International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 7 1827 – 1830

Optimization of Insertion Error using PSO with BSCS Collusion Secure Fingerprinting

Ms. Mudegol Nandinee Lingappa
Department of Computer Engineering
Dr. D. Y. Patil School of Engineering & Technology
e-mail:gmudegol@.com

Ms. Mohanpurkar Arti A.
Department of Computer Engineering
Dr. D. Y. Patil School of Engineering & Technology
e-mail: yasharti@gmail.com

***Abstract*—** Recently, most of digital data is shared among large group of users via internet so there is also growth in plagiarism, software piracy, and ownership theft. In this context, there is need of ownership protection and traitor identification. To address this issue we present a scheme by revisiting Kamran's watermarking technique. Watermarking is a technique used for protecting the ownership but fingerprinting is used for both ownership protection and traitor identification. Much more work is done on fingerprinting multimedia data but in this paper fingerprinting is done on relational data. The challenge in fingerprinting is that data must remain useful after inserting the marks. Here, marks means the collusion secure codes which are created using BSCS scheme. Optimized error insertion is achieved by using Particle Swarm Optimization (PSO). No separate tracing algorithm is required at the time of detection. All experiments will be done on medical database. The aim of the proposed system is to optimize the error insertion, make it collusion secure and identify the traitor.

Keywords- Boneh Shaw codes, Collusion secure, Fingerprint, Particle Swarm Optimization, Traitor tracing, EMR.

*****

I. INTRODUCTION

Tremendous growth in use of internet has introduces problems like plagiarism, software piracy, ownership thefts. Suppose, an owner of Relational Database (RDB) sells a database to three buyers, and if a dishonest buyer (traitor) sells that data personally to third party without taking permission from owner then there is need of ownership protection and traitor identification. There are two techniques to achieve this, watermarking and fingerprinting. Watermarking is a technique which can be used for ownership protection. There is lot of literature on watermarking multimedia content but for numeric relational data new techniques have been developed because nature of multimedia and relational data differ in number of aspects. The multimedia data mainly focuses on operations like zooming and compression but relational database mainly focuses on operations like insertion, deletion, updating. Some techniques used bit level marking and these bit positions and values are determined by algorithms that use secret key which is only known to the owner. The same secret key is used during watermark embedding as well as watermark verification [1].

The main challenge in watermarking is that after inserting watermark the data must remain useful for the intended purpose i.e. insertion of marks should not reduce the usability of original data. If we don't modify the data then we can't insert the watermark. That's why we have to limit the change which is acceptable. The acceptable nature and level of change depends on the type of application for which the data is to be used. For example, the medical data or stock market data is very sensitive, a small change in data leads to very large changes in the results.

Fingerprinting is also a class of information hiding that inserts digital marks into data. Unlike watermarking which is used for the purpose of identifying sources of data, fingerprinting is used to identify the recipient to whom the data has been provided [3]. In that the buyer's identification with secret key is given as fingerprint known as a *buyer-specific mark* which is embedded into a data copy provided to the buyer, so that the owner can detect the mark in pirated copy and the buyer-specific mark helps to detect the traitor. We can use the randomly selected numbers as buyer's Id. This buyer's identification is taken as a mark like in watermarking technique. Fingerprinting consists of two phases fingerprint embedding and fingerprint detection [6].

The usability constraints on the numeric data can be maintained by using optimization techniques such as Particle Swarm Optimization (PSO). It is a technique inspired by social behaviour of the birds and it helps to solve the continuous problems. The basic idea behind PSO is that, there are multiple solutions. Each candidate solution is called *Particle*; population of all particles is a set of vectors known as *Swarm*. The particles change their components and fly in the space. The actual position of particles can be evaluated by using a function known as *Fitness Function*, which is to be optimized. Each particle is characterized by position vector and velocity vector and has the individual knowledge as well as the social knowledge of its best neighbour. On the basis of this information the optimum solution can be obtained. Here fingerprinting is made collusion resistant using Boneh Shaw (BOSH) fingerprinting codes.

There are more techniques for fingerprinting multimedia data that are not applicable to relational database. Relational data differs from multimedia data in following aspects so there is need to generate new techniques for fingerprinting relational database: [6]

1. Multimedia data contains large number of bits providing large cover to hide fingerprint, but relational database consists of set of independent objects i. e. tuples and the fingerprint has to be embedded in tuples.
2. The relative positions of objects doesn't change in the multimedia data but there is no proper ordering among the tuples in relational database, so the tuples are out of order.
3. Multimedia data is not updated frequently i.e. any object is not dropped or replaced normally, but the operations on relational database lead to the insertion, deletion, updating of the tuples.

The objectives of the proposed system are as follows:

1. Ownership protection with Traitor identification.

2. Providing an optimized technique for fingerprinting.
3. Providing collusion secure fingerprints.

In this paper, we are proposing a system which is collusion secure and which inserts optimized error in the relational database. The acceptable change is computed using PSO which is minimal. The Boneh-Shaw's concatenated codewords are used for collusion secure purpose. In that the inner code is the Boneh Shaw Replication Scheme (BS-RS) code and the outer code is the Random Code.

The rest of the paper is organized as follows: section II describes the literature survey. Problem Statement and Description of system is given in section III. Section IV gives the System Architecture and overview of the proposed system. Section V discusses the expected results of the system. The conclusion is given in section VI.

II. LITERATURE SURVEY

Agrawal et al. [1] proposed the idea of watermarking using least significant bits (LSB). Multi-bit watermarking was not accounted in this paper which makes this technique vulnerable against simple attacks, for example shifting of only one least significant bit (LSB) results in loss of watermark.

Yingjiu Li [3] presented a marking scheme that permits an arbitrary mark bit-string to be embedded in a relation using a single secret key. The mark bit-string can be used to represent different buyers who purchase the database relation. The detection algorithm tests whether a key was used to mark a relation and, if so, it returns the actual mark bit-string that was embedded. The marking scheme can be used for both watermarking and fingerprinting. The only difference is that in watermarking same bit-string is embedded and detected but in fingerprinting different bit-strings are embedded and detected. Here the given detection algorithm is not collusion secure. Solution for collusion-secure detection using Boneh Shaw code is suggested where fingerprint is divided into two halves. First half is calculated using hash function with buyer's identification and secret key. Second half is calculated using Boneh Shaw codes.

Julien Lafaye [9], proposed WATERMILL: an optimized watermarking and fingerprinting system for databases under constraints using the watermill database. A built-in usability constraint definition language and an efficient watermarking engine is used to override the limitations of the greedy method. Tardos code is used to make the fingerprinting collusion secure but the fingerprint is applied using LSB.

Hans Georg Schaathun, Marcel Fernandez [14], improves the decoding algorithm with the help of soft output from the inner decoder, and has shown that this permits using significantly shorter codewords. It uses soft decision decoder for that purpose. Here soft information is passed from inner to the outer decoder. It achieves the equivalent performance as that of Tardos codes. A soft-decision decoder is a class of algorithm used to decode data that has been encoded with an error correcting code. A soft-decision decoder will typically perform better in the presence of corrupted data than its hard-decision counterpart, because hard decision decoder only operates on fixed set of values for example, 0 and 1 in binary code but soft decision decoder operates on all the values in the range of 0 to 1.

Jos'e Moreira S' anchez in [15], in their master thesis, presented and analyzed some of the main existing fingerprinting codes and also discussed some new constructions. Their study specifically focuses on the estimation of the minimum length of the codes, given the design parameters of the system: number of users to allocate, maximum size of the collusions and probability of identification error. They have also presented a comparative analysis of the families of codes. Their results show the regions where it is preferable to use each code, given the design parameters like number of users to allocate, maximum size of the collusions and identification error probability.

III. PROBLEM DESCRIPTION AND SPECIFICATION

A. *Problem Statement*

To create a fingerprint this is secure in opposition to collusion attack. After insertion of fingerprint the data should be remain useful for the intended purpose. Identify the traitor, if any.

B. *Problem Description*

In medical field the patients' medical information is stored in EMR (Electronic Medical Records) and will be used in future for the diagnosis purpose. If anyone changes this data by mistake or intentionally, it may lead to misdiagnosis and may also results in the patient's death. The defined problem is to protect the ownership of data and identify the buyer who illegally supplies the owner's copy of data to third party, known as *traitor*. Collusion secure fingerprinting is another important concern.

Another challenge in fingerprinting is that data should be secure against collusion attacks. Suppose an owner of the EMR system sells this data to two persons (known as **buyer**). Each of them is given a somewhat different (unique) copy of the data with inserted fingerprints. If they compare their data copies with each other and find out the fingerprint positions where fingerprints are inserted, with the help of differences. Then the buyers can remove or alter these fingerprints and resell the data as their own without any owner's consideration. So the data must be collusion secure. The buyers can detect the fingerprints only when they find out the differences in their data copies. If the data copies are same then they can't detect the fingerprints is the *marking assumption* [13].

IV. SYSTEM ARCHITECTURE

The system architecture depicted in fig. 1 represents the design of the proposed system. The architecture is divided into three parts i.e. Fingerprint computation, Fingerprint embedding, and Fingerprint decoding. The fingerprint can be inserted on the fly at runtime because it is computed before embedding process. At the time of creation PSO is used to compute the minimum acceptable change for every numeric attribute. The Boneh Shaw concatenated code is also computed.

 ISSN: 2321-8169
1827 – 1830

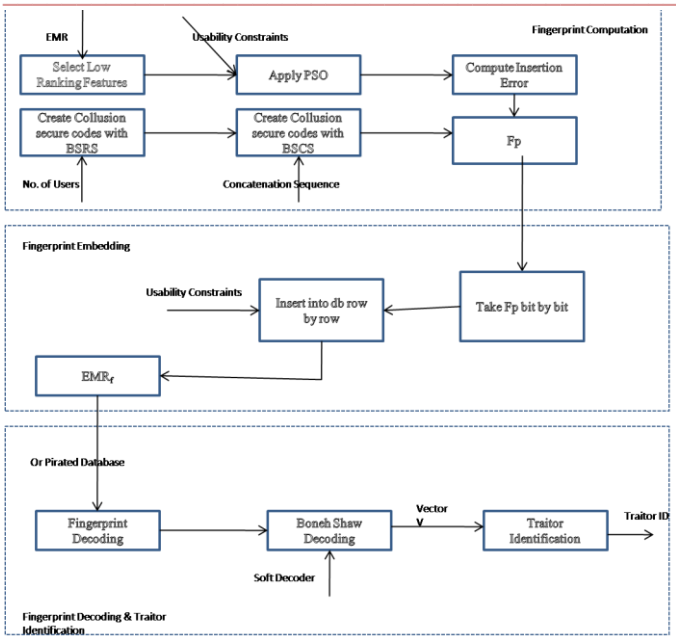

Figure 1. System Architecture

At the time of insertion, the computed acceptable change is added or subtracted from the attribute value based on the Boneh Shaw code. At the time of detection the number of ones and zeros are detected on the basis of classification potential after insertion of fingerprint and the acceptable change.

A. Fingerprint Creation:

While creating the Fingerprint, first all numeric attributes are selected. Then these attributes are ranked according to the classification potential. After that low ranking features are selected for marking because high ranking features are more sensitive to the change.

After selection of attributes, apply PSO to calculate the minimum acceptable change for each attribute. And finally calculate the Boneh Shaw collusion secure codes using BSCS Scheme.

Algorithm 1. Fingerprint Computation:
Input: EMR, No. of Users, Usability Constraints
Output: Acceptable Change, Boneh Shaw code
1. Select only Numeric Attributes
2. Calculate CP
3. Select low ranking attributes
4. Apply PSO to compute acceptable Change
5. Calculate Boneh Shaw Code

B. Fingerprint Embedding:

While embedding the fingerprint, based on Boneh Shaw codes, if the bit is 1 the insertion error is subtracted from the attribute value and if the bit is 0, insertion error is added to that attribute value.

Algorithm 2. Fingerprint Embedding:
Input: Acceptable change, Boneh Shaw Code
Output: Fingerprinted Database
1. Compute Insertion Error using Acceptable Change and Attribute Value.
2. if bit is 0
   Add Insertion Error to the Attribute value
3. else
   Subtract Insertion Error from the Attribute value
4. Store that insertion error in Δ

C. Fingerprint Decodingand Traitor Identification:

After insertion we will get the fingerprinted database. This copy is given as input to detection phase. In the detection phase inserted $F_P$is obtained without having the original database (EMR). The detection algorithm is given in [5].

Algorithm 3. Fingerprint Decoding:
Input: Fingerprinted database or Pirated database, Δ
Output: Detected Fingerprint
1. Initialize no. of one's and zero's count to zero
2. Compute change in original Db value and pirated Db value(ΔFp)
3. if $\Delta Fp \leq 0$ then
   Detected bit is 1 and increment no. of one's count by 1
4. else if $\Delta Fp > 0$ and $\Delta Fp \leq$ threshold then
   Detected bit is 0 and increment no. of zero's count by 1
5. else detected bit is an unknown mark
6. Detected fingerprint = mode(dt(1,2,……,l))

This fingerprint is then given to the tracing algorithm given in [14] which uses the soft information returned by the soft decision decoder for the decoding of inner code. The output is vector (V_O={V_{O1},V_{O2},….,V_{OP}}) which is given as:

$$V_{OJ} = \frac{F_{PJ} - F_{PJ-1}}{R} \quad (4)$$

The idea of soft decision decoder is that all V_{OJ} sum to 1, if V_{OJ} is close to zero then the J is incorrect. If V_{OJ} is larger, then J is correct decoding. If this algorithm outputs at least one guilty user if any then we says that the algorithm is successful. If the algorithm outputs the innocent user then we say that it is failed.

After decoding of inner code of all blocks a reliability matrix R_B will be formed and it is given as the input to the outer decoding algorithm. The outer decoding algorithm returns all the code-words for comparison to find guilty user, if any.[5][14]

V. RESULTS

In this section the experimental results are discussed. The experiments are performed on medical database of thyroid patients using Microsoft SQL Server which is running on Intel Core 2 Duo CPU with 2.10 GHz processor with 4 GB of RAM.

Fingerprint insertion preserves the classification potential of high ranking features means CP of features before insertion of fingerprint as well as after insertion of fingerprint does not change. The inserted fingerprint is collusion secure. The work on decoding process is in progress.

TABLE I. EXPERIMENTAL RESULTS1

Sr. No.	Attribute	CP	
		Before	*After*
1	A1	39.804	39.804
2	A2	30.194	30.194
3	A3	30.003	30.003

TABLE II. EXPERIMENTAL RESULTS2

Sr. No.	Attribute	Mean	
		Before	*After*
1	A1	1.145	1.145
2	A2	11.305	11.316
3	A3	3.838	3.849

TABLE III. EXPERIMENTAL RESULTS3

Sr. No.	Attribute	Variance	
		Before	*After*
1	A1	0.273	0.273
2	A2	27.224	27.278
3	A3	19.224	19.335

VI. CONCLUSION

The Fingerprint Insertion technique described here preserves the Classification Potential of high ranking features. The fingerprint is calculated prior to the insertion process, so fingerprint can be inserted on the fly in real time. The codes used for collusion secure purpose are generated using Boneh Shaw method and decoded using soft decision decoder. Hence the decoding complexity is as that of Tardos code.

REFERENCES

[1] Rakesh Agrawal, Peter J. Haas, Jerry Kiernan:"Watermarking relational data: framework, algorithms and analysis", The VLDB Journal (2003) / Digital Object Identifier (DOI) 10.1007/s00778-003-0097-x (2003).

[2] Radu Sion, Mikhail Atallah, Fellow, IEEE, and Sunil Prabhakar: "Rights Protection for Relational Data", IEEE Transaction on Knowledge and Data Engineering, Vol. 16, No. 6, June 2004.

[3] Yingjiu Li, "Fingerprinting Relational Databases: Schemes and Specialties", IEEE Transactions On Dependable And Secure Computing, Vol. 2, NO. 1, January-March 2005.

[4] M. A. Panduro: "A Comparison of Genetic Algorithms, Particle Swarm Optimization and Differential Evolution Method For The Design of Scannable Circular Antenna Arrays", Progress In Electromagnetics Research B, Vol. 13, 171-186, 2009.

[5] M. Kamran and Muddassar Farooq, "An Information-Preserving Watermarking Scheme for Right Protection of EMR Systems", IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 11, November 2012.

[6] Raju Halder, Shantanu Pal, Agostino Cortesi, "Watermarking Techniques for Relational Databases: Survey, Classification and Comparison", Journal of Universal Computer Science, Vol. 16, no. 21 (2010).

[7] Xinwei Li, Baolong Guo, Long Chen, Xianxiang Wu, Leida Li, "A high capacity and strong robust fingerprinting for compressed images", Computers and Electrical Engineering 38 (2012) 1249–1261.

[8] Jooyoung Lee and Deukjo Hong, "Collusion Resistance of the JH Hash Function, IEEE Transactions On Information Theory", Vol. 58, No. 3, MARCH 2012.

[9] Julien Lafaye, David Gross-Amblard, Camelia Constantin, and Meryem Guerrouani, "Watermill: An Optimized Fingerprinting System for Databases under Constraints, IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 4, April 2008".

[10] Min Wu, Wade Trappe, Z. Jane Wang, and K.J. Ray Liu, "Collusion-Resistant Fingerprinting for Multimedia", IEEE Signal Processing Magazine 15 1053-5888/04/©2004IEEE.

[11] Benny Chor, Amos Fiat, Moni Naor, and Benny Pinkas, "Tracing Traitors", IEEE Transactions on Information Theory, Vol. 46, No. 3, May 2000.

[12] Wade Trappe, Member, IEEE, MinWu, Member, IEEE, Z. JaneWang, Member, IEEE, and K. J. Ray Liu, Fellow, IEEE, "Anti-collusion Fingerprinting for Multimedia", IEEE Transactions on Signal Processing, Vol. 51, No. 4, Aril 2003.

[13] Dan Boneh and James Shaw, "Collusion-Secure Fingerprinting for Digital Data", IEEE Transactions On Information Theory, VOL. 44, NO. 5, September 1998.

[14] Hans Georg Schaathun, Marcel Fernandez, "Boneh-Shaw Fingerprinting and Soft Decision Decoding", In the Proc. of IEEE ISOC ITW2005 on Coding and Complexity; editor M.J. Dinneen; co-chairs U. Speidel and D. Taylor; pages 183-186.

[15] Jos´e Moreira S´anchez, master thesis, "Properties and evaluation of fingerprinting codes", Escola T`ecnica Superior d'Enginyeria de Telecomunicaci´o de Barcelona, Universitat Polit`ecnica de Catalunya September 2009.

.