A New Watermarking Approach for Non-numeric Relational Database

Prof. Rajneeshkaur Bedi
Department of Computer Engineering, MIT COE, Paud Road, Pune -38

Prof. Anita Thengade
Department of Computer Engineering, MIT COE, Paud Road, Pune -38

Dr. Vijay M.Wadhai
Department of E & TC, MIT Alandi, Pune-38

ABSTRACT
In this paper, a novel watermarking technique is proposed for data authentication and integrity of Relational Database. For integrity verification of tables in the database, the watermark has to depend on a secret key and on the original copy of that table. It is important that the dependence on the key should be sensitive. The proposed method makes use of the concept of eigen values by constructing a tuple -Relation matrix for each tuple. The eigen values are used for generating the watermark for a record in the table. Watermark embedding is done by using eigen values in a non numeric attribute of a tuple. Detection of the watermark prove the authenticate and integrity of data. We will show that our approach leads to an effective technique that is robust against different forms of malicious attacks as well as benign updates to the data.

General Terms
Security

Keywords
Watermarking, Relational Database, Eigen values;

1. INTRODUCTION
Digital images, video and audio are examples of digital assets which have become easily accessible by ordinary people around the world. However, the owners of such digital assets have long been concerned with the copyright of their digital products, since copying and distributing digital assets across the Internet was never easier and possible as its now a days. Digital watermarking technology was suggested lately as an effective solution for protecting the copyright of digital assets[2,3]. This technology provides ownership verification of a digital product by inserting imperceptive information into the digital product. Such 'right witness' information is called the watermark and it is inserted in such a way that the usefulness of the product remains, in addition to providing it with robustness against attempts to remove the watermark.

A watermarking of database systems started to receive attention because of the increasing use of database systems in many real-life applications. There is a need to preserve originality, ownership and integrity of database systems in a way that cannot be identified by everyone. It in turn arises the need of developing a well secured watermark technique to protect the database systems against the piracies.

Many researchers concentrated on watermarking numeric attribute relational database systems and also for non numeric attribute which stores image in it[1,2,6]. In this paper, we present an effective watermarking technique for non numeric relational data that is robust against various attacks. The main contributions of the paper are summarized as follows:

- We select non numeric low impact attributes which are used to watermark a relation.
- During watermark generation process, we design an algorithm for :
    1. Secret Key generation using Eigen values of tuple-Relation matrix for a tuple.
    2. Watermark insertion.
    3. Watermark detection is the reverse process of watermark insertion.

As any change to non-numeric attribute values of the tuple must be a change in vowel, consonant, numbers and special character of tuple-Relation matrix, we consider their individual counts to compute the weighted ASCII sum A for developing a new watermarking scheme. It provides a secret key to the authenticated user to verify the received database relation. If the database has undergone tampering during transmission, it can be recognized by the authenticated user.

2. RELATED WORK
Here we briefly discuss three of the previous approaches related to our work for watermarking relational databases. First, the method given in Agrawal et al[1] utilizes the pseudorandom number generator algorithm to identify the marked tuples and attributes, and also the degree of error to the marked attributes.

Second, is the approach proposed by Zhang et al using embedded images [4.] In other words, in their approach, they embed images into relational database as the watermarks.

While previous techniques have been mainly concerned with introducing errors into the actual data.

The approach proposed by T. Rethika , Ivy Prathap, R. Anitha and S.V. Raghavan these authors contributes a novel secure and efficient algorithm using the mathematical concept of eigen values for text watermarking. This concept motivated use to create tuple Relation matrix[6].

Other approach proposed by Vahab Pournaghshband[5] inserts new tuples that are not real and they call them "fake" tuples, to the relation as watermarks. this approach uses fake tuples and utilizes the insertion and detection watermarking algorithms. Evaluating watermarks for relational database is a challenge and requires further consideration. However, the persistency of the watermark after both malicious and benign updates, as a sub problem, might be evaluated by acquiring access to a log of user queries on a particular database over a reasonably long period of time, and then run the log on the watermarked database and observe whether the watermark detection algorithm will confirm

the watermark. While this evaluation process sounds plausible, it is application-specific and may not be generalized very well.

In our approach, we introduced the concept of eigen value based watermark generation to watermark non numeric attribute in the relational database.

3. OUR APPROACH

3.1 Process of Generating Watermark

The watermark generation process involves secret key generation using eigen values of tuple-Relation matrix for a tuple in the given relation.

In this paper, in subsequent sections, we use the Employee's personal Database as an example.

Secret Key Generation

Consider the Employee database, select low impact non numeric attributes such as address and city to be watermarked. Then compute the no. of vowels, consonants and special characters occurring in each tuple for selected non numeric low impact attributes of the relation.

As per the notation defined in figure 1, the weighted consonant sum C and weighted vowel sum V of high impact non numeric attribute of a tuple is calculated. Now the weighted ASCII sum A of each tuple is computed as below,

$$A = \frac{\sum_{i \in n} ASCII(c) \div n}{k} \qquad 0<i<n \qquad (1)$$

where ASCII(c) is ASCII value of the character c in the tuple. Tuple vectors are constructed with V, C, P, A as its components.

k	Number of tuples in the relation
n	Number of non numeric low impact attributes in the relation
V	ASCII values of the vowels of selected non numeric attribute are summed up to give the weighted vowel sum V
C	ASCII values of the consonant of selected non numeric attribute are summed up to give the weighted consonant sum C
P	the count of special characters of selected non numeric attribute
A	The weighted ASCII sum of the all character of selected non numeric attribute is calculated using formula (1) to give the weighted sum A
X	Concatenate all the eigen value
m	Secret key

Figure 1: Notation

The tuple-Relation matrix D is

$$D = [\, d_{i\,j} \,]_{n \times 4} \qquad (2)$$

where di1, di2 and di4 denote the weighted vowel sum, consonant sum, and ASCII sum of selected non numeric attribute in the given relation respectively and di3 denote the number of special characters of selected non numeric attribute in the given relation respectively. Each vector in the tuple-Relation matrix is not a unit vector and hence it is normalized as below,

$$N = [\, n_{i\,j} \,]_{n \times 4} \qquad (3)$$

$$n_{i\,j} = \frac{d_{i\,j}}{(\sum d2_{i\,j})1/2}$$

The normalized tuple-Relation matrix N is then pre-multiplied with its transpose NT to yield the watermark matrix W which is a square matrix of order 4. Let e1, e2, e3, e4 be the 4 eigen values of the watermark matrix W. Some of the eigen values may be zero when the rank of the matrix is less than or equal to 4. The precision of the eigen values is increased by multiplying each of the eigen values by 10.

The generated secret key is easily computable once we arrive at the tuple-Relation matrix. In O(n) time, the secret key can be generated from the relation. On the other hand it is not easy to form the tuple-Relation matrix even if the secret key is known. It is hard to find the tuple-Relation matrix even from the eigen values.

1) For each tuple r € R do
2) Get number of non numeric low impact attributes in n
3) Compute the count of special characters P for selected non numeric attribute in the tuple. Also calculate the weighted vowel sum V, consonant sum C and weighted ASCII sum A for selected non numeric attributes of tuple in the relation.
4) Construct the tuple-Relation matrix D for the selected tuple in given relation.
5) Normalize the matrix to get N.
6) Compute the watermark matrix W=NT*N, where NT denotes the transpose of N.
7) Find the eigen values of the watermark matrix, W. If they are floating points, convert into integers by multiplying by 10 to get two digit eigen value .
8) Concatenate all the eigen value in X.
9) If X < 8 digits then padding of zeroes to right
   else
   Take first 8 digit from right
10) Secret key m = X

Figure 2. Key Generation Algorithm

This algorithm produces a unique secret key of a tuple in given relation.

3.2 Watermark Insertion

Figure 2. gives secret Key for a tuple to be watermark. We now present a technique that marks only non numeric attributes and assumes that the marked attributes are such that small changes in some of their values are acceptable and no obvious. The data owner is responsible for deciding the impact of attributes suitable for marking.

For simplicity, assume that attributes address and city are of low impact from owner point of view which is used marking.

/* The secret key of a tuple is known only to the owner of the database. */

1) For each tuple r € R do
2) Use secrete key m, split the key into two digit integer to get n values, assign it to watermark vector (WV)= [wi…wn, i<n]
3) Select last n non numeric low impact attributes that will be marked
4) Sort the WV and get the ASCII character of WV.
5) Insert these WV character in n low impact non numeric attribute prefix and postfix alternatively.

Figure 3. Watermark Insertion Algorithm

In the step1 determines one of tuple in the relation. Take the secret key m which is generated by key generation algorithm, then split the key into two digit integer to get n different values, store these values in a watermark vector (WV) and assign the variable name to them as wi…wn, where i<n . The data owner has to decide the low impact attributes suitable for marking. The values within watermark vector need to sort out and get ASCII character of corresponding values. Then Insert these ASCII character in the selected low impact non numeric attribute prefix and postfix alternatively. When we insert marks in these attributes it won't be changing that much so it would not affect the actual data of the attribute.

3.3 Watermark Detection

/* The secret key of a watermarked tuple is known only to the owner of the database and also to the authenticated user. */

1) For each tuple r € R do
2) Get secrete key m, split the key into two digit integer to get n values, assign it to watermark vector (WV)= [wi…wn, i<n]
3) Select last n non numeric low impact attributes that has been marked
4) Sort the WV
5) Extract prefix and postfix character alternatively from n low impact non numeric attributes and store it's ASCII value in vector WV'.
6) If WV=WV', then watermark detected else suspicious data found.

The watermark is detected by using secrete key. The secret key is shared between the owner of the data and authenticated user. This will ensure the authenticate and integrity of data. If they differ the authenticated user can understand that some tampering has happened in between.

The detection algorithm is simple and less time consuming.

4. APPICATION

1. In the semiconductor industry, parametric data on semiconductor parts is provided primarily by three companies: Aspect, IHS, and IC Master. They all employ a large number of people to manually extract part specifications from datasheets. They then license these databases at high prices to design engineers. Companies like Acxiom have compiled large collections of consumer and business data.

2. In military applications where the information of personnel, armaments have to be kept secure, this database can be used.

3. The ACARS meteorological data used in building weather prediction models. The wind vector and temperature accuracies in this data are estimated to be within 1.8 m/s and 0:5◦ C respectively. The errors introduced by watermarking can easily be constrained to lie within the measurement tolerance in this data.

4. Consider experimentally obtained gene expression datasets that are being analyzed using various data mining techniques. The nature of some of the data sets and the analysis techniques is such that changes in a few data values will not affect the results.

5. The customer segmentation results of a consumer goods company will not be affected if the external provider of the supplementary data adds or subtracts some amount from a few transactions.

5. CONCLUSIONS

In this paper, we presented a new approach to watermark a non numeric attribute in the relational database and discussed the insertion and detection watermarking algorithms in details. A novel secure and efficient algorithm using the mathematical concept of eigen values for non numeric relational database is proposed. Its various features are discussed in detail. Any tampering to the data can successfully found. This algorithm can be used effectively where a huge amount of relational data is transferred between owner and authenticated users.

6. REFERENCES

[1] Agrawal, R., Haas, P., and Kiernan, J. 2003. Watermarking relational data: framework, algorithms and analysis. The VLDB Journal 12, 2 (Aug. 2003), 157-169. DOI= http://dx.doi.org/10.1007/s00778-003-0097-x.

[2] J.T. Brassil, S. Low, N.F. Maxemchuk , and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying ", IEEE Journal on Selected Areas in Communications", vol. 13, No. 8, October 1995, pp.1495-1504.

[3] Ding Haung, Hong Yan, "Interword distance changes represented by sine waves for watermarking text images", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, No.12, pp. 1237- 1245, Dec 2001.

[4] Zhang, Z., Jin, X., Wang, J., Li, D. 2004. Watermarking Relational Database Using Image. In Proceedings of the Third International Conference on Machine Leaning and Cybernetics, (Shanghai, August 26 – 29, 2004).

[5] Vahab Pournaghshband 2008 A New Watermarking Approach for Relational Data ,ACM-SE'08 March 28-29,2008,Auburn,AL,USA,ACM ISBN 978-1-60558-105-7/08/03.

[6] T. Rethika , Ivy Prathap, R. Anitha and S.V. Raghavan 2009 ESRGroups France A Novel Approach to Watermark Text Documents Based on Eigen Values.

[7] L. Boney, A. H. Tewfik, and K. N. Hamdy. Digital watermarks for audio signals. In International Conference on Multimedia Computing and Systems, Hiroshima, Japan, June 1996.

[8] M. Atallah and S. Lonardi, "Authentication of LZ-77 Compressed Data," *Proc. ACM Symp. Applied Computing*, 2003.

[9] M. Atallah, V. Raskin, C. Hempelman, M. Karahan, R. Sion, K. Triezenberg, and U. Topkara, "Natural Language Watermarking and Tamperproofing," *Proc. Fifth Int'l Information Hiding Workshop*, 2002.