International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-04, July 2018

A Review on Reversible Watermarking Techniques for Relational Database

Ms. Akshata A. Churi, M.E Student, Shree.L.R Tiwari College of Engineering, Mira Road, India, Churiakshata333@gmail.com

Dr. Vinayak D. Shinde, H.O.D Computer Department, Shree. L.R. Tiwari College of Engineering, Mira Road, India.

***Abstract*— Many real world applications uses open databases which are available on the internet to extract information based on their needs. The relational databases which are freely available are used by research community for mining new information regarding their research works. These databases are vulnerable to security issues such as ownership of data and data tampering. The reliability of the data source must be verified before using it for any research or application purpose. Therefore to check ownership and reliability of data, watermarking is done to the data. When watermark is embedded to the database it reduces the quality of the data therefore that data is not useful for information retrieval. For applications areas such as the Government Organizations, Hospitals, Industry it is important to restore the original data as it is without any distortions. The watermarking techniques which satisfies these requirements are known as 'Reversible watermarking'. Reversible watermarking is designed to completely restore the original data without any distortion. There are many effective approaches that performs reversible watermarking to ensure ownership along with data recovery. Here, we reviewed some of the Reversible Watermarking techniques to identify each techniques pros and cons and to identify which technique will be useful for us and what are the future trends which need to be implemented.**

Index Terms*—. *Difference Expansion Based Watermarking (DEW), Genetic Algorithm (GA), Genetic Algorithm & Difference Expansion Based Watermarking (GADEW), Mutual Information (MI), Reversible Watermarking, Robust Reversible Watermarking (RRW).

INTRODUCTION

The advancement of information technology has boosted the growth of business and research. In many fields, data are extracted widely from various sources for information retrieval and decision making. Many real world application mine data available in different formats like text, audio, video, images and relational data to gather new ideas and information. Especially relational data which is more prominent among the scholar community is shared extensively by the researchers. Open databases are available on the internet for reference to scholars. However there is a possibility that these databases are hacked by attackers. The data are illegally copied by the attackers and causing threat to its ownership rights. The personal information of customer is also retrieved by the attacker causing major security issue for the data. In order to resolve these scenario, and to prove ownership of data, watermarking technique are being used for many years which effectively protects data being illegally copied by others. The watermark generated will be embedded to the original data which helps to identify the ownership of data. The data owner can easily identify their data if it contains a unique watermark. The issue regarding watermark is that, while embedding the watermark to the data, the database undergoes certain modification based on the bandwidth of the watermark causing the quality of data to be compromised. To overcome this disadvantage of watermarking techniques, reversible watermarking technique is introduced in which the embedded watermark can be retrieved by the data owner and the original data can be decoded from the watermarked data therefore the data quality is kept intact. For applications areas such as the Government Organizations, Hospitals, Industry it is important to restore the original data as it is without any distortions. The watermarking techniques satisfying these requirements are known as 'Reversible Watermarking'. Reversibility is ability to generate the original data from the watermarked data using a secret key. Moreover in Reversible watermarking the data owner can specify the distortion tolerance i.e. the amount of change in the data that can be allowed by owner while embedding watermark. Based on the distortion tolerance the watermark is embedded to the data.

 DOI : 10.18231/2454-9150.2018.0516

LITERATURE REVIEW

1. Reversible and Blind Database Watermarking Using Difference Expansion [1]

Proposed Technique => Difference expansion based watermarking (DEW) [1] technique is used to achieve reversibility in context of relational databases. DEW is able to restore the original database exactly. It also allows adding distortion into the database using distortion tolerance of the attribute. It also encourages the owner to distribute the trial version of the database, which can only be reverted by those users who have purchased the key.

Difference expansion watermarking techniques (DEW), performs some arithmetic operations on numeric features and perform transformations. The watermark information is embedded in the LSB of features of databases to minimize distortions.

Proposed reversible watermarking technique introduces distortions as a result of the embedding process. Changes in the data are controlled by placing certain bounds on LSB. On the other side, to limit the distortions, the data outside the limited bounds is left unwatermarked.

Difference expansion based watermarking (DEW) technique was used to achieve reversibility. DEW is able to restore the original database exactly. Additionally, it also allows adding distortion into the database using distortion tolerance of the attribute. It also encourages the owner to distribute the trial version of the database, which can only be reverted by those users who have purchased the key. Based on calculating difference value between 'two attributes' of selected tuples.

For Watermarking: (TVx = Target Value of x,
TVy = Target Value of y)
a = [TVx +TVy / 2] , d= TVx –TVy .....1)
d' = 2*d+b .....2) where b is watermark bit
CVx= a+[d'+1 / 2] , CVy= a – [d' /2].....3)
where CVx , CVy = watermarked data

For reverse watermarking:
a = [CVx + CVy / 2] , d = CVx – CVy .....4)
b = d'–2*[d' /2].....5) ← watermark bit b extraction
TVx = a+ [d'+1 / 2],
TVy = a – [d'/2]… 6) ← Original data extraction.

Table 1.1 Sample dataset for DEW

Emp_id	E_Name	Dept_id	D_name
101	Amit	200	ADMIN
104	Gaurav	100	HR
103	Rohit	300	ACCOUNTS

Two attributes = Emp_id, Dept_id on which watermark is going to be Embed
Two Target values TVx =104, TVy =100
By applying formulas we get:
a = (104+100)/02 = 102 d =104 – 100 = 4
d' = 2*d+b .....2) where b is watermark bit
d'= 2 × 4 + 1 = 9, if b = 1.
& CVx= a+[d'+1 / 2] , CVy= a – [d' /2].....3)
where CVx , CVy = watermarked data
CVx= 102 + ((9 + 1)/2) = 107 ,
CVy= 102 – (9/2) = 98.
CVx= 107 CVy= 98 ….. (Values after watermark)

Table 1.2 Watermarked Dataset by using DEW

Emp_id	E_Name	Dept_id	D_name
101	Amit	200	ADMIN
107	Gaurav	98	HR
103	Rohit	300	ACCOUNTS

For reverse watermarking :
a= [CVx + CVy / 2],
d'= CVx – CVy .....4)
a = ((107 + 98)/2) =102 and d'= 107 - 98 = 9
b = d' – 2*[d' /2].....5) ← watermark bit b extraction
b = 9 − (2 × (9/2)) = 1
d = (d' − b) / 2 so d = (9 − 1) / 2 = 4
TVx = a+ [d+1 / 2],
TVy = a – [d /2].....6) ← Original data extraction
TVx= 102 + ((4 + 1)/2) = 104,
TVy = 102− (4/2) = 100
Two Target Values received are
TVx = 104 TVy = 100

2. Genetic Algorithm and Difference Expansion Based Reversible Watermarking For Relational Databases [2]:

DEW [1] only checks two TVs in tuple rather than checking different combinations of TVs in the same tuple. If the CVs do not fulfill the distortion tolerance of their attributes, then the tuple is left unwatermarked and algorithm proceeds to the next selected tuple. This means less number of watermark bits are embedded in selected tuples.

Proposed technique (GADEW) [2] Genetic Algorithm and Difference Expansion Based Watermarking used (GA) Genetic Algorithm for watermark insertion in the selected tuple of a relation. Instead of just checking capacity of the selected attributes [1], it checks combination of different attributes in same selected tuple, and tries to finds the optimal pair of attributes.

Proposed approach maximizes watermark capacity as well as tries to reduce attribute and tuple-wise distortions. Watermark capacity increases when less tuples are selected for watermark. Comparing with DEW [1] if selected tuples for watermark increases watermark capacity also increases. Distortion introduced in data is high with DEW as compare with GADEW so GADEW is more secure than DEW.

 DOI : 10.18231/2454-9150.2018.0516

3. Robust and Reversible Watermarking Technique for Relational Data [3]:

Genetic algorithm is used in the Robust and Reversible Watermarking technique to achieve an optimal solution that is feasible for the problem and does not violate the defined constraints. An optimal watermark value is created through the Genetic Algorithm and inserted into the selected feature of the database in such a way that the data quality remains intact.

Proposed technique RRW [3] introduces concept of Mutual Information (MI) which measures the amount of information that one feature contains about the other features in a database,& after that mutual information(MI) value is used to select a suitable feature from the database for watermarking. Secure against attack such as subset alteration, subset deletion, and subset addition. It is applicable only to Relational Database having Numerical Values, not applicable to database having Textual databases.

4. Enhanced Robust and Reversible Watermarking For Supervised Learning Data [4]:

The proposed system [4] provides the process for watermarking the relational database using numeric and non-numeric attributes of database aiming to recover the watermarked data successfully.

Initially the textual data is converted to numeric values, Rest Process is same as RRW [3] , and at end again numerical attribute is converted to textual attribute. When we consider database of supervised learning training data there is no need to consider all other features. If we consider supervised learning data then importance of the feature is depends only on its co-relation with class feature therefore whenever there is supervised learning data then there is no need to find co relation of feature with all other feature except class feature. Therefore MI will be calculated only n times if there are n features.

Enhance feature selection for supervised learning dataset:

In this step co-relation of feature Fi is calculated only with class feature by using same formula as that in RRW [3]. In RRW this step is carried out for each feature with all other features i.e. if there are n features then this step is executed $2^{n-1} - 1$ times.

For example, n are 5 then 15 times this equation 1 will be executed. If we consider supervised learning data then importance of the feature is depends only on its co-relation with class feature therefore whenever there is supervised learning data then there is no need to find co-relation of feature with all other feature except class feature. Therefore equation 1 will be executed only n times if there are n features. For example there are 5 features then equation will be executed 5 times only. This will be very efficient when data is high dimensional when n is like greater than 100.

Proposed method works on non-numeric data and gives better results, reduces computation, storage and time requirements.

5. Genetic Algorithm Based Reversible Watermarking Approach for Numeric and Non-Numeric Relational Data [5]:

The proposed system [5] provides the process for watermarking the relational database using numeric and non-numeric attributes of database aiming to recover the watermarked data successfully.

Initially the database having textual data values are converted to numerical values and weight of each attribute is calculated. This weight is used for calculating the Mutual information (MI) value amongst the features. A suitable non-numeric data feature is selected for watermark embedding.

6. A Survey on Reversible Watermarking Techniques for Relational Databases [6]:

Most of Reversible watermarking technique applies on numeric data only, Non numerical database watermarking is a big research area for recent years.

Reversible Watermarking technique uses appropriate embedding mechanism to embed some information that can help to recover original data along with embedded watermark.

In distortion free techniques watermark is embedded in permuted or hash values of original data, therefore original data remains intact.

In reversible watermarking technique, a balance should be considered between the amount of secret information embedded into the original data and the distortion introduced in the underlying data.

7. A Robust & Reversible Watermarking Techniques for Relational database [7]:

Paper defines Reversible watermarking as Invertible Watermarking or Erasable Watermarking which Enables one to generate the first information after the substance have been verified.

Proposed approach [7] guarantees 100% regeneration of first database content after watermark has been recognized & confirmed.

8. An Advanced Watermarking Technique on Relational Database [8]:

This paper evaluates RRW [3] technique through attack analysis & prove that watermark is detected with maximum decoding accuracy in different conditions. This Paper states that even if an attacker attacks the database &

 DOI : 10.18231/2454-9150.2018.0516

alter the content of database, RRW [3] is able to recover both embedded watermark & original data. RRW is proved to be best among DEW & GADEW techniques.

Future concern is to watermark shared database in distributed environments where different members are connected through distributed systems and they Share their data in different proportions.

9. **Architecture for Reversible Watermarking For Distributed Database** [9]:

Paper provides watermarking technique for shared database in distributed environment. Proposed system [9] consists of four servers. These Servers are connected to each other via LAN and connected to respective client directly. Database design starts from global schema and Processed by designing fragmentation and Allocating fragments to different servers. Data is distributed by fragmenting the data and storing at different servers. Here horizontal fragmentation is used, each replica is located at minimum 2 servers. Each server add watermark to

Numerical data in relational database. Only authorized client can access watermarked data. Here read authorization is provided to client and log is maintained for read only access.

When authorized client send request to server, server checks validity of client. For this, server maintains client table and checks user identification number and password match. If match found then client is valid otherwise refused to access the database.

10. An Efficient Reversible Watermarking Technique for Textual Data [10]:

Reversible Watermarking on Textual data is performed by using UTF-8 & ASCII value Select the textual field from the database, each character from the word is identified separately based on the retrieved character the UTF-8 corresponding to the word is then generated.

The ASCII value of the alphabets for both the upper and lower case is stored in a separate matrix. Data owner can specify an optimized value represented by β. To encode the optimal value the first alphabet of the word is selected, Based on the alphabet the corresponding ASCII value is then changed using the UTF-8 and the optimal value.

.

If UTF-8 bit code of character is 0 then
Encoded bit of character =
ASCII value of character + optimal value (β)
Else
Encoded bit of character =
ASCII value of character − Optimal value (β)
This process is repeated until all the 8 bit string of the UTF-8 is fully traversed. The final value thus obtained is then stored to a separate matrix that is used for the decoding purpose

III. Comparison of Reversible Watermarking Techniques

Table 1.3 Shows the Comparison of three Reversible Watermarking Techniques as DEW [1]

,RRW [3] , & RWTD [10] based on certain parameters as Relational data Format used, feature value selection procedure used , watermark generation & key generation process for identifying the watermark & level of watermark applied , amount of distortion allowed & level of security achieved through this technique.

DEW & RRW works for relational database having only numerical values, whereas RWTD works for relational database having numerical as well as non-numerical databases. Tuple selection for embedding watermark value is decided by Difference Expansion between two numerical attributes in DEW, Mutual information ratio is calculated for each numerical attribute in RRW & in RWTD any attribute can be selected depending on owner for embedding watermark. Watermark is generated based on average & difference expansion of two selected features in DEW. Genetic Algorithm is used in case of RRW. Function of ASCII value & optimum value is selected for RWTD. Key value for identifying watermark is generated by using hash of secret key & primary key in DEW. Based on genetic algorithm defined objective value i.e. β in RRW & owner defined value of β in RWTD. Watermark embedding capacity is less in DEW than in RRW & RWTD where as it is high using RRW & optimum using RWTD. Amount of distortion allowed while embedding watermark is less in DEW as compare to RRW & RWTD and is high in RRW & Average in RWTD. After embedding watermark DEW is still vulnerable to attacks & data loss where as RRW & RWTD are robust against attacks.

 DOI : 10.18231/2454-9150.2018.0516

Table 1.3 Comparison of DEW, RRW & RWTD

Proposed Technique	Data Format	Tuple/ Feature selection	Watermark Generation	Key Generation	Watermark Capacity	Distortion Tolerance Mechanism	Security
DEW [1]	Numeric	2 features are selected based on Difference Expansion	Based on average & difference of 2 features	Based on hash of secret key & primary key	Less	Tolerable up to LSB	Vulnerable to attack
RRW [3]	Numeric	Based on Mutual Information (MI)	Based on Genetic Algorithm	Based on GA defined Objective i.e. β	High	GA based optimum value is Embedded (more Tolerable than DEW)	Robust against attack
RWTD [10]	Textual	Depending on Owner	Function of ASCII value & Optimum value	Owner defined value of β	Optimum	Optimum	Robust against attack

SUMMARY

Paper [1] describes Difference Expansion based Watermarking technique used for reversible watermarking. Paper [2] describes Genetic Algorithm based Difference Expansion based watermarking which is more secure than DEW and also increases watermarking capacity of database. Paper [3] describes RRW technique which uses Genetic Algorithm and Calculate Mutual Information for each attribute which is used to identify which attribute is used to embed the watermark. Paper [4] describes enhanced RRW technique which is used to watermarked database having textual attributes, and also uses supervised learning dataset which minimizes the calculation of Mutual information of attribute which needs to be calculated only for class feature. Paper [5] defines the new technique which is used for watermarking non-numerical data as by converting it to numerical values and then weight for each attribute is considered for calculating Mutual Information (MI) amongst the other features and whose MI value is less than threshold value is selected for Watermark embedding. Paper [6] discuss about reversible watermarking applications, requirements, generalized structure of reversible watermarking, & classification of reversible watermarking techniques. Paper [7] analyze Robust and reversible watermarking technique and guarantees that original data can be recovered in presence of active attack. Paper [8] also analyze RRW [3] against different available reversible watermarking techniques and prove that RRW is best among all. Paper [9] describes reversible watermarking technique for shared database in distributed environment where different users share their data. Paper [10] introduces new reversible watermarking technique which is used by databases having textual data to watermark textual field.

V. CONCLUSION

It is important to preserve data quality of the relational data since they are highly used for research and application purpose. The purpose of this paper is to study available literature till date regarding reversible watermarking technique & find the area where this technique can be improved so that the level of security provided by reversible watermarking technique can be increased. The different techniques reviewed in this paper are some of the Reversible Watermarking technique used for relational database. These method can be used in many real world applications which shares its database for business purpose & these techniques are used to provide security to those databases, but there are lack of reversible watermarking techniques which can be applied on relational database having non-numerical values & such database can become target of an attacker. So in the future we need to direct our research to find reversible watermarking technique for non-numerical data which can be robust against any database attacks.

ACKNOWLEDGMENT

With the completion of this paper, I would like to take this opportunity to express my gratitude and deep regards to thank my guide Dr.Vinayak D.Shinde, H.O.D Computer Engineering Department, for his constant support and valuable guidance. I also express my sincere appreciation to all the Professors of Computer Engineering Department as well as lab assistants for their support. I am also thankful to my parents, my husband Mr. Suraj K. Raut and my well-

 DOI : 10.18231/2454-9150.2018.0516

wishers for their support & continuous upliftment at each and every step in fulfillment of my work.

REFERENCES

[1] G. Gupta and J. Pieperzyk, "Reversible And Blind Database Watermarking Using Difference Expansion Based Watermarking," *Institute For Computer* Sciences ,*Social Informatics & Telecommunication Engineering*, 2008.

[2] J. Khurram and A. Khan, "Genetic algorithm and difference expansion based reversiblewatermarking for relational databases," *Elsevier*, 2013.

[3] S. Iftikhar, M. Kamran and Z. Anwar, "RRW - A Robust and Reversible Watermarking," *IEEE*, vol. 27, no. 4, 2015.

[4] M. V. Gaikwad and R. A. Kudale, "Enhanced Robust and Reversible Watermarking," *International Journal of Science and Research (IJSR)*, vol. 5, no. 11, 2016.

[5] G. R. Ghogare and A. Junnarkar, "Genetic Algorithm Based Reversible Watermarking Approach for Numeric and Non-Numeric Relational Data," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 7, pp. 1817-1822, 2007.

[6] S. Iftikhar, K. M and Z. Anwar, "A survey on reversible watermarking techniques for Relational Database," *Wiley*, 2015.

[7] H. R and V. S, "A Robust and Reversible Watermarking Technique for Relational Data," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 5, 2017.

[8] A. Khan, D. Narodiya and M. Korde, "An Advanced Watermark Technique on Relational Database," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 6, no. 9, 2017.

[9] S. R. Hatture and S. Raut, "Architecture for Recoverable Watermarking for Distributed Databases," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 6, 2016.

[10] A. N. Vaidyanathan , D. S. Subasree and M. P. B. , "An Efficient Reversible Watermarking Technique for Textual Data," *International Journal of Innovative Research in Science & Technology*, vol. 3, no. 2, 2016.

 DOI : 10.18231/2454-9150.2018.0516