contents[7].

On the other hand, the problem of relational database watermarking is not given the sufficient attention. In the real world there are many fields where data denotes a very important entity. For example, weather data, stock market shares, scientific research data. This realistic data can often tolerate some amount of errors. There can be maximum of 4% of degradation in the usability of data. For e.g. In salary of an employee change from 28,375 to 28, 382 is acceptable [7].

To date, there are very less watermarking approaches that are designed for relational database watermarking. But they are not resilient to watermarking attacks like tuple addition attack, tuple deletion attack, tuple alteration attack, etc [3].

1.1. Related Work

Most of the previous work has been done on LSB of the tuple. Agarwal and Kiernan suggested an algorithm to add watermark to the selected LSB bit of selected tuple. This technique doesn not offer for multibit watermarks[1]. Sion et el. Proposed a watermarking technique to insert the watermark into the data statistics[3]. But this technique again allows the data to be vulnerable to the synchronization attacks. Also this techniques makes watermark loose its property of blind detection.

1.1.1. Drawback of Existing System

- Existing algorithms focus only on LSBs of the tuple.
- Multi bit and multi attibute watermarking is not offered by existing system.
- The existing algorithms are not resilient to different watermarking attacks.
- Joint ownership concept is not supported yet.

1.2. Proposed System

In the proposed system, our main focus is to design a robust watermarking technique against different kinds of attacks. We are proposing the algorithm which partitions the data according to the given key into a number of partitions. So that number of guesses for the attacker increases. Also we are introducing the multi column watermark along with multi-bit watermark. The algorithm uses threshold based watermark decoding algorithm. So the probability of 100% recovery of watermarked data increases. One more thing the proposed algorithm offers is option for Joint Ownership. There can be a number of owners for the database. In this case they can insert their own private keys which then will be combined into a single private key. In this way every owner can feel secure for his/her ownership on data and no single person can claim ownership on it.

1.2.1. Advantage of Proposed System

- The algorithm supports blind decoding. That is in order to decode the watermark, knowledge of neither original data nor watermark is required.

- The watermarking techniques uses partitioning technique so it makes difficult for an attacker to guess the watermarked data.
- The algorithms uses multibit and multi column watermarking technique. So it is less prone to different watermarking attacks.
- The algorithm supports joint ownership. So it becomes difficult for any single person to claim for ownership.

2. APPROACH OVERVIEW

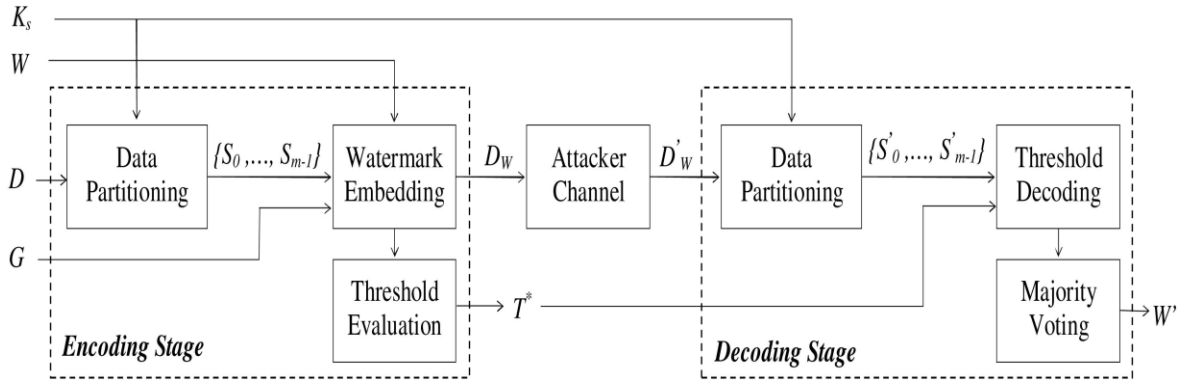

Fig2. *Watermark encoding and decoding stages.*

Fig. 2 shows the block diagram for stages of the watermark encoding and decoding. The input to the watermark encoding algorithm is dataset D, private key K_s which is known only to the owner and the watermark W. The encoding algorithm transforms the dataset D into D_w. The encoding algorithm also evaluates the threshold value which will be later used in watermark decoding algorithm

In the decoding phase of the watermark, the suspicious database is taken as input to the algorithm. Based on threshold evaluation, the algorithm identifies which rows are marked and decodes the watermark which is output of the system.

For data partitioning and encoding decoding, the algorithm makes use of highly encrypted Message Digest 5 algorithm.

3. ALGORITHM

There are two algorithms presented using this technique as given below.

3.1. Encoding Algorithm

3.1.1. Data Set Partitioning

First the owners of data have to enter their secret keys which will be combined into a single secret key. By using the secret key K_s, the data set D is partitioned into m non overlapping partitions {S0……Sm-1}. For the purpose of partitioning it uses the MD5 algorithm which takes variable length input K_s and generate a fixed length 128 bit value.

$$\text{Partition}(r) = H(K_s \,||\, H(P_k \,||\, K_s)) \bmod m$$

Where K_s is the secret key P_k is primary key and m is the number of partitions.

3.1.2. Watermark Embedding

A watermark bit is embedded in each partition by altering the partition statistics while still verifying the usability constraints in G. This alteration is performed by solving constrained optimization

problem. Also while inserting the watermark the X_{max} and X_{min} statistics are evaluated which will be used in stage 3.

3.1.3. Optimal Threshold Evaluation

The bit embedding statistics are used to compute the optimal threshold T using the X_{max} and X_{min} that minimizes the probability of decoding error.

3.2. Decoding Algorithm

3.2.1. Data Set Partitioning

First the owners of data have to enter their secret keys which will be combined into a single secret key. By using the secret key K_s, the data set D is partitioned into m non overlapping partitions {S0……Sm-1}. For the purpose of partitioning it uses the MD5 algorithm which takes variable length input K_s and generate a fixed length 128 bit value.

$$\text{Partition}(r) = H(K_s \parallel H(P_k \parallel K_s)) \bmod m$$

Where K_s is the secret key P_k is primary key and m is the number of partitions.

3.2.2. Threshold-based Encoding

The statistics of each partition are evaluated, and the embedded bit is decoded using a threshold-based scheme based on optimal threshold T.

3.2.3. Majority Voting

The watermarked bits are decoded using a majority voting technique. This technique evaluates the hiding function θ and compares it with threshold statistics T. If hiding function θ is greater than threshold T the bit is evaluated to 1. Otherwise the bit is evaluated as 0.

4. CONCLUSION

Our system introduces the concept of joint ownership that allows a number of owners to feel secure about their data. A major advantage of using this approach is a larger bit capacity is available to hide the watermark bits. The proposed technique must be suitable for different areas like, e-banking, multimedia industries, film industries etc. The proposed algorithm keeps the data safe from additive attacks.

REFERENCES

[1] R. Agrawal and J. Kiernan, "Watermarking Relational Databases," Proc. 28th Int'l Conf. Very Large Data Bases, 2002.

[2] F. Petitcolas, R. Anderson, and M. Kuhn, "Attacks on Copyright Marking Systems," LNCS, vol. 1525, pp. 218-238, Apr. 1998.

[3] R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 6, June 2004.

[4] L. Vaas, "Putting a Stop to Database Piracy," eWEEK, Enterprise News and Revs., Sept. 2003

[5] Halder R., Pal S., and Cortesi A.: "Watermarking techniques for relational databases: survey, classification and comparison". *Journal of Universal Computer Science (JUCS)* Vol.16, No.21, pp.3164-3190, 2010.

[6] Li Y., Guo H., and Wang S.: "A multiple-bits watermark for relational data". *Proceedings of the Principle Advancements in Database Management Technologies*, pp.1-22, 2010.

[7] Shehab M., Bertino E., and Ghafoor A.: "Watermarking relational databases using optimization-based techniques". *IEEE Trans. Knowl.Data Eng. (TKDE)*, Vol. 20, No.1, pp.116-129, 2008.

[8] Zhou X., Huang M., and Peng Z., "An additive-attack-proof watermarking mechanism for databases' copyrights protection using image". *Proceedings of the SAC 2007*, pp.254-258, 2007.

[9] Wang H., Cui X., and Cao Z.: "A Speech Based Algorithm for Watermarking Relational Databases". *Proceedings of the ISIP 2008*, pp.603-606, 2008.

[10] R. Agrawal, P. Haas, and J. Kiernan, "Watermarking Relational Data: Framework, Algorithms and Analysis," The VLDB J., vol. 12, no. 2, pp. 157-169, 2003.

[11] 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.

[12] S. Shah, S. Xingming, H. Ali, and M. Abdul, "Query Preserving Relational Database Watermarking," Informatica, An Int'l J. Computing and Informatics, vol. 35, no. 3, pp. 391-396, 2011.

AUTHORS' BIOGRAPHY

Akshay Mayekar is pursuing Bachelors of Engineering Degree from Dnyanganga College of Engineering and Research, Pune-58. He is responsible for giving the module ideas.

Megha Jha is pursuing Bachelors of Engineering Degree from Dnyanganga College of Engineering and Research, Pune-41. She is responsible for design analysis of the system.

Sheetal Mule is pursuing Bachelors of Engineering Degree from Dnyanganga College of Engineering and Research, Pune-41. She is responsible for overall idea of the system.

Chidvilasinee Shridattopasak is pursuing Bachelors of Engineering Degree from Dnyanganga College of Engineering and Research, Pune-41. She is responsible for overall idea of the system.