ISSN:0975-9646

Database Watermarking Using SHA 512 Signature Generation Technique

Ranjana Waman Gore, Rucha Tare

Department of Information Technology, Marathwada Institute of Technolgy,
Aurangabad, Maharashtra, India

Abstract— **The use of relational databases is growing rapidly over the internet and also in many applications worldwide and so it faces various issues like ownership right issues, tampering issues and many more. In this paper we present a technique which can eradicate these issues using the SHA512 signature generation technique in which we embed a watermark in the tuples of the database without changing the original data. That is by adding an extra column of signature bit to the database. This technique is resilient to all sorts of attacks on the databases which include deletion attack, alteration attack, and insertion attack too. We use a checksum to verify the tampering in the database. This checksum verifies the signature extracted. The database signature is then encrypted with public key using RSA algorithm to obtain security and authorization. This paper shows the implementation of our technique experimentally and also proves it better than the existing systems as we have compared our system with already existing systems which clearly shows the merits of our system. The advantage of our system is that it doesn't change the original data unlike all other watermarking techniques that modify the bits in the tuples which may alter the usefulness of the data.**

Keywords—**database, signature, watermark.**

I. Introduction

Internet is an important and widely used medium for digital media as it is free of cost, available always, efficient and easily accessible. Through the Internet we can access databases also. Undoubtedly this feature has numerous merits but it comes with several demerits too. The threat that your data may get stolen is one of the demerits. Thus there is a need to find a new technique for the users that identifies illegal copies of their databases. Watermarking relational databases is one technique which can overcome the problem of copyrights, false ownership issues; detection of the tampering in the databases. The watermarking technique we used is hashing algorithm using SHA512 which adds a signature to tuple of the databases. This watermark can be summarized as adding an extra column to the rows of the databases without even touching the original data in the database. As we are not changing the original data it becomes an advantage for us over all other watermarking systems. This signature is then encrypted using the admin or owner's public key and is sent to the other end. This signature is extracted on the other end by the authorized receiver. And then the verification of data is done.

Internet can distribute contents all over the web in such cases user has to take security measures for the database ownership and database tampering. In these cases watermarking solves the problem to a large extent. Watermarking techniques allows the owner of the database to embed an imperceptible watermark into tuple of the database. A watermark consists of marks that can prove the ownership of database and the owner can claim that the database belongs to him. The watermarking technique also helps to identify the sender of the data, and is the data sender genuine. The watermarking technique we have used here does not get affected to any attacks by the attacker. Imperceptible watermark means the watermark will be invisible and is not noticeable in the data. At present we have watermarking techniques for video, images, audio, and text but watermarking for the relational databases is not yet developed and this problem remains unnoticed. But protection of data is an important issue for many applications like database for university, banks or for example here we have used stock market database and weather monitoring database. We have also specified authorization that gives different access permissions to different users in the same system to make it more secure. The owner here specified as admin of the system has all the access permissions (read, write, update, send) where as the receivers have can view and extract the watermark but cannot manipulate it. And lastly the end users can only view the data as they are unauthorized users without username and password. The attackers or the hackers are treated as the end users only so they do not have access to any data. In this paper we have proposed an efficient technique which overcomes all the attacks .And without the data in the database getting affected.

Here we have used hash algorithm SHA (Secure Hash Algorithm) which was developed by National Institute of Standards and Technology (NIST) and published as federal information processing standard [1] in 1993.SHA offers many algorithms as its newer version for example SHA1, SHA2, and SHA512. We have used SHA512. SHA denoted authorization of the system. Signature uses "Asymmetric Cryptography" which employs an algorithm uses two different keys one for creating signature, and, another for verification of the signature.

A. Related Previous Work

1. Introducing Minor Changes in the Database

The idea of watermarking in this technique is to make minor changes in the existing database, these changes are too minor to identify by any attacker and are tolerable as the databases normally can tolerate few changes in it. The most important thing to be taken into consideration is that these changes should not affect the usefulness of the data and no adverse effect is caused.

2. Watermark Based On the Bit Pattern
Here in this technique bit pattern plays an important role in watermarking. The bit pattern its attributes and its positions are algorithmically defined to create a watermark
There are certain limitations of in the discussed systems such as hackers can easily access the data, there are no security measures are provided, access permissions are not specified, they are not resilient to any sort of attacks by the attacker and also cannot predict the tampering.

II. Our Approach

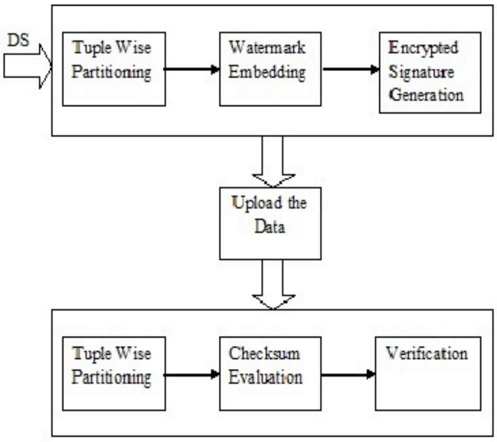

Fig. 1 Watermarking concept diagram

Fig. 1 shows a diagram summarizing the main concept of the watermarking system model used. A database Db is transformed into a watermarked version W by applying a watermark embedding function that also takes an input of public key PK only known and used by the admin of the database. Watermarking changes the original data but these changes are tolerable
The watermark encryption stage can be explained by the following steps:
Step 1. **Tuple wise partitioning:** The database Db is partitioned into x number of partitions
Step 2. **Signature generation:** The sha512 algorithm is used to generate the signature that is used for further encryption. The encryption is done using RSA algorithm with the public key of the admin or the owner. This encrypted signature is sent to the other end along with the original data.
Step 3. Watermark embedding: A watermark bit is embedded in the tuple as mentioned above The watermarked version W is delivered to the desired recipient.
Watermark decryption is the way of extracting the watermark using the database Db and the private KP and the signature generated. The decryption algorithm is not blind as the original database Db and the signature generated is required for the successful decryption of the embedded watermark.
The watermark decryption is divided into three main steps:
Step 1. **Tuple wise partitioning:** The database Db is partitioned: by using the data partitioning algorithm used in Encryption stage, the data partitions are generated.
Step 2. **Checksum evaluation**: The signature is extracted using private key of the receiver and tuples of each partition are checked and checksum is evaluated.
Step 3. Verification: The watermark bits are verified and tampering is indentified.
We have implemented the watermarking technique into following two applications.

1. Stock Market
2. Weather Report

These two applications contain a centralized database with security. We have implemented watermarking techniques to these applications. First the given database will be partitioned based on the tuples. The watermark will be embedded by modifying these partitioned sets. The owner uses public key to embed watermark. On the receiver side first it will check the user authentication. If the receiver is admin it has all the access permissions. The watermark can be decrypted using the private key, the original database, and the signature generated. The checksum is evaluated and the data is verified.

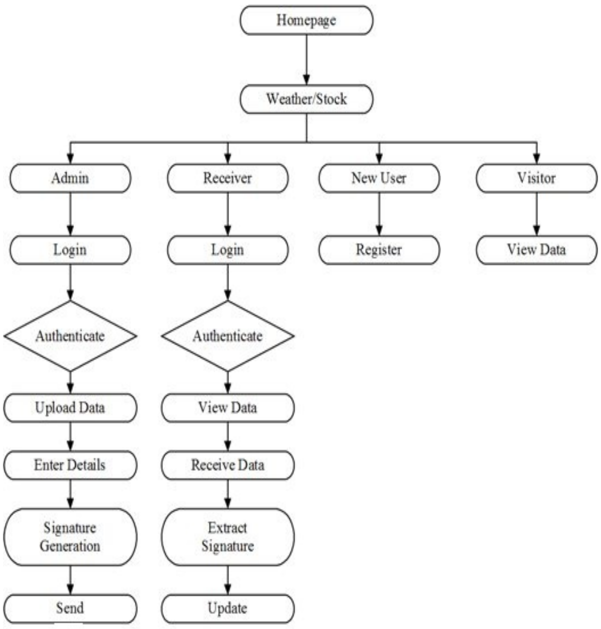

Fig. 2 Flow Chart of the System

The above Fig. 2 shows the flow chart of the system. The first page that user will access to, is the homepage. The homepage has two modules, weather and stock. These weather and stock are the real time databases that we have used in the system. Each module has two different login pages that are for Admin and Receiver, also it has a Signup page for new user. The admin has all access permission. Only admin can upload the data and embed watermark. A signature is generated and this signature in encrypted form is send to the receiver. The receiver logs into the system and can see the receiver homepage. The receiver has two options, either to view the data or to receive the data. The receiver has to select the data and extract the signature and update the data into the database. The visitors/end user can only view the data

Module Description**:** There are three main modules in the system as summarized below:

A. *Data Partition:*

Data partition module contains two sub modules for collecting data from admin side. First one is “stock information” second one is “weather report”. In stock information sub module admin will insert details about the company and final stock range of these companies. Each and every change of the stock details should updated by the admin. Like these weather reports too. This is module fully handled by the admin. After this process the flow is directed to watermark encryption module.

B. *Watermark Encryption:*

This module plays an important role in the system. Here we embed watermark to the inserted data tuple using the SHA512. After the watermark is embedded the data will be hidden and the signature will be generated. This encrypted signature will be sent to the intended recipient. RSA with its Public key is used to embed the watermark that is encrypting the signature.

C. *Watermark Detection:*

At the time of retrieval the watermark is decrypted using the original data, the signature generated and the private key used by the recipient. But before that access permission of the end users are checked, based on their username and password. If the user is admin [receiver side] he has all the access permissions. If it is an end user [visitor] he can only view the data.

If hackers want to access our data in-between the network, our application will check the access permission and the key. Hackers will be considered end users so they can’t access the data without help of admin. We have used SHA512 and RSA algorithm. The data partition and the rsa algorithm are used [2].

III. Advantages And Analysis

A. *Usefulness*

The modifications made during the watermark embedding in the tuples that is the generation of signature bit in the database does not affect it and is more reliable as we directly do not change the original data ,we just add a signature bit thus the usefulness of the system is not degraded.

B. *Resilience*

The system is resilient to all the external and internal attacks. This system is resilient to attacks like deletion, alteration and insertion.

C. *Determination of Ownership*

As the signature will be extracted from the receiver’s side using its private key the data will e verified and thus it will determine the ownership as it is authorized with SHA512 and is confidential with RSA.

D. *Structure*

A database is made of correlated tuples. The watermark does not alter the structure of these tuples as no change in the data is made.

E. *Authorizations and Access Permissions*

As stated earlier we have specified different access permissions to different users. If the user is admin [receiver side] he has all the access permissions. If it is an end user [visitor] he can only view the data. If hackers want to access our data in-between the network, our application will check the access permission and the key. Hackers will be considered end users so they can’t access the data without help of admin

TABLE I: COMPARISON OF VARIOUS EXSTING TECHNIQUES WITH OUR TECHNIQUE

	R. Agrawal And J. Kiernan (2002)	R. Sion ,M.Atallah And S. Prabhakar (2004)	Our Technique Using Sha 512
Watermark Information	Bit Pattern	Binary String	Signature Generation
Intention	Proof Of Ownership	Proof Of Ownership	Proof Of Ownership Tamper Detection
Resilient To Deletion Attack	No	Yes	Yes
Resilient To Alteration Attack	No	Yes	Yes
Resilient To Insertion Attack	No	No	Yes

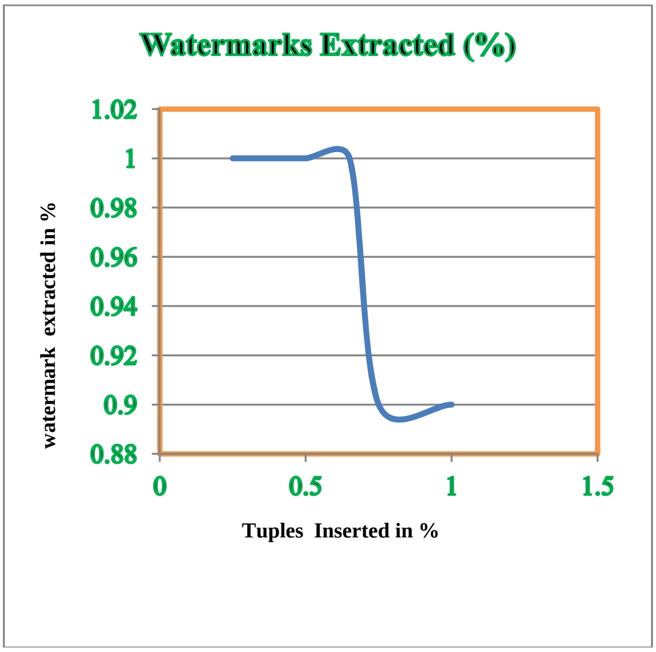

Fig. 3 Resilience to insertion attack

Fig. 3 shows that how the system designed is resilient to the insertion attack where x-axis shows number of tuples inserted against watermark extraction percent on y-axis.

Applications:

Applications of watermarking are following:

1. Monitoring the Broadcast: Identifying when and where your data is broadcasted by decoding watermarks embedded in them.
2. Determination of Ownership: Embedding signature of yourself as your copyright
3. Evidence of Ownership: Using watermarks to provide evidence in ownership disputes.
4. Tracking Redistribution: We can use watermarks to identify who has redistributed our data without permission
5. Identifying the Tampering: With the checksum evaluation we can indentify tampering in the original dataset.

IV. Conclusion and future scope

In this paper, we have presented a resilient watermarking technique for database using watermarking by generating the signature. This system uses the checksum evaluation that checks the signature generated using the private key of the recipient. This system does not affect the performance and the usefulness of the database and is resilient to attacks of insertion, alteration and deletion. We have used SHA 512 for signature generation and the checksum evaluation process algorithm and RSA algorithm for watermarking and the decoding process. This has minimized the errors from the existing systems. We have compared our technique [3] and other with previous approaches and shown the superiority of our technique with respect to all types of attacks. This proves that our technique is secure and efficient than existing systems. We have tested the system against numerous attacks numerous times. Moreover we do not modify any tuple data for embedding watermark so it is more reliable secure and the usefulness of the system is not tampered

Future Scope:

In this paper, watermarking technique for all the data types is shown. Our system is resilient to the attacks made in the databases but in future it can be tested for external and internal attacks from the front end also.

REFERENCES

[1] National Institute of Standards and Technology, Fips 180, Federal Information Processing Standards,Secure Hash Standard (SHS), April 1993.

[2] Shehab Mohamed, Elisa Bertino, and Arif Ghafoor, “Watermarking Relational Databases Using Optimization-Based Techniques”, IEEE Transactions On Knowledge And Data Engineering, Vol. 20, No. 1, January 2008.

[3] R. Sion, M. Atallah, and S. Prabhakar. “Rights Protection for Relational Data”. IEEE Transactions on Knowledge and Data Engineering, 16(6), June 2004.