A Review on Watermarking Techniques for Relational Data

[1]Sharafunisa S, [2] Smitha E S
[1]M.Tech Student, [2]Associate Professor
[1]Departmentof Computer Science and Engineering,
[1]LBS Institute of Technology for Women, Thiruvananthapuram, India

Abstract—Nowadays internet is providing many web based ser-vices such as digital repositories, libraries and e-commerce etc. For using these applications the digital information such as image, audio, video, relational data are made publicly available. So there is a need to protect these digital assets from various threats like ownership claiming, piracy, theft etc. Watermarking is a solution to overcome these issues. Using watermark we can implement both tamper protection and ownership proof. Encryption is also a possible solution, but the encrypted data cannot be used for data mining applications and ownership protection cannot be achieved. The watermarking techniques encode the data in such a manner that it remains useful for recipients. The data owner watermark the data and decodes later if required when, ownership proof is required. The recipient can also get valuable information from the watermarked data without need for any pre-processing. This paper presents review on some of the watermarking technique for relational database.

Index Terms—Watermark, Genetic Algorithm (GA), Prediction Error Expansion (PEE), Difference Expansion Watermarking (DEW), Reversible and robust watermarking (RRW).

I. INTRODUCTION

Internet is offering a wide range of web based services, such as database as a service, digital repositories and libraries, e-commerce, online decision support system etc. These applications make the digital assets, such as digital images, video, audio, database content etc., easily accessible by ordinary people around the world for sharing, purchasing, distributing, or many other purposes. As a result of this, such digital products are facing serious challenges like piracy, illegal redistribution, ownership claiming, forgery, theft etc. Digital watermarking technology is an effective solution to meet such challenges. A watermark is considered to be some kind of information that is embedded into underlying data for tamper detection and ownership proof, traitor tracing etc. Initially, most of the work on watermarking was concentrated on watermarking still images, video, audio, etc. However in the recent years watermarking of database system started to receive attention. In general database watermarking technique consists of two phases: watermark embedding and watermark verification. During watermark embedding phase, a private key K is used to embed the watermark into the original database. The private K is only known to the owner of the database. The watermarked database is made publicly available. To verify the ownership of a suspicious database, verification process is performed where the suspicious database is taken as input and by using the private key K the embedded watermark is extracted and compared with the original watermark information.

II. WATERMARKING FOR RELATIONAL DATABASE

Digital watermarking for relational databases emerged as a candidate solution to provide copyright protection, tamper detection, traitor tracing and maintaining integrity of relational data [1].

Reversible Watermarking Techniques

Irreversible watermarking helps to prove ownership rights, but the embedding process usually alters data to a large extent. These techniques generally comprises of two phases: watermark embedding and watermark extraction. Reversible watermarking on the other hand recovers the original data and provides water-mark information for ownership protection. Reversible watermarking technique for relational database can be generalized into four phases: Pre-processing, Encoding, Decoding and Data Recovery [2].

1.Data Pre-processing:

In data pre-processing step two sub modules perform different tasks to accomplish desired goals such as selection of suitable feature for watermark embedding and formation of a watermark.

a. Feature analysis and selection:

In these statistical measures are used to rank the relational database features according to their importance in the information extraction process. Ranking of features is usually performed by measuring the mutual information (MI), information gain and entropy.

b. Watermark Generation:

The watermark can be generated using different techniques such as evolutionary algorithms, pseudo-random sequences and hashing techniques.

2.Watermark encoding:

During watermark encoding phase, watermark information is embedded in the selected feature. A number of parameters are also computed in the encoding phase for use in watermark encoding and in the subsequent decoding phase.

3.Watermark Decoding:

In the watermark decoding phase, the embedded watermark is decoded from the suspicious data. The pre-processing step is performed, and decoding strategies are used for recovering watermark. In verifying the watermark information, the original watermark and the detected watermark are compared. Watermark detected should be the same as the watermark inserted into the data to prove ownership rights.

4. Data Recovery:

The original data are recovered in the data recovery phase through post-processing steps for error correction and recovery. After recovery, the recovered data might also be compared with the original data to ensure that the data quality is not compromised.

III. LITERATURE SURVEY

In this section we summarize some of the watermarking techniques for relational database. Agarwal et. al [3] proposed the first irreversible watermarking technique for relational databases. The algorithm proposed is based on the primary key and a secret key. The algorithm works on numeric attributes. They are inserting a single bit into the numerical field of the database and detecting it with the help of detection algorithm. The tuples for watermark insertion is selected using one way hash function. The message authentication code (MAC) is a one way hash function that depends on a key. In this algorithm for selecting the tuple for watermark insertion is selected using a MAC that uses the value of primary key attribute and the secret key K that is only known to the owner of the database. The attribute within the tuple, the bit position to be marked is also found using MAC, different parameters are used to take the modulus value. The selected bit positions are set 0 or 1 based on the hash value obtained using the primary key and the secret key. The watermark detection is probabilistic in nature. The detection algorithm takes a subset of tuples and using hash function determines whether the tuples are marked or not. The detection algorithm is blind also, that means the detection algorithm does not need the knowledge of neither the original data nor the watermark.

Ali and Ashraf [4] proposed a robust and blind technique for database watermarking. The proposed technique is an image based watermarking scheme. In image based watermarking algorithms, an image is used to watermark the database. The image is transformed into bits which represents watermark bits. The bits are embedded in carefully chosen locations in database and if recovered correctly can be used to reconstruct the embedded image. In the watermarking technique proposed by Ali and Ashraf a binary image is embedded in spaces of non-numeric multi word attributes of subsets of tuples, instead of numeric attributes at bit level. The bits of image are segmented into short binary strings. The database is also logically divided into subset of tuples. The embedding process of each short binary string is based on creating a double space at a location determined by the decimal equivalent of the short string. Extraction of the short string is done by counting number of single spaces between two separated double space locations. The image watermark is then constructed by the converting the decimals into binary strings. A major advantage of using the space-based watermarking is the large bit capacity available for hiding the watermark. Since, the proposed algorithm embeds the same watermark for all non-intersecting subsets of the database; it is robust against subset deletion, subset addition, subset alteration and subset selection attacks. This is an irreversible technique. Extraction of a short binary string is done by counting number of single spaces between two separated double space locations.

Farfoura et al. [5], proposes a blind reversible watermarking technique for owner-ship protection. A reversible data hiding technique known as prediction error expansion has been used to ensure reversibility. In this technique an image selected by the owner of the database is used as the watermark information. The identification image is converts into a stream of bits 0s and 1s. This techniques works on numeric attributes. Embedding watermark into database may incur distortion in the values. A tolerable change in the attributes value can be accepted, and will not affect the overall data usability. To minimize the amount of distortion that will be introduced to original data, watermark bits will be embedded in the fraction portion of the numeric attributes. Record and features for marking are selected on the basis of a one-way cryptographic hash function. In the watermark insertion phase, the converted bit stream of identification image is embedded into relational data for representing copyright information. . In this algorithm prediction error expansion (PE) is uses for watermark embedding. PE assumes two intensities for each pixel, the original intensity y and the predicted one $y^{\wedge}$. Here we assume that y is the fraction portion of an attribute and $y^{\wedge}$=LSB (H (ti.P||K)), is any value known at encoding and decoding time. The difference between y and $y^{\wedge}$ has been calculated. This converted into to a binary value to concatenate the watermark bits with the expanded difference and is converted back to an integer value. The expanded difference is added with $y^{\wedge}$ and finally it is concatenated with the integer portion to achieve the watermark encoding. In the watermark detection phase, tuples with marked features are identified again by taking modulus. Fractional and integer portions of the feature values are again extracted. In the detection phase a majority voting scheme is applied to find the final watermark information.

Genetic algorithm based difference expansion watermarking proposed by Jawad and Khan [6], is a reversible watermarking technique that recovers both watermark and cover work exactly as it was before watermark insertion. In this technique, the watermark insertion is divided into three modules preprocessing module, GA module and insertion module. Pre-processing module is responsible for providing initial values to DEW technique and to select appropriate fitness parameters for GA module. Selection of tuple is performed by hashing technique, known as MAC. It requires primary key of selected tuples and secret key chosen by the owner. DEW technique only considers two features for a selected record based on the amount of distortion that they can tolerate and thus result in low watermark capacity and high amount of distortion.GA module is responsible for obtaining best possible chromosome according to the selected fitness function. Best chromosome of GA is passed to insertion module of GADEW. The watermark insertion is implemented using difference expansion technique. The watermark bits are computed using Message authentication code and embedded into the selected tuple using DEW. If the change in value of data is within the distortion tolerance ranges, then the dataset is updated accordingly. In the watermark detection process, the original data are restored by sorting out the record and features based on the primary key and ascending order of their names, respectively. Watermark bits are again detected on the basis of MAC. The original values of features and watermark bits are detected by applying the DEW technique. If these values are within the distortion limits, and restored watermark bits are the same detected through MAC, then the original data are restored.

In RRW [7], 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.

IV. CONCLUSION

In this paper, a survey of watermarking techniques for relational database has been presented.The databases are used effectively in collaborative environments for information ex-traction; consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Watermarking is advocated to enforce ownership rights over shared relational data and for providing a means for tackling data tampering. When ownership rights are enforced using watermarking, the underlying data undergoes certain modifications; as a result of which, the data quality gets compromised. There are two types of watermarking techniques reversible and irreversible. In reversible methods

the original data is again recovered from the watermarked relation. Reversible watermarking is employed to ensure data quality along with data recovery. Both categories are robust against various malicious attacks. There for this Survey has been presented for an overview of water-marking technique for relational database and some existing watermarking techniques.

IV. ACKNOWLEDGMENT

We are greatly indebted to our principal Dr. K. C. RAVEENDRANATHAN, Dr. SHREELEKSHMI R., Professor, Head of the Department of Computer Science and Engineering, Mr. MANOJ KUMAR G., Associate Professor, Department of Computer Science and Engineering, LBS Institute of Technology for Women who has 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, Shantanu 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] S.Ifthikar, M.Kamran and Z.Anwar", A Survey on Reversible Watermarking Techniques for Relational Databases" ,Security and communication networks ,2015

[3] R.Agarwal and J.Kiernan, "Watermarking relational databases", Proc. 28th Int. Conf. Very Large DataBases , 2002,pp.155-166.

[4] Al-Haj.A and Odeh.A ," Robust and blind watermarking of relational database systems" , Journal of Computer Science , vol.4, 2008,pp.1024-129

[5] Farfoura ME and Horng S-J," A novel blind reversible method for watermarking relational databases", 2010 International Symposium on Parallel and Distributed Processing with Applications (ISPA), IEEE, Taipei, Taiwan, 2010,pp.563-569.

[6] Jawad K, Khan A, "Genetic algorithm and difference expansion based reversible watermarking for relational databases",Journal of Systems and Software , 2013,pp.2742-2753.

[7] S.Ifthikar, M.Kamran and Z.Anwar."RRW-Reversible and Robust Watermarking for relational data", IEEE transactions on Knowledge and Data Engineering, 2015, Volume: 27, Issue: 4, pp: 1132-1145.