Database Security using Watermarking Technique

Pooja Shelar[1], Priyanka Patil[2], Amrita Khamkar[3], Anushri Patil[4], Trupti Malavi[5], Nilam Parabkar[6]

[123456]Shivaji University, Department of Computer Science and Engineering, Sanjay Ghodawat Institutes, Kolhapur, Maharashtra, India

---***---

Abstract - *Databases mostly contains critical information. In today's internet-based application environment, ownership rights protection on relational database is decisive issue because unauthorized changes to data may have serious consequences and results of significant losses for the organization. Ownership protection on relational databases that is shared with having receiver desires to develop a watermarking technique that must be robust against different types of attacks and it should continue to have the knowledge in the databases in order to make them effective for knowledge-aware decision support systems. It is desirable that a database owner need not define usability constraints for every application and for every recipient distinctly. Most of the study available in this field is focused on images, audio, video etc. However, with the requirement of relational database security solution, this paper presents the watermarking technique on relational data with certain constraints and analyzes their strengths and weaknesses.*

***Key Words*: Relational Database System; Copywrite Protection; Watermarking; Database attacks.**

1.INTRODUCTION

A watermark is recognizable image or pattern in paper that appears as various shades of lightness/darkness when viewed by transmitted light, caused by thickness variations in the paper. A watermark stored in a data file refers to a method for ensuring data integrity which collaborate aspects of data hashing and digital watermarking. Both are useful for tamper detection, though each has its own benefits and disadvantages. Digital watermarking is the process of possibly once and for all embedding information into a digital signal. The signal may be audio, pictures or video, for example. If the signal is copied, then the information is also passed in the copy. In this paper watermark is applied to any relational database having attributes which changes in their values will not affect the applications. Databases have different types of attributes like, numeric, text, etc. The proposed scheme applied for non-numeric attributes to preserve the value of attributes in database. Generally, database watermarking techniques consists of two main phases: watermark insertion and watermark detection as shown in figure 1.

This paper is organized as follow: in section 2; related work is presented, in section 3; outline and flow charts of proposed technique is described in details, in section 4; some important features of the proposed scheme watermark are discussed, in section 5; analysis and the performance of the proposed system are estimated with reference to different attacks and in Section 6; conclusion and future work.

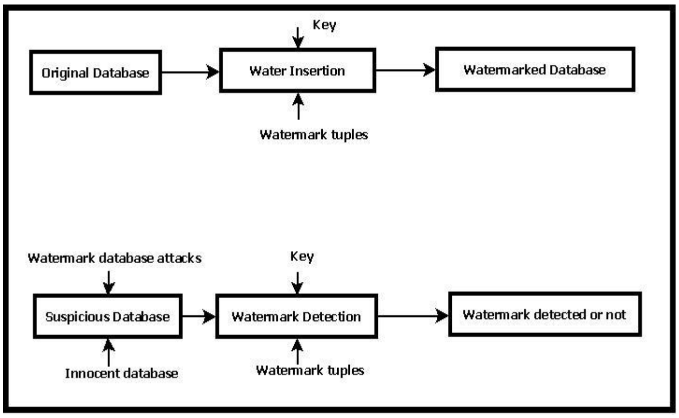


Fig 1: Watermark insertion and Detection

2. RELATED WROK

Set Bedi et al., proposed watermarking technique used for data authentication and integrity of relational database. For integrity confirmation of tables in the database, the watermark is depending on a secret key and the original copy of relation. This method used the concept of Eigen values to generate the watermark for a record in tuple. Watermark embedding is done by using Eigen values in a non-numeric attribute of a tuple. Detection of the watermark proves authenticate and integrities of data.

Pramod et al., presented new scheme for watermarking non-numeric relational databases. This technique uses voice of a copyright holder to create watermark, then corresponding insertion and detection algorithm had been applied.

Rajneesh et al., proposed a secure method which uses both semantic and syntactic techniques to watermark the tuple in a relation. The Watermarking technique is reliant on secret key and on the relation. The proposed algorithm is based on the idea of predefined signals of ASCII characters. A secret

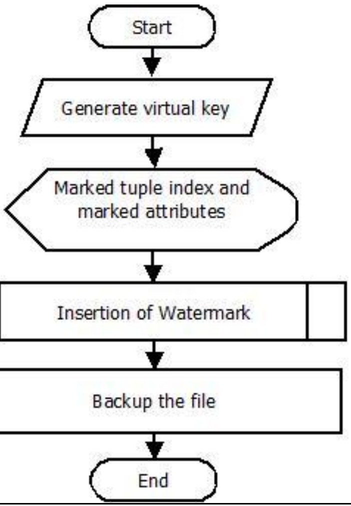
key is generated by using these signals only. To embed a watermark they used the concept of contractions for words and also one of syntactic approach.
Irjet Template sample paragraph .Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

3. PROPOSED SYSTEM

Proposed watermarking techniques for text exploit the special properties of formatted text. Watermark is introduced by altering the spacing between words and lines of text. The main goal is to design a technique which has less transfer time, fully blind, robustness, and reliability. Below are described two stages of watermark technique.

3.1 Watermark insertion stage

The following flowchart shows insertion stage as shown in figure 2.



Notation	Description
R	Relation to be marked
r	record of a relation
K	Secret Key known only to owner
r.p	Primary Key
Vp	virtual primary key
V	Number of attributes in the relation available for marking
Y	Used to determine the number of tuples to be marked. If w denotes number of tuples to be marked, then w=n/y
1/y	Fraction of tuples marked
r.Ai	Attribute Value
Totalcount	Number of tuples are tested
Matchcount	Number of tuples contain the expected value
A	Significance level of the test for detecting a watermark
Ґ	Minimum number of correctly marked tuples needed for detection
S	Suspected relational database
Attribute marked i	Index of selected attribute that will be marked
N	Number of tuples in relation

Fig 2: Watermark insertion

In this stage, the watermark is embedded as the space between the nonnumeric attributes to preserving the query results.

The watermark insertion stage composed of the following steps:

Step 1: Generate virtual key

To get the virtual key the following equation has been used
hash1 (r.p, Қ) = H (Қ &H(r.p& Қ)) (1)

Where:

- hash1 is defined as a message authentication code which is a one-way hash function H that depends on a key. It operates on an input message (M) of arbitrary length and returns a fixed length hash value h= H (M).
- & represents concatenation.

Step 2 : Generate marked tuples index and marked attribute

The algorithm generates marked tuple index for insert watermark in its selected attributes by using the following equation
tuple marked =Vp mod У (2)

Where mod operation had been applied on Vp and y then, marked tuple had been checked if it is equal to zero or not. If it is not equal to zero anther tuple will be chosen and then mod operation will be applied again. If it is equal to zero, the following equation will be applied to get marked attribute.

Attribute marked i = Vp mod ν (3)

Where mod operation had been applied on vp and v to get index marked attribute.

Step 3: Insertion watermark

In this step the algorithm gets the value of attribute after insert watermarking by using the following function

r.Ai=EmbedWm (r.Ai,tuple marked, attribute marked, Қ) (4)

Where : "EmbedWm" is insertion watermark function

This function has two main rules.

a) Generate watermark the algorithm uses the following equation to generate watermark

hash2(V, Қ) = H(Қ&H(H(Қ) ¤ H(V)) (5)

Where:hash2 a one-way hash function represents XOR operation. V is the variable. In The proposed system "V" is considered as marked tuple to get "T1". And it is considered as marked attribute to get "T2",

Where:

T1= hash2 (tuple marked, Қ) (6)

T2= hash2 (attribute marked, Қ) (7)

Watermark "T" is the value of 160 digit generating by XOR operation between T1 and T2.

b) Change case of selected attribute depending on value of 1th bit position of T.

Step 4: Backup the file.

3.2 Watermark Detection Stage

The following flowchart illustrates detection stage as shown in figure 3.

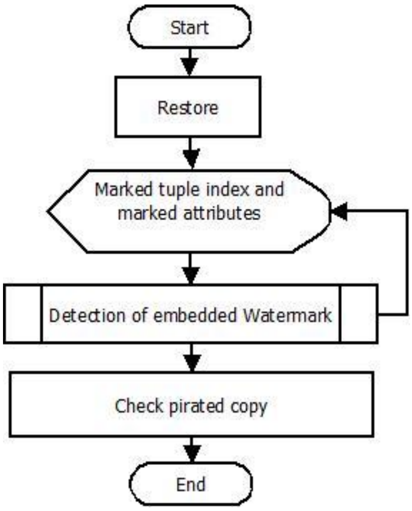

Start
Restore
Marked tuple index and marked attributes
Detection of embedded Watermark
Check pirated copy
End

Fig 3: Watermark Detection

The watermark detection algorithm has the following steps:

Step 1: Restore Get the original database.

Step 2: Generate marked tuples index and marked attribute.

The algorithm gets selected attribute as explained in insertion stage then increase (totalcount) by one.

Step 3: Detection the embedded watermark

This operation is done by a function called "WmDetection"

WM detection (τ,tuple marked, attribute marked, k) return number (8)

This function has two main steps:

a)Generate watermark as explained in insertion stage.

b)Compare attributes case: the function "WmDetection" compares the current attribute case with the case that must have been set for that attribute by the watermark insertion algorithm. If it returns "1" then "matchcount" will increase by one, as shown below: matchcount = matchcount + WmDetection (s.Ai, tuple marked,attribute marked, Қ) (9)

Step 4: repeat the previous steps for each tuple in relational database.

Step 5: check pirated copy

In this step, the algorithm checks if this database is pirated or not by getting the result of threshold function as follow:

Ґ=threshold (totalcount, α) (10)

Then compare (matchcount) with the minimum count returned by the threshold function for the test to succeed at the chosen level of significance α.

4. CHARACTERISTICS OF PROPOSED SYSTEM

4.1 Security

The secret key K only known by the owner has been used. Selection of tuples and attributes is based on K. Also, using a secure one-way hash function (SHA-1) for selecting marked attribute and generating watermark makes this scheme more secure.

4.2 Blind Detection

The proposed technique is fully blind since watermark detection should neither require the knowledge of the original database tables nor the watermark itself.

4.3 Query preserving watermark

Watermark embedding is done by changing the case of selected attribute according to algorithmic rules, which does not change the value of attribute, so the result of the queries will not be changed after embedding.

5. ANALYSIS

The detectability of a watermark in the proposed system depends on the significance level, the number of marked tuples, the number of tuples in the relation and the fraction of tuples marked.

5.1 Robustness

Watermarks should be robust against attacks to erase them. Robustness is a very important issue as it proves ownership of the data. Users have to sense their watermark in data without any damage or defect, i.e. the rate of correctly detection of watermark is also called robustness of the watermark. Below subsections will explain the advantages and disadvantages of the proposed system against numerous forms of malicious attacks.

a. *Advantage of proposed system*
   1. Preventing Subset Selection attack.
   2. Preventing Deletion attack.

b. *Disadvantage of proposed system*
   1. Subset insertion attack

In this attack, an attacker might insert many tuples in the original database. The system can be executed with various percentages of insertion. These experiments run on N=13270, fraction =10, v=3 and α =0.01. The system had been run twice at each percentage then the results each time are recorded, then the maximum percentage of detection had been taken.

   2. Modification attack

In this type of attack, the attacker changes the tuples of the database randomly. Attacker hopes by doing so to erase the watermark from the database. The system is executed with various percentages of modification twice the results are recorded each time then take maximum percentage of detection.

5. CONCLUSION AND FUTURE WORK

In this paper, we reviewed the technique on watermarking relational databases that embeds the watermark tuples in the database. Every author worked for the robustness of the technique. Many watermarking techniques are based on different watermark information; many of these techniques are designed for numerical database and are distortion based. Finally, we observe that usability of the watermarked database and deterministic detectability leaves so many queries in mind for future research. Many of these techniques used a single attribute of a tuple to embed a watermark. So, this work will be extended towards embedding the same watermark at different attributes at different places. Therefore, it will be hard for attacker to remove watermarks from different places from the database. The proposed technique is depend on presence of primary key. So we will also extend the work to find solution if there is no primary key.

REFERENCES

[1] Nahla EI_Haggar, Mahmoud M. EI-khouly, Samah S. Abu EI Alla, "A New Technique for Relational Database Protection, Faculty of Computers & Information Helwan University.

[2] Rakesh Agrawal Jerry kieman, "Watermarking Relational Databases", IBM Almaden Research.

[3] Rajneesh Kaur,Bedi,Purva,Gujarathi,PoonamGundecha,shishKulkarni, "A Unique Approach for Watermarking Non-numeric Relational Database", International Journal of Computer Applications (0975 – 8887) Volume 36– No.7, December 2011.

[4] Bedi R., Thengade A.,Wadhai V., "A New Watermarking Approach for Non Numeric Relational Database", 2011.

[5] Pramod a. kharade,vikramsinh m. pokale, vijay d. chougule and vishal v. mangave,"speech based watermarking for non-numeric relational database" , international journal of innovative research in engineering &science,april 2013.

[6] Snehal S. Kshatriya, Prof. Dr. S. S. Sane, "A Study of Watermarking Relational Databases", International Journal of Application or Innovation in Engineering & Management (IJAIEM)Volume 3, Issue 10, October 2014.

[7] Brijesh B. Mehta," A Novel approach as multiplace watermarking for security in database", Dept. of Computer Engineering, S. V. National Institute of Technology, Surat, Gujarat, INDIA-, 2010-2011.

[8] Burepalli V.S. Rao, Munaga V. N. K. Prasad,"Subset selection approach for watermarking relational databases", Institute od development and research in Banking Technology, Hyderabad, India

[9] Hazem El-Bakry, Mohamed hamada, "A novel Watermark Technique for relational Databases", Mannsura Univ., Egypt.

[10] Ali Al-Haj, Ashraf Odeh, and ShadiMasadeh, "Copyright Protection of Relation.