ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

A Robust and Semi-blind Reversible Watermarking for Relational Database

G.V.Sathiyapriya, S.Banumathi

M.E Student, Department of C.S.E, Arunai Engineering College, Thiruvannamalai, India

Asst. professor, Department of C.S.E, Arunai Engineering College, Thiruvannamalai, India

ABSTRACT: A range of watermarking techniques are being used over the path of the last few decades for ownership security of digital media such as images, audio, video, and natural language processing software. With the increasing fame of sharing data stores across the Internet, the same condition has evolve for relational databases. Data owners allow their data to be accessed and used at all; thereby revealing them to threat of data theft. though watermarking technology is helpful in proving ownership through discovery of data privacy, the problem is that it introduces permanent modifications in the data which are irreversible and cause the follow-on watermarked data to become different from the original content. accordingly, data analysis and decision making on the unclear version of data becomes impossible. Reversible watermarking of relational databases is a moderately new and promising area. The techniques for ownership protection and data recovery of relational data needs to think about new constraint based on facts such as (i) a database consists of tuples /records, which is where the watermark requests to be fixed; (ii) the ordering of the records in a database relation; and (iii) data operation such as insertion, deletion, and alteration that normally occur in a database. Moreover, depending on the type of data stored in the database, the available bandwidth or capacity for watermark embedding may vary and the same power be utilize to make the technique more tough but, the bandwidth consumed during watermark embedding should not be too huge so as to compromise the data quality. Data modification are allowed to such an extent that the quality of the data before and after embedding the watermark is acceptable. normally the data owner decides how much modification can be tolerated without compromise the quality.

KEYWORDS: Reversible watermarking, genetic algorithm, data recovery, data quality, robust.

I. INTRODUCTION

Data mining is the study of the "Knowledge Discovery in Databases" , where process of ascertain pattern in large data sets relating method at the meeting point of artificial intelligence, machine learning, and shared database systems etc. The main goal of the data mining process is to dig out the information from a data set and transform it into an understandable structure for future use. The large datasets generated from databases are being mined to extract secreted knowledge that are useful for decision makers to take effective, efficient and timely decisions in a competitive world.

The intention is to work in a shared environment and make data openly available so that it is useful for knowledge mining and decision making. Take the case of Walmart a large cosmopolitan retail firm that has made its sales list accessible openly over the Internet. However these openly available datasets make attractive targets for attacks.. Rakesh and Jerry did initial hard work in the domain of database watermarking to hide watermark bits using least significant bit. Their technique is based on watermarking. It was assumed that the relational database can allow small quantity of alteration in a database. conversely, inserting watermark bit into LSB may affect data as it can be easily manipulate by the attacker. In another study, Sion et al.targeted the selected tuples to hide watermark bits into partition statistics. Statistics were changed according to usability constraints. It is responsible to keep check on the values of the attributes so that the change did not go beyond the limit. Shehab et al. further enhanced the above method by using optimization techniques.

The major contributions of the work are: (1) the design of an intelligent reversible watermarking technique for relational data that ensures data recovery without compromising data quality, and (2) a robust data recovery scheme that is tough against subset modification, subset deletion and subset insertion attacks. It detects the watermark fully and recover the original data. The strength is evaluated through attack analysis, considering the attacker channel. It is value

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

mention that sound watermarking techniques exploit redundancy in the data to embed the watermark in a manner that it does not force the overall size. For example consider the case of image watermarking that embeds information in the least significant bit (LSB) of pixel values which is invisible as well as it does not effect the image size.

This is popular because changes in the size of the original data can cooperate the presence or absence of a watermark. Similarly one other contribution of the technique is that it avoids detection by keeping the size of the original relational data unaltered.

II. RELATED WORKS

The earliest irreversible watermarking technique for relational databases be proposed by Agrawal et al. In this technique, histogram extension is use for reversible watermarking of relational databases. Zhang et al. proposed a technique of circulation of error between two equally discrete variables and chosen some initial non-zero digits of errors to form histograms. Histogram expansion technique is used to reversily watermark the chosen non-zero initial digits of errors. This technique is keep path of transparency information to validate data quality. However, this technique is not robust beside heavy attacks .

Difference Expansion Watermarking techniques (DEW) utilize methods of arithmetic operations on numeric features and make transformation. The watermark information is normally fixed in the LSB of features of relational databases to reduce distortion. while, in robust reversible watermarking, a GA based best possible value is fixed in the selected feature of the dataset with the objective of preserving the data quality while minimizing the data distortions as a result of watermark embedding. Another reversible watermarking technique is based on difference expansion and support vector regression (SVR) prediction to defend the database from being tamper. The intent following the design of these techniques is to provide ownership proof. Such techniques are in danger to change attacks as any change in the lengthened value will fail to identify watermark information and the original data.

Genetic Algorithm based on Difference Expansion watermarking (GADEW) method is used in a intended robust and reversible solution planned for relational databases. GADEW recover upon the drawback mention above by minimize distortions in the data, increasing watermark capacity and lowering fake positive rate. To this end, a GA is occupy to increase watermark capacity and minimize introduce distortion. This is for the reason that the watermark capacity increases with the increase in numeral of features and the GA runs on added features to explore the best one for watermarking. On the other hand, watermark capacity decrease with the raise in watermarked tuples. GADEW used the alteration actions (AWD and TWD) to manage distortions in the resultant data. In this state, the robustness of GADEW can be cooperate when AWD and TWD are set high values.

Prediction-error expansion watermarking techniques (PEEW) add in a judge as apposed to a difference operator to select applicant pixels or features for embedding of watermark in order. The PEEW proposed technique by Farfoura et al. is easily broken against malicious attacks as the watermark information is fixed in the partial part of numeric features only. In this exacting situation, the plan works because the intent of the attacker is to preserve the usefulness of the data; otherwise, the attacker can easily cooperate the partial part. Reversible Watermarking is robust, as the watermark information is fixed in the values of numeric features, to make the plan flexible beside such attacks.

In RRW, all the tuples of the selected feature can be marked thanks to the selection of a low distortion watermark; therefore, the attacker will have to attack all the tuples to corrupt the watermark to mitigate the effect of the majority voting scheme. Attacking all the tuples is not a viable option for the attacker because he has no knowledge of the original data or the usability constraints and that would completely compromise its usefulness. Moreover, since RRW can afford to embed watermark bits in all or a large fraction of the tuples of the selected feature; it achieves high robustness against heavy attacks.

However, marking all tuples is not a requirement. RRW is configurable in that the data owner can choose a fraction for watermarking if it is required. RRW outperforms existing state of the art reversible watermarking techniques including DEW, GADEW and PEEW. These techniques embed the watermark in partitions of the data to ensure minimum distortion; therefore, recover original data with degraded data quality and lack robustness. Reversible Watermarking has defeat drawbacks of these techniques and is also resilient against heavy attacks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

III. SYSTEM ARCHITECTURE

This section discusses Robust and semi-blind for reversible watermarking of relational databases that improves data recovery ratio. The main architecture of RRW is presented in Fig. 1. Robust and semi-blind Reversible Watermarking includes the following four major phases: (1) watermark preprocessing; (2) watermark encoding; (3) watermark decoding; and (4) data recovery.

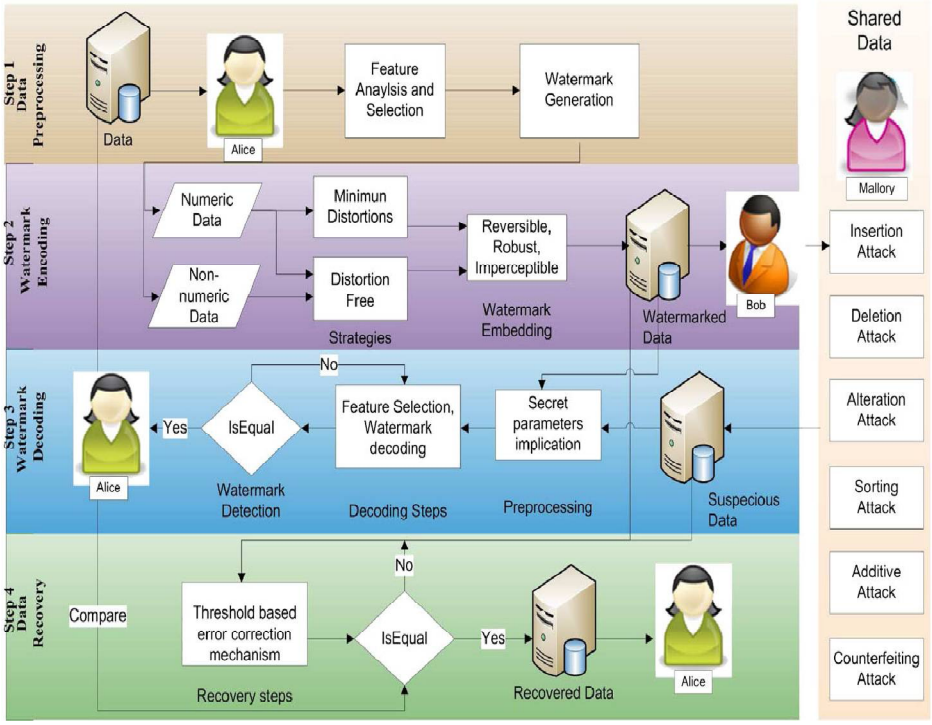

Fig1: System Architecture

A. *Watermark Encoding*

The watermark encoding algorithm starts the embedding method with the most significant bit MSB of the watermark. For this cause the algorithm workings with one tuple at a time. If the MSB of the watermark is 1, the new value of D_r (recoverd data) denoted by D_{Wr} and β (optimal value) is calculated using equation as follows.

$$D_{Wr} = D_r - \beta$$

In order to embed the second MSB of the watermark, the algorithm is again employed using the same procedure, but the updated value D_r of the feature (D_{Wr}) is used for calculating new values of D_r and D_{Wr} . If the algorithm encounter a watermark bit that is 0 then the new value of D_{Wr} is calculated using equation.

$$D_{Wr} = D_r + \beta$$

B. *Watermark Decoding*

The watermark decoding algorithm mainly consists of two steps: Step 1. For every candidate feature A of all the tuples in the watermark bits are detected starting from the LSB (least significant bit) and moving towards the MSB (most significant bit). The bits are detected in the reverse order compared with the bits encoding order because it is easy to detect the effect of the last encoded bit of the watermark. This process is carried out using the change matrix . Step 2. The bits are then decoded according to the percentage change values of watermarked data.The final watermark information is retrieved through a comparison of data.

C. *Attacker channel*

After watermarking, the data is released to the intented recipients over a communication channel that is assumed to be insecure and termed as “attacker Channel” in this research domain. The data may undergo several malicious attacks

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

in the attacker channel. The efficiency and effectiveness of Waternarking is describe during robustness analysis determined by its reaction to following attacks,

i. Subset insertion
ii. Subset alteration
iii. Subset deletion

D. Data Recovery

The original data are well again in the data recovery phase through post-processing steps for error correction and recovery. After recovery, the in good health data might also be compare with the original data to make sure that the data quality is not cooperate.

IV. ROBUSTNESS ANALYSIS

In the robustness study, experiments were performed with three types of attacks: (1) Insertion; (2) Deletion; and (3) Alteration. The original data recovery is reported for Robust Reversible Watermarking only and watermark detection rate is compared with DEW, PEEW , and GADEW techniques with different division of tuples. Robust and Reversible Watermarking is highly robust the same as compared to PEEW, GADEW and DEW techniques in the analysis of the three types of attacks. In Reversible Watermarking all the tuples of the selected feature were watermarked to attain robustness. GADEW watermarks a large number of tuples in two exacting features as compared to DEW and thus improve robustness in scenario of point-wise and tuple-wise complicated attacks. However, the addiction of robustness of GADEW on AWD and TWD requires a trade-off between the watermark robustness and the data distortion introduced throughout watermark embedding. As a result, the decoding correctness of GADEW degrade if data distortion are minimize. In exacting, its presentation is not acceptable in case of heavy attacks because watermarked tuples are affected by such attacks; as a result, it becomes difficult to decode the watermark correctly from the residual unaltered tuples. On the opposing, Robust Watermarking watermarks all the tuples of the chosen element and still provide a way of calculating data distortions although ensure data quality after watermark embedding.Following graphs describe the detection of values from various attacks.

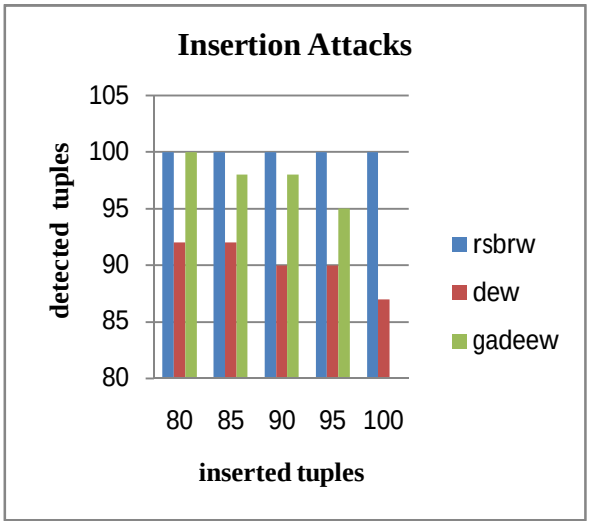

Fig 2 a.Insertion Attacks

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

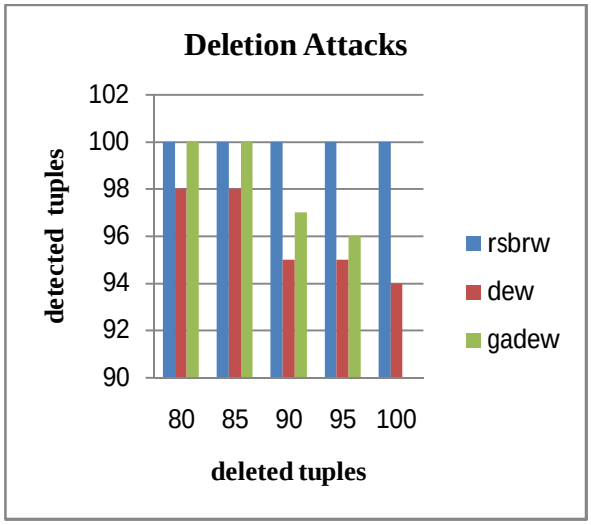

Fig 2 b. Deletion Attacks

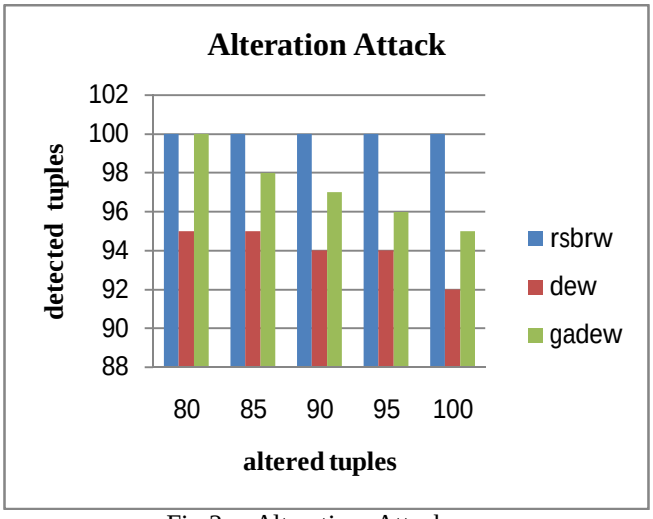

Fig 2.c. Alteration Attacks

V. CONCLUSION

In this paper, I have projected a Semi-blind reversible watermarking for relational database proposal. This proposal can prove the true possession of the database's owner, and attain full recovery of the original database relation once the watermark information is detected and valid. The watermarks are fixed into a database relative under the control of a secure embedding key. A majority voting mechanism was useful to correct the watermark bits detected from the data at watermark detection phase. Our future research will be directed towards increasing the level of attack flexibility against numerous sources of attacks in a classic blind reversible watermarking way, and propose fresh technique used for watermarking database relations with no primary keys in both numeric and non-numeric data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

REFERENCES

1. Swathi melkundi, chaitali chandankhede "A robust technique for relational database watermarking and verification," IEEE Transactions on Dependable , vol. 10, no. 2, pp. 70-83, March-April 2015.
2. Imamoglu " A watermarking technique for relational database based on partioning" IEEE transactions on dependable and secure computing, vol. 8, pp. 670-684, 2015.
3. D.srinath, "reversible watermarking technique based on time stamping", IEEE transaction on information forensics and security, vol. 7, n0. 5, 2015.
4. Desheng Fu, Shu Zhou, Chenglong Cao ,"A Study of relational database" international conference on information science and Technology, vol. 10, 2015.
5. Junghwan Rhee, Ryan Rileym Zhiqiang Lin,Xuxian Jiang, and Dongyan Xu. "A robust multiple watermarking technique for information recovery",IEEE transactions on Information Forensics and Security ,VOL. 9,No. 1,Jan 2014.
6. Thomas Arnold and T. Andrew Yang ,"Robust lossless watermarking of relational database based on circular histogram modulation"ACM Journal of Computing Sciences in Colleges, Vol 26,pp. 122-129, May 2014
7. Jeffrey Bickford, Ryan O'Hare, Arati Baliga, Vinod Ganapathy and Liviu Iftode "An empirical study on robustness of a fragile watermark for relational database" , Proc. Int'l ACM Conf. Computer and Comm. Security, *2014*
8. Eric Uday Kumar, "Watermarking relatrional database" springer journal in computer virology,vol 6,pp.123-141,may 2013
9. "Walmart to start sharing its sales data," http://www.nypost.com/p/news/business/walmart-opensup, last updated: 09:55 AM on February 4, 2012, last accessed: July, 20 2013.
10. "Identity theft watch," http://www.scambook.com/blog/2013/ 04/identity-theft-watch-customer-passwords-stolen-fromwalmart- vudu-video-service/, last updated: April 11, 2013, last accessed: July, 20 2013.