PROPRIETORSHIP PRESERVATION OF RELATIONAL DATABASE USING WATERMARKING

Arti Ganesh Rajguru
ME Computer Engineering, Matoshree college of Engineering and Research Center, Maharashtra, India

ABSTRACT

The watermarking plans to be produced for accomplishing 'ownership protection on relational database' needs to fulfill the subsequent difficulties a] Resistant to various sorts of assaults b] Able to protect and maintain the integrity of the data in databases so it can be used effectively by any decision support system c] Maintain equalization between the inverse requests of database proprietors and users d] Ensurement of minimum records loss inside the information e] Usability restrictions for capacities and users. The work puts forward a sturdy watermarking system for relational database that tries to conquer the above cited difficulties. Watermark is placed in the information of characters, integers and unsigned integers data types. To approve the robustness, the framework will be tried for different assaults. This method cannot stop the unlawful copying of the data, but still it can assist to achieve it by presenting a manner to identify the original owner of the duplicated data.

Keyword: *- Data quality, Robust watermarking, Relational data*

1. INTRODUCTION

Watermarking can be defined as a way by which several kinds of information can be interlaced into underlying data for ownership authentication, tamper detection, localization, and/or defector tracing purposes. Watermarking methods can be applied to various types of mass content. In relational databases, Rights/Ownership protection of such data is critical in situations where data are sensitive, valuable and on the other hand they need to be delegated. E.g. Data mining application, where data are sold in sections to the parties, specialized in taking out it. In the environment most of the data it is hard to connect rights of the inventor over it. So Watermarking can help us to solve these problems. Unlike hash decryption and encryption, typical watermarking techniques modify the ordinal data by making it a modulation form of the watermark information. This predictably causes permanent distortion to the original data and therefore these data cannot meet up the integrity requirement of some applications [6].

Innovative watermark decoding algorithm make sure that its decoding accuracy is independent of the usability constraints Our approach make possible Alice to define usability constraints only once for a particular database for every possible type of suggested application. it also take care that the watermark set ups the smallest amount possible distortions to the original data without losing the robustness of the inserted watermark.[1]

This watermark scheme consist of various types of attacks which are as follows

Deletion Attack

Attacker can also arbitrarily delete the tuples or selects them in a complicated manner on the basis of their statistical allocation of attribute values

Insertion Attack

There are two types of insertion attack 1)fixed insertion and 2) constraint reliant insertion. In the first attack, attacker inserts new n tuples by duplicating values of existing g tuples. In the second attack, he makes the Tuple values based on the mean μ and the standard deviation σ of watermarked data set.

Alteration Attack
This attack try to destroy the watermark by altering one or more bits in the watermarked data. More additional information about the marked bit position makes attack more successful[1]

2. RELETED WORK

In Agrawal and Kiernan [3] projected a bit resetting algorithm .It sets the LSB of the candidate attribute of the selected subset of tuples. The selection of watermarking parameter is based on calculating Message Authenticated Code (MAC), where MAC is calculated using the Tuple primary key and the secret key. This technique assumes unconstrained LSB exploitation during watermark embedding process. However, unwanted results might also get generated due to such out-of-bound modification of data. Although LSB-based data hiding techniques are efficient, the disadvantage of this technique is that an attacker can easily eliminate watermark by simple manipulation of data by using Bit-resetting techniques.
In [10] Sion et al, it projected a statistical-based algorithm in which a database is divided into a large number of unique, nonintersecting subsets of tuples. The concept of data partitioning is based on the use of special marker tuples, making it disposed to watermark synchronization errors, particularly in the case of Tuple insertion and deletion attacks, as the position of marker tuples is troubled by these attacks. Such errors may be minimized if marker tuples are stored during watermark embedding stage .Also the marker tuples can be used for creating the data partitions again in the watermark decoding stage. Disadvantage of this technique is that using the stored marker tuples to reconstruct the partitions violates the requirement of blind decoding of watermark
Content characteristics as watermark information [9] In these techniques, the extracted information from the database contents are treated as watermark and it can be embedded in the data in the same database. But since the content keeps changing under various operations in que the modification of the content may lead to the extracted information differently in the verification phase from that in embedding phase .We can use this method in order to take out only the invariant properties and embed it into the stable part of the database content.

2. PROPOSED SYSTEM

Following are the flow of proposed system:

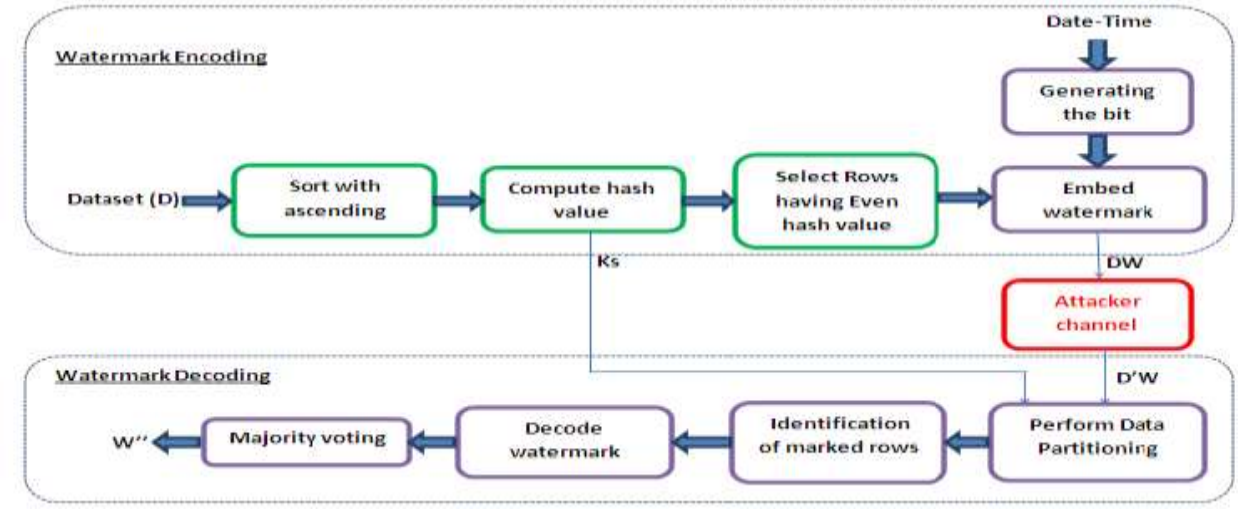

Fig 1 Architecture of the system

Fig. 1 shows that we select dataset which is freely available in the number format after that with the help of Algorithm 1 we make the partition by computing their hash value with the help of primary key, secret key, no of partitions and it can be divided into different clusters and then sort it in Ascending order. Then apply Algorithm 2 to calculate Threshold to make an attacker more difficult to gain on to the systemthen select only those rows whose value are greater than the threshold. i.e. we are selecting only even hash value rows whose values are comes out to be zeros. Then apply Algorithm 3 to select only those rows whose hash value are Even. After selecting even rows apply Algorithm 4 to start embedding the data with the help of Date and time. then the watermark bit length is generated .its bit length is depend upon total no of bits are use for watermarking then Apply Algorithm 5 to Decode the data still if any error remains, apply Majority Voting to remove that error.

3. RESULT ANALYSIS

Here dataset use is forest cover type .This dataset is in the form of numbers i.e. 0's & 1's .for implementation we have selected 1000 Rows and 55 Columns. another dataset use for character is Artificial character dataset which is UCI Machine learning dataset. Main concept is that when attacker delete the no of rows from dataset still it achieves 100% decoding accuracy and give the original data back.[1]
We also try to prove that our results produce minimum information loss and less data distortion as compared to previous paper .results are shown as follows:

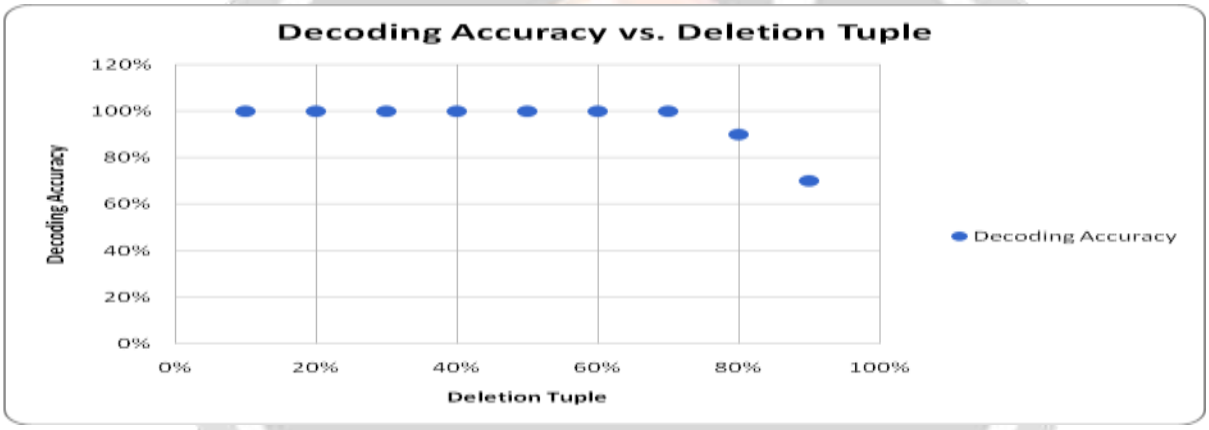

Fig 3.1 Decoding accuracy v/s Deletion Tuple

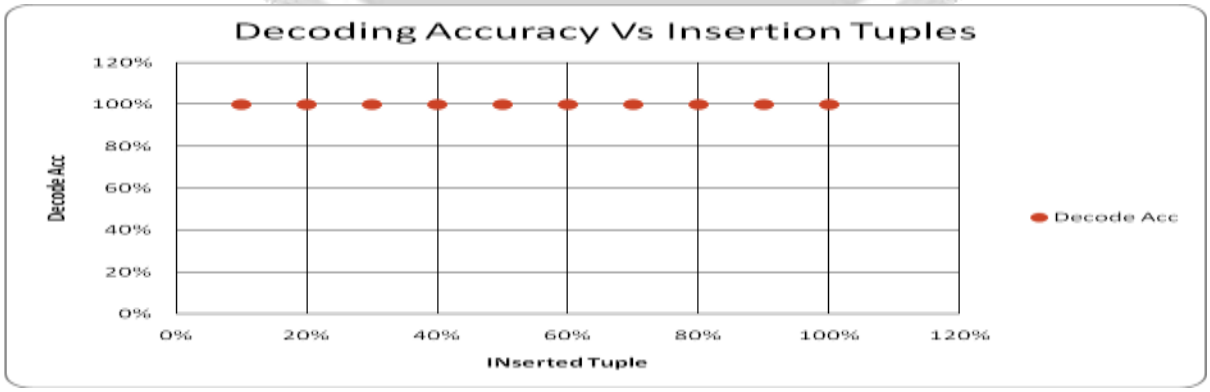

Fig 3.2 Decoding accuracy v/s Insertion Tuple

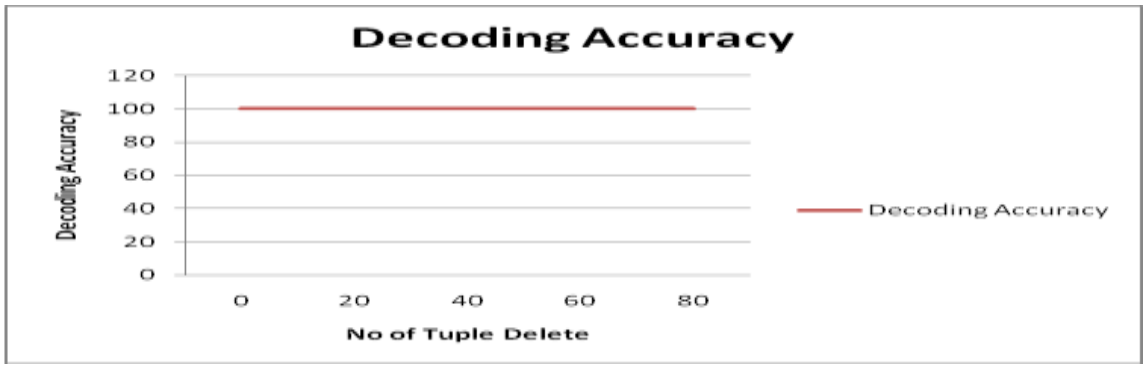

Fig 3.3 Decoding accuracy

4.CONCLUSIONS

In proposed system which is highly flexible against different types of attack, i.e. insertion, deletion, alteration attack which results in minimum information loss in the original data set. In spite of the severity of malicious attack on the watermarked data, the watermark bits are successfully interpreted with 100 percent accuracy because the accuracy of the system is independent of the usability constraints. Our technique provides maximum possible robustness and minimum data distortion.

5. REFERENCES

[1] Kamran, M., Suhail, S.; Farooq, M., "A robust, distortion minimizing technique for watermarking relational databases using once-for-all usability constraints," IEEE Transactions on Knowledge and Data Engineering, on ,vol.PP, no.99, pp.1-1, Dec 2013.

[2] M .Shehab, E. Bertino, and A. Ghafoor, Watermarking Relational Databases Using Optimization-Based Techniques, IEEE Trans. Knowledge and Data Eng., vol. 20, no. 1, pp. 116-129, Jan. 2008.

[3] 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.

[4] S. Bhattacharya and A. Cortesi, A Distortion Free Watermark Framework for Relational Databases, Proc. Fourth Intl Conf. Software and Data Technologies (ICSOFT 09), pp. 229-234, 2009.

[5] R. Halder and A. Cortesi, A Persistent Public Watermarking of Relational Databases, Proc. Intl Conf. Information Systems Security, pp. 216-230, 2011.

[6] 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.

[7] 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. 2004.

[8] Y. Wang, Z. Zhu, F. Liang, and G. Jiang, Watermarking Relational Data Based on Adaptive Mechanism, Proc. Intl Conf. Information and Automation (ICIA 08), pp. 131-134, 2008.

[9] R. Halder and A. Cortesi, A Persistent Public Watermarking of Relational Databases, Proc. Intl Conf. Information Systems Security, pp. 216-230, 2011.

[10] 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. 2004.