Reversible Watermarking Technique for Relational Data using Ant Colony Optimization and Encryption

Sharafunisa S, Smitha E S

Abstract— Data is stored in different digital formats such as images, audio, video, natural language texts and relational data. Relational data in particular is shared extensively by the owners with communities for research purpose and in virtual storage locations in the cloud. The purpose is to work in a collaborative environment where data is openly available for decision making and knowledge extraction process. So there is a need to protect these data from various threats like ownership claiming, piracy, theft, etc. Watermarking is a solution to overcome these issues. Watermark is considered to be some kind of information that is embedded into the underlying data. While embedding the watermark, the data may modify, to overcome this we use reversible watermarking in which owner can recover the data after watermarking. In this paper, a reversible watermarking for relational data has been proposed that uses ant colony optimization and encryption for more accuracy and security.

Index Terms— Ant colony optimization (ACO), Mutual information (MI), Reversible watermarking, Data recovery, Genetic Algorithm (GA).

I. INTRODUCTION

Watermark is some kind of information that is embedded into the underlying data for ownership proof .Watermarking technique[1] are used to ensure security in terms of ownership protection and tamper proofing for different data formats. Initially the watermarking techniques are used for multimedia data like image, audio, video [2] etc., but recently the watermarking techniques for relational data received attention because of its importance in real life applications. Database watermarking can be used to enforce ownership rights of relational data. But watermarking modifies the underlying data while embedding watermark. To overcome this problem we use reversible watermarking [3], that recover the original data from watermarked relation.

In this paper a reversible watermarking technique for numerical relational database has been proposed that uses ant colony optimization [4] for watermark creation. In the early techniques for watermarking relational data does not consider about the importance of attributes in knowledge extraction process. But in this technique a Mutual Information [5] statistics is used to select suitable feature for watermark embedding. So the data will be useful after embedding the watermark information. Like other reversible watermarking techniques the proposed technique also comprises of following phases, (i) Data pre- processing (ii) Watermark encoding (iii) watermark decoding (iv)Data recovery. In addition to this an encryption technique is also added to improve the security of the technique.

Manuscript published on 30 August 2016.
***** Correspondence Author (s)

Sharafunisa S, M.Tech Scholar, Department of Computer Science and Engineering, LBS Institute of Technology for Women, Thiruvananthapuram, India.

Smitha E S, Associate Professor, Department of Computer Science and Engineering, LBS Institute of Technology for Women, Thiruvananthapuram, India.

II. RELATED WORKS

Agarwal and Kiernan[6] first irreversible watermarking technique for relational data has been proposed. It works only on numerical relational data. The technique is based on message authentication code (MAC) and primary key attribute of the relation. This watermarking technique ensures that some bit positions of some of the attributes of some of the tuples contain specific value. The specific bit locations and values are algorithmically determined under the control of secret key known only to the owner of the data. This bit pattern constitutes the watermark. The problem with this technique is it is irreversible, that is we cannot get the original data from the watermarked relation.

To overcome the problems with irreversible watermarking technique, reversible watermarking techniques for relational data has been developed. Difference expansion watermarking techniques (DEW) [7] exploits methods of arithmetic operations on numeric operations and performs transformations. Gupta and Pieprzky, [8] propose a reversible watermarking scheme which is the modified version of Agarwal and Kiernan's [6] one. In this scheme, during the detection phase, the original un-watermarked version of the database can be recovered along with the ownership proof. The operation first extracts a bit *OldBit* from the integer portion of the attribute value before replacing it by the watermark bit and inserts it in the fraction portion of the attribute value. Thus, the watermark bit can be recovered during detection and the attribute can be restored to its unmarked value by replacing the watermark bit with the original bit *OldBit* extracted from the fraction part.

Genetic algorithm based difference expansion technique [9] uses Genetic algorithm to choose suitable attributes for embedding watermark. Using DEW technique the watermark bits selected by MAC are embedded into selected features. This technique tries to minimize the distortions in the data.

Farfoura and Horng proposed prediction error expansion watermarking technique [10], in which prediction error expansion is used instead of difference expansion. Predictor is used to select the candidate features for embedding watermark.

This method is fragile against attack, because in this the watermark is embedded into the fractional part of numerical data.

In RRW [11], proposes a reversible and robust watermarking scheme for relational data. It works on numeric attributes. In this technique the features are selected using mutual information. That is the features are selected according to their importance in knowledge discovery process. The watermarking bits are generated using genetic algorithm. Using this evolutionary algorithm the chromosome having best fitness value is selected. Based on the value of fitness function, the bits are embedded in to the selected features.RRW provides a robust solution for data recovery that is reversible and resilient against heavy attacks. The proposed method is similar to existing RRW; watermark creation phase is modified to improve the performance and robustness of the watermarking technique.

III. PROPOSED SYSTEM

The reversible watermarking technique for relational data using ant colony optimization and encryption includes the following four major phases: (1) Watermark pre-processing (2) Watermark encoding (3) Watermark decoding and (4) Data Recovery. Fig. 1 shows the architecture of proposed technique.

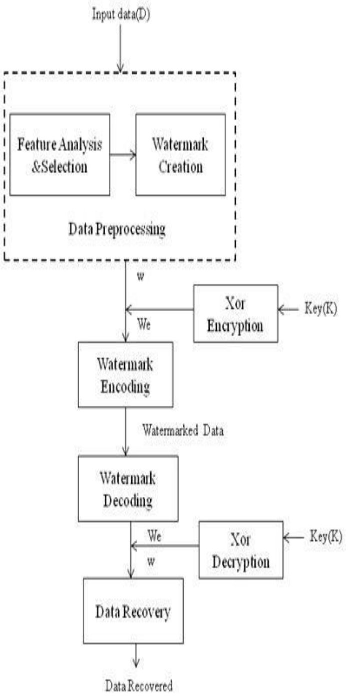


Fig.1: Proposed System Architecture

A. *Watermark Pre-Processing phase*

In this phase two tasks are accomplished (i) Feature analysis and selection (ii) Watermark information creation [11].

During feature analysis and selection, the owner has to select a suitable feature for watermark embedding. While embedding watermark the usability of data may decrease. We need a watermarking technique in which the features for embedding watermark have to be chosen based on their importance in knowledge extraction process. Like RRW, in proposed method also mutual information (MI) [5] is used for selecting features. Mutual Information is a well known information theory concept, statistically measures the amount of information that one feature contains about other feature in the database. Mutual information of every feature with all other feature is calculated by using equation (1).

$$MI(A,B)=\sum_a\sum_b P_{AB}(a,b)\log(\frac{p(x,y)}{P(x)P(y)} \qquad (1)$$

The value of MI of each feature is used to rank the features. The owner can define a secret threshold based on the MI of all the features in the database. The features having MI lower than that threshold can be selected for watermarking.

For the creation of optimal watermark information, that needs to be embedded in the original data, an evolutionary technique is used. In this method, ant colony optimization (ACO) [4] is used for the creation of optimal watermark information. ACO is a probabilistic technique for solving computational problems which can be reduced to finding good path through graphs. This algorithm is a member of the ant colony algorithms family, in swarm intelligence methods, and it constitute some meta-heuristics optimizations. In the natural world, ants wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep travelling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food. Ant colony optimization is used to create the optimal watermark information and the best fitness value β. The optimal watermark information is created using the steps: initialization of different variables and population, creating random solutions, evaluate the solutions based on the fitness function, and selection of best solution. Figure 2 shows the implementation steps of ACO for the proposed technique. The watermark information is created using some constraints such as the difference between MI, mean, standard deviation of original and watermarked data should be minimum.

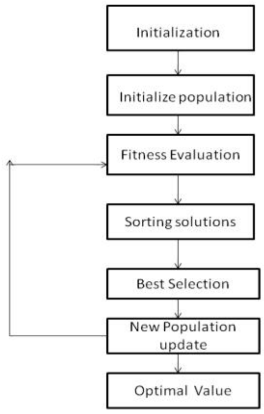


Fig 2: ACO implementation steps.

XOR encryption is used in the proposed reversible watermarking technique for encrypting the watermarking information created using ant colony optimization. Encryption helps to improve the robustness of the watermarking technique.

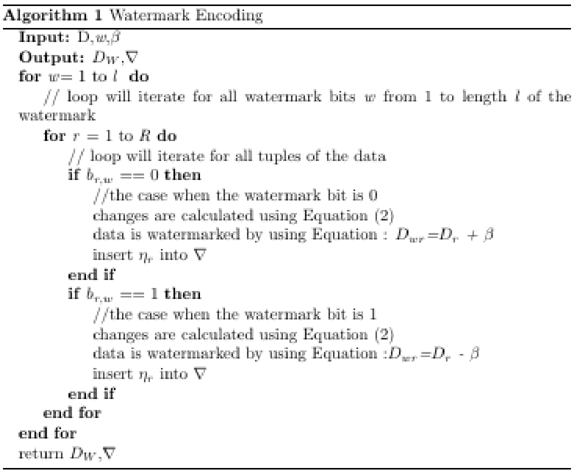
In this technique, a secret key is selected by the owner that has to be XORed with the watermark information generated using ant colony optimization algorithm. The secret key and watermark information should have same number of bits. In the watermark decoding phase, to decrypt the watermark information the owner has to enter the same key that is used for encryption.

B. Watermark Encoding Phase

In the watermark encoding phase the optimal watermark information calculated using ACO and encryption is embedded into the features selected using MI. The watermark encoding phase of proposed technique is similar to the encoding phase of RRW [11]. The optimum value β is added into every tuples of the selected feature when the watermark bit is 0; otherwise, its value is subtracted from the value of the feature. The watermark is inserted into every tuple of the selected feature of the dataset. After finding the optimum value of β a parameter η_r is calculated according to the equation (2) that represents the percentage change in the watermark encoding. This parameter is calculated for a tuple r as:

$$\eta_r = D_r * \zeta \qquad (2)$$

Figure: 3 show the algorithm for watermark encoding. The algorithm iterates for all watermark bits and tuples of the selected feature.



```
Algorithm 1 Watermark Encoding
Input: D,$w$,$\beta$
Output: $D_W$,$\nabla$
for $w$= 1 to $l$  do
    // loop will iterate for all watermark bits $w$ from 1 to length $l$ of the
watermark
    for $r$ = 1 to $R$ do
        // loop will iterate for all tuples of the data
        if $b_{r,w}$ == 0 then
            //the case when the watermark bit is 0
            changes are calculated using Equation (2)
            data is watermarked by using Equation : $D_{wr}$=$D_r$ + $\beta$
            insert $\eta_r$ into $\nabla$
        end if
        if $b_{r,w}$ == 1 then
            //the case when the watermark bit is 1
            changes are calculated using Equation (2)
            data is watermarked by using Equation :$D_{wr}$=$D_r$ - $\beta$
            insert $\eta_r$ into $\nabla$
        end if
    end for
end for
return $D_W$,$\nabla$
```

Fig 3: Watermark Encoding Algorithm

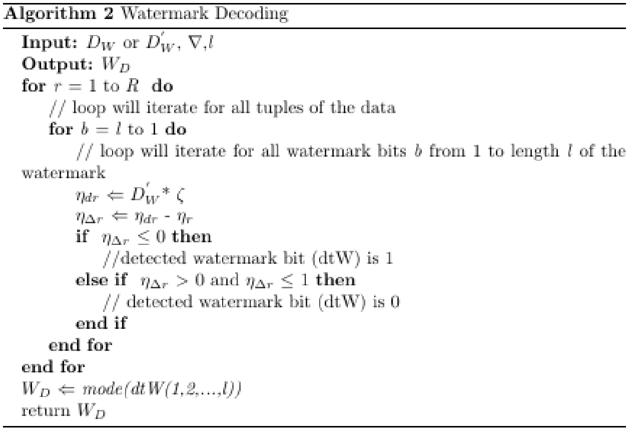
C. Watermark Decoding Phase

In the watermark decoding process, the first step is to locate the features which have been marked. We use a watermark decoder ζ, which calculates the amount of change in the value of a feature that does not affect its data quality. The watermark decoder decodes the watermark by working with one bit at a time. The decoding phase is also similar to RRW [11]. But the encrypted watermark bits are decoded from the watermarked data. In the decoding phase, η_{dr} is calculated using equation (3) and represents the percent change detected in the watermarked data. The value of η_{dr}, η_r and $\eta_{\Delta r}$ is calculated using the values of tuple r and therefore might be different for every r. The parameter $\eta_{\Delta r}$ is computed by calculating the difference between the original data change amount η_r and the watermark detected change amount η_{dr} using Equation (4).

$$\eta_{dr} = D_W^{'} * \zeta \qquad (3)$$

$$\eta_{\Delta r} = \eta_{dr} - \eta_r \qquad (4)$$

The watermark decoding algorithm is described in figure4.



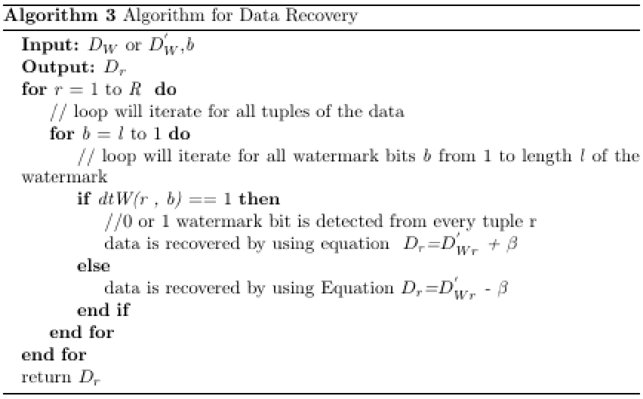
```
Algorithm 2 Watermark Decoding
Input: $D_W$ or $D_W^{'}$, $\nabla$,$l$
Output: $W_D$
for $r$ = 1 to $R$  do
    // loop will iterate for all tuples of the data
    for $b$ = $l$ to 1 do
        // loop will iterate for all watermark bits $b$ from 1 to length $l$ of the
watermark
        $\eta_{dr} \Leftarrow D_W^{'} * \zeta$
        $\eta_{\Delta r} \Leftarrow \eta_{dr}$ - $\eta_r$
        if  $\eta_{\Delta r} \leq 0$ then
            //detected watermark bit (dtW) is 1
        else if  $\eta_{\Delta r} > 0$ and $\eta_{\Delta r} \leq 1$ then
            // detected watermark bit (dtW) is 0
        end if
    end for
end for
$W_D \Leftarrow mode(dtW(1,2,...,l))$
return $W_D$
```

Fig 4: Watermark Decoding Algorithm

Using the watermark decoding algorithm the encrypted watermark bits are extracted. Before performing data recovery we have to calculate the original watermark information using XOR decryption. The decoded bits and secret key entered by the owner is XORed to get the original watermark bits.

D. Data Recovery Phase

During data recovery phase, the original data values of features are recovered from the watermarked relation. The optimized value of β computed through the ACO is used for regeneration of original data .The data recovery algorithm is presented in figure 5 [11] . If the watermark bit is 1, the β value is added to the watermarked data to get the original data. Otherwise the β value is subtracted from the watermarked data to get original information.



```
Algorithm 3 Algorithm for Data Recovery
Input: $D_W$ or $D_W^{'}$,$b$
Output: $D_r$
for $r$ = 1 to $R$  do
    // loop will iterate for all tuples of the data
    for $b$ = $l$ to 1 do
        // loop will iterate for all watermark bits $b$ from 1 to length $l$ of the
watermark
        if $dtW(r , b)$ == 1 then
            //0 or 1 watermark bit is detected from every tuple r
            data is recovered by using equation  $D_r$=$D_{Wr}^{'}$ + $\beta$
        else
            data is recovered by using Equation $D_r$=$D_{Wr}^{'}$ - $\beta$
        end if
    end for
end for
return $D_r$
```

Fig 5: Algorithm for data recovery After executing the data recovery algorithm we get the original data values in the relational data.

IV. EXPERIMENTAL EVALUATION

Matlab is used to implement the proposed watermarking technique. The data set used is Cleveland heart disease data set [12].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This is a numerical data set containing more than 300 tuples. A system with Intel core i3 and 32-bit operating system is used for the implementation purpose. The data set is given as input and pre-processed the dataset to get it in required format. Mutual information is used to select the candidate features.ACO is used to create the optimal watermark information and best fitness value. For reducing the data distortion the length of watermark bits is taken as 3. The change incorporated into the tuple value, ζ is set as 10\%. Before watermark encoding, the watermark bits are encrypted using XOR encryption. The owner can input encryption key, using this key XOR encryption is performed and the encrypted bits are encoded into the selected features. During watermark decoding phase first we decode the encrypted watermark information. For getting the original watermark we have to input the same key that is used for encryption. Using the watermark bits we can recover the original data in the data recovery phase.

For evaluating the performance of the proposed technique, it is compared with existing reversible watermarking technique, RRW. Test cases are selected for comparison purposes. The test case includes small number of tuples selected from the data set. For comparing the genetic algorithm used in existing technique is compared with ant colony optimization in proposed technique. The time consumption and error rate during data recovery phase is used for performance evaluation purpose. The graph shown in figure 6 shows the comparison between the execution time of GA and ACO. The result shows that ACO has faster convergence time compared to GA [11].

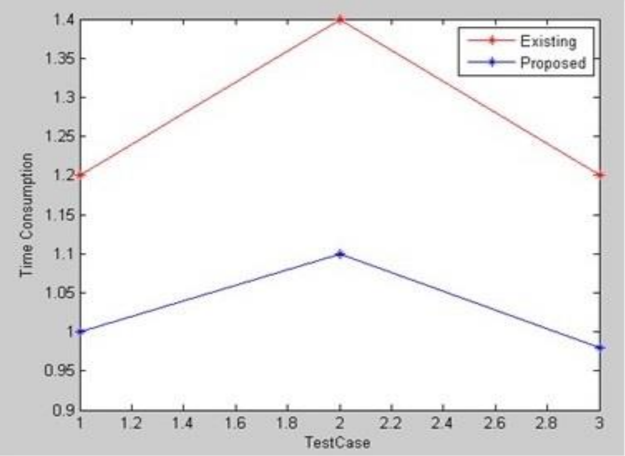


Fig.6: Time consumption (GA Vs ACO)

The graph in figure 7 shows the comparison of existing and proposed technique based on the error rate during data recovery phase. The result shows that the proposed technique recovers the data more accurately.

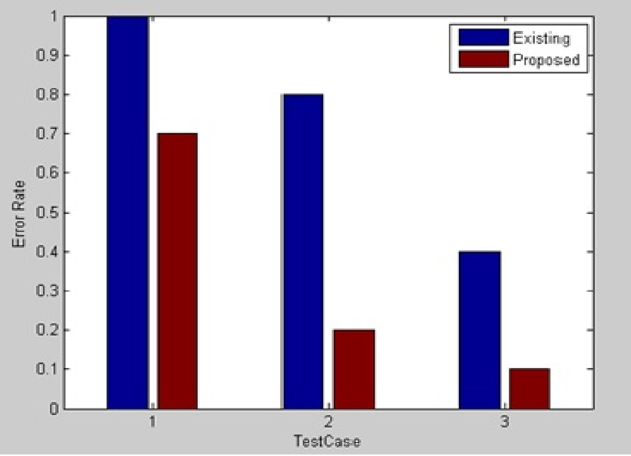


Fig.7: Error in data recovery

V. Conclusion

Watermarking is advocated to enforce ownership rights over shared relational data for providing means for tackling data tampering. Irreversible watermarking techniques, make changes in the data to such an extend data quality get compromised. But, reversible watermarking techniques are able to recover the original data from the watermarked data. In this work, a reversible watermarking technique for numerical relational data has been proposed that ensures data quality along with data recovery. The encryption technique used in this technique also improves the security of the data. The experimental results show that it has faster execution time and lower error rate during data recovery compared to existing technique.

Acknowledgment

We are greatly indebted to our principal, Dr. Jayamohan J, Dr. V. Gopakumar, Professor, Head of the Department of Computer Science and Engineering, Mrs. Smitha E.S, Associate Professor, Department of Computer Science and Engineering, LBS Institute of Technology for Women who have been instrumental in keeping my confidence level high and for being supportive in the successful completion of this paper. We would also extend our gratefulness to all the staff members in the Department; also thank all my friends and well-wishers who greatly helped me in my endeavor. Above all, we thank the Almighty God for the support, guidance and blessings bestowed on us, which made it a success.

References

1. Raju Halder, Shanthanu Pal and Agostino Cortesi ,“Watermarking Techniques for Relational Databases: Survey, Classification and Comparison,” Journal of Universal Computer Science, Vol 16 ,2010, Number 21, pp.3164-3190
2. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia”, IEEE Trans. Image Process., vol. 6, no. 12, pp. 16731687, Dec. 1997
3. Ifthikar, M. Kamran and Z. Anwar, “A Survey on Reversible Watermarking Techniques for Relational Databases,” Security and communication networks, 2015.
4. Marco Dorigo and Thomus Stultze, ”Ant Colony Optimization“, 2004.

5. T. M. Cover, J. A. Thomas, and J. Kieffer,'Elements of information theory," SIAM Rev., vol. 36, no. 3, pp. 509510, 1994.
6. R. Agarwal and J. Kiernan, "Watermarking relational databases", in Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 155166.
7. G. Gupta and J. Pieprzyk, "Reversible and blind database watermarking using difference expansion," in Proc. 1st Int. Conf. Forensic Appl. Tech. Telecommun., Inf., Multimedia Workshop, 2008, p. 24.
8. G. Gupta and J. Pieprzyk, "Database relation watermarking resilient against secondary watermarking attacks," in Information Systems and Security. New York, NY, USA: Springer, 2009, pp. 222–236.
9. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," J. Syst. Softw., vol. 86, no. 11, pp. 2742–2753, 2013.
10. M. E. Farfoura and S.-J. Horng, "A novel blind reversible method for watermarking relational databases," in Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl., 2010, pp. 563–569
11. Iftikhar S, Kamran M, Anwar Z.,“ RRW-a robust and reversible watermarking technique for relational data, IEEE transactions on Knowledge and Data Engineering , 2015, Volume: 27,Issue: 4, pp: 1132 – 1145
12. K. Huang, H. Yang, I. King, M. R. Lyu, and L. Chan,"Biased minimax probability machine for medical diagnosis“, AMAI, 2004.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)