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 10, October 2016

A Secure Watermarking Technique for Numeric and Non-Numeric Relational Data

Uzma A. Shaikh[1], Prof. Kishor N. Shedge[2]

M.E. Student, Dept. of Computer Engineering, SVIT, Chincholi, Nashik, Maharashtra, India[1]

Assistant Professor, Dept. of Computer Engineering, SVIT, Chincholi, Nashik, Maharashtra, India[2]

ABSTRACT: The rapid development of internet has led to the easy access and distribution of digital data. With the advancement of such technologies, protecting the ownership and controlling the copies of digital data is becoming very important. Digital Watermarking is used to impose ownership rights on the shared relational data and to provide a means for tackling the data tampering. In the existing watermarking techniques the data gets degraded while embedding the watermark. In this paper, a watermarking approach based on robust and reversible watermarking technique is used. This technique is used to watermark the numeric as well as the non-numeric features present in the relational database.

KEYWORDS: Digital Data, Robust, Reversible Watermarking.

I. INTRODUCTION

Due to the rapid growth of the Internet, the wide development of digital multimedia contents, and the easier distribution, copyright protection of owners is becoming more important. Digital watermarking is an important area in information hiding, thus providing a promising method of protecting digital data from illegitimate copying, and manipulation by embedding a secret code directly into the data. Digital watermarking allows the user to add a layer of protection to the digital media content by identifying copyright ownership and delivering a tracking capability. Accordingly, it monitors and reports where the user's digital media contents are being used.

Cryptography allows secure delivery of content to the consumers only. All legitimate consumers are not trustworthy, and untrustworthy consumers may modify or copy the decrypted contents illegally. However, cryptography provides no protection once the content is decrypted, which is required for human perception. Watermarking complements cryptography by embedding a message within the content.

Digital watermarking was exploited in other digital media like protecting software, natural language and sensor data. The basic watermarking procedures like watermark insertion and detection to multimedia objects cannot be applied directly to watermarking relational databases due to the differences in the characteristics of multimedia and relational data. Unlike highly correlated multimedia data whose relative positions are fixed, database relations contain tuples with little redundancy. The tuples can be added, deleted or modified frequently in either benign updates or malicious attacks.

In the conventional irreversible watermarking schemes, only the embedded watermark information can be extracted from the suspected data; however, in reversible watermarking schemes, the original objects can be recovered along with the embedded watermarks. Most watermarking schemes have been irreversible (the original relation cannot be restored from the watermarked relation). One major motivation for reversible watermarking is some real-life applications such as outsourced medical and military data. These kinds of data do not allow any losses.

In this, a robust and reversible watermarking technique is proposed. This method is used to watermark numeric as well as the non-numeric data in the relational database. This method tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information.

II. RELATED WORK

R. Sion, M. Atallah, and S. Prabhakar [2] proposed an encoding method for rights protection for categorical data.In this technique, if watermarks are altered then the percentage of data loss increases.

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 10, October 2016

Diljith M. Thodi and Jeffrey J. Rodriguez [3] proposed a prediction-error expansion and histogram shifting technique as an alternative to embedding the location map. The proposed technique improves the distortion performance at low embedding capacities and mitigates the capacity control problem.

Agrawal and Kiernan [4] proposed the first irreversible watermarking technique for relational databases. They presented an effective watermarking technique geared for relational data. This technique uses the pseudo-random distribution of watermark based on keyed-hash function. This technique ensures that some bit positions of some of the attributes of some of the tuples contain specific values. The scheme does not provide security against secondary watermark attacks.

M. E. Farfoura and S.-J. Horng [5] proposed a reversible data embedding technique called Prediction-error Expansion (PE) on integers to achieve reversibility. This technique is fragile against malicious attacks as the as the watermark information is embedded in the fractional part of numeric features only.

A. M. Alattar [6] proposed watermarking algorithm based on difference expansion of colored images. The algorithm uses spatial and spectral triplets of pixels to hide pairs of bits, which allows algorithm to hide large amount of data.

G. Gupta and J. Pieprzyk [7] proposed an improvement over the reversible and blind watermark scheme proposed by Gupta, G., Pieprzyk, J. in Reversible and blind database watermarking using difference expansion.

. Zhang, B. Yang, and X.-M. Niu [8] proposed the first reversible watermarking scheme for relational databases. In this technique, histogram expansion is used for reversible watermarking of relational database. Zhang et al. proposed a method of distribution of error between two evenly distributed variables and selected some initial non-zero digits of errors to form histograms. Histogram expansion technique is used to reversibly watermark the selected nonzero initial digits of errors. This technique keeps track of overhead information to authenticate data quality. However, this technique is not robust against heavy attacks.

K. Jawad and A. Khan [9] proposed a Difference Expansion Watermarking based on Genetic Algorithm (GADEW), a reversible solut ion for relational database. GADEW improves upon the drawbacks mentioned by G. Gupta and J. Pieprzyk, in Reversible and blind database watermarking using difference expansion by minimizing distortions in the data, increasing watermark capacity and lowering false positive rate.

III. PROPOSED SYSTEM

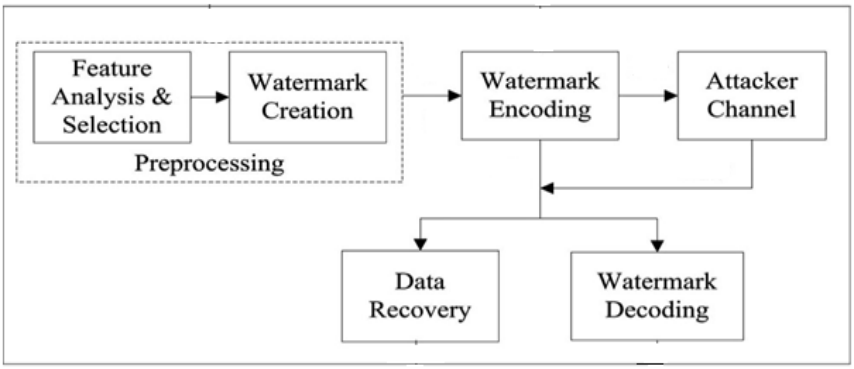

Figure 1: System Architecture

There are four major phases :

***A. WATERMARK PREPROCESSING*:** In the watermark preprocessing phase, two tasks are accomplished. In this firstly, the suitable features for watermark embedding are selected and analysed. After feature selection, the watermark is created. For selecting the features, the mutual information of each feature with the other features is calculated. The mutual information is calculated using the formula given below:

$$MI(A,B) = \sum_a \sum_b P_{AB}(a,b) \log P_{AB}(a,b) / P_A(a)P_B(b)$$

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 10, October 2016

After the calculation of mutual information, the secret threshold is defined. The features having mutual information less than the defined threshold are selected for watermarking. After the selection of features, optimal watermark string is created using genetic algorithm. Also the optimal fitness value (β) is obtained using the genetic algorithm.

***B. WATERMARK ENCODING* :** The main focus of watermark encoding phase is to embed watermark information in such a way that it does not affect the data quality. Watermark embedding phase embeds the watermark in accordance with the encoding strategies and data usability. During this phase, firstly ,the changes while embedding the watermark into the data are calculated using equation(1). For the non-numeric data, the ASCII value of each character of a string is generated. After the calculation of amount of change while watermark encoding, the data is watermarked. If the MSB of watermark string is 1, then data is watermarked using the equation (2). If the MSB is 0,then data is watermarked using the equation (3)

$$\eta_r = D_r * \zeta \text{ ---------- (1)}$$

ζ is the watermark decoder.

$$D_{Wr} = D_r - \beta \text{ ----------(2)}$$
$$D_{Wr} = D_r + \beta \text{----------(3)}$$

After watermarking, the data is released to the attacker channel. The attacker channel refers to all such possible attacks. The attacks can modify some contents of the data with an aim to corrupt the embedded watermark; making the data suspicious.

C. WATERMARK DECODING: The Watermark decoding phase recovers watermark information effectively for detection of the embedded watermark. In this phase the percent change in the watermarked data is calculated using equation (4). After this, the difference between the original data change and the watermark detected change amount is calculated using equation (5). Final watermark information is retrieved through a majority voting scheme using Equation (6).

$$\eta_{dr} = D'_W * \zeta \text{---------- (4)}$$

$$\eta_{\Delta r} = \eta_{dr} - \eta_r \text{---------- (5)}$$

$$WD \leftarrow \text{mode}(dtW(1,2,\ldots..,l)) \text{---------- (6)}$$

***D. DATA RECOVERY*:** After detecting the watermark string, some post processing steps are carried out for error correction and data recovery. The main responsibility of post-processing is to use the decoded watermark bits, and convert these bits into the watermark information that was embedded as the watermark. If the detected watermark bit is 0, the data is recovered using equation (7). If detected watermark bit is 1, then data is recovered using equation (8).

$$D_r = D'_{Wr} - \beta \text{ ----------(7)}$$
$$D_r = D'_{Wr} + \beta \text{----------(8)}$$

IV. ALGORITHM

A. WATERMARK ENCODING ALGORITHM:

Input: Database, watermark string, β
Output: Watermarked database, matrix

1. For all watermark bits 1 to length l
2. For all the tuples of the data
3. If watermark bit is 0 ,then compute changes using equation (1).
4. Watermark data using equation (3).

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 10, October 2016

5. insert η_r into the matrix.
6. End if.
7. If watermark bit is 1, then compute changes using equation (1).
8. Watermark data using equation (2).
9. insert η_r into the matrix.
10. End if
11. End For
12. End For

B. WATERMARK DECODING ALGORITHM:

Input: Watermarked Database, matrix containing change in the data values, length of watermark string
Output: Decoded Watermark

1. For all tuples of the watermarked data
2. For all watermark bits b from 1 to length l of the watermark
3. Compute percent change in the watermarked data using equation(4)
4. Compute the difference between original data change amount and the watermark detected change amount using equation (5)
5. If difference computed using equation(4) ≤ 0
6. Detected Watermark bit is 1
7. Else if difference is > 0 and ≤ 1
8. Detected Watermark bit is 0
9. End If
10. End For
11. End For
12. Compute final watermark string using equation (6)

V. RESULTS

To evaluate the performance of the proposed methodology, this methodology is implemented on Visual Studio 2010 C# using the .Net 4 framework. The Adult Dataset is considered for the result analysis. The results of watermarking are shown on the two features of the dataset. The Mutual Information of features before and after watermarking is given in the table 2. An Optimum Value of β =0.401 is obtained using genetic algorithm. The watermark string 100 is used for watermarking. The Watermarked Data is shown in table 3.

In the table below the original data with two features and four tuples is shown. One feature contains numeric data and the other feature contains non-numeric data. The proposed system is used to watermark both numeric and non-numeric data.

Table 1: Original Data

Sr. No.	hours_per_week	Workclass
1.	40	State-gov
2.	30	Self-emp-not-inc
3.	16	Private
4.	45	Private

The mutual information of all the features is calculated. The mutual information of features before watermarking and after watermarking is shown in table below. The MI_o represents the mutual information of the original data. MI_W represents the mutual information of the watermarked data. ΔMI represents the change in the value of mutual information. It is calculated by taking the difference between MI_W and MI_0.

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 10, October 2016

Table 2: Mutual Information of Features

Sr. No.	Features	MI_0	MI_W	ΔMI
1.	Age	0.8796	0.8796	0
2.	Education	0.4279	0.4279	0
3.	Education_num	0.4279	0.4279	0
4.	Martial status	0.1537	0.1537	0
5.	Occupation	0.4886	0.4886	0
6.	Relationship	0.2038	0.2038	0
7.	Race	0.1537	0.1537	0
8.	Gender	0.6447	0.6447	0
9.	Capital gain	0.1537	0.1537	0
10.	Capital loss	0.2767	0.2767	0
11.	Hours_per_week	0.3689	0.3689	0
11.	Native country	0.2566	0.2566	0
12.	Workclass	0	0	0

The table below shows the value of features after watermarking. The optimal fitness value 0.401 is used for watermarking. In the existing system the non-numeric features could not be watermarked. The result of watermarking numeric and non-numeric features is shown in the proposed system.

Table 3: Watermarked Data

A1	Existing System	Proposed System
40	40.401	40.401
30	30.401	30.401
16	16.401	16.401
45	45.401	45.401
State-gov	Not Applicable	Uvcvg/iqx
Self-emp-not-inc	Not Applicable	Ugnh/gor/pqv/kpe
Private	Not Applicable	Rtkxcvg
Private	Not Applicable	Rtkxcvg

VI. CONCLUSION AND FUTURE WORK

Nowadays the security of data is becoming significant especially in the collaborative environments. In irreversible watermarking techniques the original data could not be recovered thus degrading the quality of data.

In this a robust and distortion free watermarking technique has been proposed that is capable of recovering the original data. It allows recovery of large amount of the data and embedded watermark even after being subjected to malicious attacks. The existing watermarking technique is used to watermark only the numeric features of the database. But the proposed system allows watermarking of numeric as well as non-numeric features of the relational database. One of the future concerns is to regenerate the true data for all the tuples on deletion of data.

REFERENCES

[1] Saman Iftikhar, M. Kamran, and Zahid Anwar, RRW-A Robust and Reversible Watermarking Technique for Relational Data, IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 4, April 2015.
[2] R. Sion, M. Atallah, and S. Prabhakar, Rights protection for categorical data, IEEE Trans. Knowl. Data Eng., vol. 17, no.7, pp. 912926, Jul. 2005.
[3] D. M. Thodi and J. J. Rodriguez, Expansion embedding techniques for reversible watermarking, IEEE Trans. Image Process., vol. 16, no. 3, pp. 721730, Feb. 2007.
[4] R. Agrawal and J. Kiernan, Watermarking relational databases,in Proc.28th Int. Conf. Very Large Data Bases, 2002, pp. 155166.

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 10, October 2016

[5] 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. 563569.
[6] A. M. Alattar, Reversible watermark using difference expansion of Y. triplets, in Proc. IEEE Int. Conf. Image Process., 2003, pp. I501,vol.1.
[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] Zhang, B. Yang, and X.-M. Niu, Reversible watermarking for relational database authentication, J. Comput., vol. 17, no. 2,pp. 5966, 2006.
[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. 27422753, 2013.
[10] F. A. Petitcolas, Watermarking schemes evaluation, IEEE Signal Process. Mag., vol. 17, no. 5, pp. 5864, Sep. 2000.
[11] J. T. Brassil, S. Low, and N. F. Maxemchuk, Copyright protection for the electronic distribution of text documents, Proc. IEEE, vol. 87, no. 7, pp. 11811196, Jul. 1999.
[12] Y.-R. Wang, W.-H. Lin, and L. Yang, An intelligent watermarking method based on particle swarm optimization, Expert Syst. Appl., vol. 38, no. 7, pp. 80248029, 2011.
[13] 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. 222236.
[14] E. Sonnleitner, A robust watermarking approach for large databases,in Proc. IEEE First AESS Eur. Conf. Satellite Telecommun.,2012, pp. 16.

BIOGRAPHY

Uzma A. Shaikh received the B.E degree in Computer Engineering from SVIT, Nashik. Currently pursuing M.E degree in Computer Engineering at SVIT,Nashik.

Kishor N. Shedge received the B.E degree in Computer Science & Engineering from RGPV, Bhopal and received the M.Tech degree in Computer Science & Engineering from LNCT, Indore. He is currently working as a Asst.Professor at SVIT, Nashik in Computer Engineering Department.