# Reversible Watermarking on Database Images using Difference Expansion Method

Ajay Shelar
B.E. Student of Computer Engineering
Atharva College of Engineering,
Mumbai University
Mumbai, MH, India

Abhishek Nanarkar
B.E. Student of Computer Engineering
Atharva College of Engineering,
Mumbai University
Mumbai, MH, India

Sushil Raipelly
B.E. Student of Computer Engineering
Atharva College of Engineering,
Mumbai University
Mumbai, MH, India

Nida Parkar
Prof. of Computer Engineering
Atharva College of Engineering,
Mumbai University
Mumbai, MH, India

***Abstract*— Proving ownership rights on outsourced relational databases is a crucial issue in today internet-based application environments and in many content distribution applications. In the cancerology domain, we were brought to make periodic mammography images to monitor tumour patients. MS SQL Database Management system (DBMS) is a solution to manage these images with patient's data recorder. Knowing the large size of medical images of mammograms, the MS SQL DBMS saves these images outside the MS SQL database using external LOBs. The link between these images and MS SQL is done through the BFILE. At this level, two problems are raised: the first problem is that access to these images can become impossible because the link is likely to be broken. The second problem is security, the fact that the images are saved outside the MS SQL database, they do not benefit from its powerful security. The protection of the integrity and confidentiality of data and patient images are a necessity defended by laws and they must be preserved against any unauthorized access, alteration or destruction. In this paper, we propose the method of reversible watermarking technique based on the difference expansion to resolve these two problems and explore its use in search and retrieval strategy of images**

***Keywords*— SQL; LOBs; Watermaking; BFILE; Difference Expansion.***

## I. INTRODUCTION

As far as the cancer treatment domain is concerned, doctors are likely to make a large number mammography images to monitor tumor in patients. For the management of these images multiple databases images are developed in terms of the size and variety. In recent years, various search engines based on Content Based Image Retrieval (CBIR) have been developed by different research teams. Generally, these systems use CBIR retrieval algorithms based on some vector descriptors of low-level features such as pixel, texture, color and shape. The computing time of such algorithms is important to extract an image that is similar to the query image. The reason is the need to calculate the vectors of feature descriptors of all the images in the database and compare them to the vector descriptor characteristics of the required query image so as to keep only the closest matching one. Management system oracle database (DBMS) helps with the solution ORD Image signature by doing the calculating and recording the characteristics of the describing vector (signatures) for every image in the database. On one hand, the parameters of the describing vector are not suited for patient's medical images. On the other hand, given the large size of these images of mammograms, the MS SQL DBMS stores these images outside of the MS SQL database. The link between these medical images and LOBs in MS SQL is done via a pointer BFILE. At this stage, there are two problems: the first one is the possibility of relationship breakage between LOBs and images. The second problem is that the images that are placed outside the MS SQL database cannot benefit from its powerful security. Therefore one of the best security measures that can be used is the watermarking. The latter is an important area of research for the security, confidentiality and integrity of data. Watermarking can be done by hiding the electronic patient's medical record in its images. Hence, the image and patient data become a single entity. However, in medical diagnostic, quality is very critical and cannot be compromised. Therefore, any change in the content of the image is not allowed. For this reason, we adopted the method of reversible watermarking technique using difference expansion.

## II. LITERATURE REVIEW

### *1. Watermarking Relational Database:*

The proposed algorithm for database watermarking based on primary key and private(secret) key. It consists of inserting a single bit watermark into the numeric field of database and then detecting it with the use of detection algorithm. Generally, in all database watermarking techniques we assume that Database relations that can be watermarked contain attributes which are such that changes in algorithm: Ŋ - Number of relations in the tuple υ - Number of attributes available in the relation for marking ζ – Number of least significant bits a few values do not affect the application[1].

Following are the notation used in the available for marking in an attribute 1/ϒ – Fraction of tuples marked ϑ – Number of tuples marked α - Significance level of test for detecting a watermark ζ – Minimum number of correctly marked tuples required for detection Here, watermarking database relation R where the scheme is R (P, A0,....., Av-1), where P is primary key attribute. This algorithm uses Message Authentication Code.

*Message Authentication Code*

MAC is a type of one-way hash function that depends on the key. One way hash function has some characteristics like,

- With given message M, it's easy to calculate hash value ħ
- With given ħ, it is difficult to compute M such that ĦJ(M) = ħ
- Given M, it is tough to find M‟ such that ĦJ(M) =ĦJ(M‟)

Here Ƒ be a MAC that randomizes values of the primary key attribute r.P of tuple r and returns an integer value in a wide range.

- Ƒ is seeded with a private key κ known only to owner.
- So Ƒ(r.P) = ĦJ(κ ○ ĦJ(κ ○ r.P)), where ○ represents concatenation.

*Watermark Insertion*

// the private key κ is known only to owner of the database.

//parameters ϒ, υ, and ζ are private to the owner.

- Foreach tuple r ϵ R do
  - if ( Ƒ(r.P)mod ϒ equals 0) then //mark this tuple
  - attribute_indexi= Ƒ(r.P)mod υ //mark attribute Ai
  - bit_index j = Ƒ(r.P) mod ζ //mark jth bit
  - r.Ai = mark(r.P, r.Ai,j)
  - mark(primary_keypk, number υ, bit_index j) return number
  - first_hash = ĦJ(κ ○ pk)
  - if(first_hash is even) then
  - set jth least significant bit of υ to 0
  - else
  - set jth least significant bit of υ to 1
  - return υ

Line 2 determines if the tuple under consideration is marked. Due to use of MAC, only the owner who has the knowledge of the private key κ is able to determine which tuples have been marked[4][5][7]. For any selected tuple, line 3 determines the attribute that will be marked amongst the υ candidate attributes. For a selected attribute, line 4 determines the bit position amongst ζ least significant bits that are marked. The results of the tests in lines 3 and 4 depend on the private key of the owner. For erasing a watermark, here, the attacker will need to guess not only the tuples, but also the marked attribute within a tuple along with the bit position.

The mark subroutine sets the selected bit to 0 or 1 depending on the hash value obtained in line 7. Thus, the result of line 9 (line 11) either keeps the attribute value unchanged or decrements (increments) the value. Consequently, marking decrements some values of an attribute while it increments some and leaves some unchanged. Databases generally allow attributes to assume null values. If a null attribute value is if found while marking a tuple, this method does not apply the mark to the null value, hence leaving it unchanged.

*Watermark Detection*

// κ, ϒ, υ, and ζ have the same values used for watermark insertion. // α is test significance level that the detector preselects.

- total count=match count=0
  - for each tuple s ϵ S do
  - if ( Ƒ(s.P)mod ϒ equals 0) then //this tuple marked
  - attribute_indexi= Ƒ(s.P)mod υ //attribute Ai marked
  - bit_index j = Ƒ(s.P) mod ζ //jth bit was marked
  - totalcount=totalcount+1
  - match count=match count + match (s.P, s.Ai,j)
  - ζ = threshold (totalcount, α)
  - if (matchcount>= ζ) then suspect piracy
  - match(primary_keypk, number υ, bit_index j) return int
  - first_hash = ĦJ(κ ○ pk)
  - if(first_hash is even) then
  - return 1 ifjth least significant bit of υ is 0 else return 0
  - **else**
  - return 1 if jth least significant bit of υ is 1 else return 0

Assume, Ramesh suspects that the relation S published by Suresh has been pirated from his relation R. Here set of tuples and attributes inside S can be a subset of R. Here assumption is that Suresh does not drop the primary key attribute or change the value of primary keys since the primary key contains valuable information and changing it will render the database less useful from the users point of view. The watermark detection algorithm is probabilistic in nature [6]. Line 3 determines that if the tuple S under consideration must have been marked at the time of insertion of the watermark. Lines 4 and 5 determine the attribute and the bit position that should been marked. The subroutine match then compares the current bit value with the value that must have been set for that bit by the watermarking algorithm.

So at Line 8 how many tuples were tested (total-count) and how many of them contain the expected bit value (match-count). In a probabilistic framework, only certain minimum number of tuples has to contain matching marked bits. The match-count is compared with the minimum count returned by the threshold function for the test to succeed at the chosen level of significance α.

## III. METHODOLOGY

Using C# as the language and the web development platform being ASP.NET, team is using .NET framework as programming model for building application on windows client and server. And using MS SQL server a relational database management system to stored images of users. For security purpose we use watermarking technique. This project works on a layer basis throughout the system. Firstly there is an interface layer which is a webpage for displaying mammographic images where the patients and doctor can query and retrieve data. Behind the presentation layer is the main and most important part of the project i.e. reversible watermarked images database retrieval engine that will do the work of providing fast and secured access to the images and reports of the patient's history.

As the user or the doctor inputs a patient-id or any watermarked mammographic images to the interface, the input query will be searched through the database and all the information and images related will be retrieved.

## IV. FUTURE SCOPE

The main goal of database with watermarking is to develop an efficient, robust and impersistant watermark for security in database. Therefore, here we are studying different methods developed by many authors and trying to find advantages and disadvantages of all. Finally we have come with a new method of database watermarking which is more robust with respect to other methods developed by other researchers in past. Our method can be applied to numerical as well as categorical databases.

## V. CONCLUSION

In this paper, we use multipurpose watermarking composed with following parameters. The objective is to enhance the safety management system of the patient's data recorder and medical images using MS SQL DBMS. The advantage of this system resides in three aspects. (A)The system begins to check the image for authentication. (B) It allows the search and retrieval of images based on their characteristic features embedded in images. (C)The fact that the images are stored outside of the SQL database DBMS, the system can restore any broken links between these images and LOBs in MS SQL based on the BFILE extracted from the watermarked image.

## REFERENCES

[1] Rakesh Agrawal, Peter Haas and Jerry Kiernan: " Watermarking Relational Data: Framework, Algorithms and Analysis" The VLDB Journal, 12(2), August 2003.
[2] D. Abraham Chandy and J. Stanly Johnson. Selection of Haralick Texture Features using an Evolutionary Algorithm for Content-based Mammogram Retrieval.17-24 European Journal of Scientific Research Volume 86 Issue 1 September, 2012
[3] Florea FI, Rogozan A, Bensrhair A and Darmoni SJ. Medical image retrieval by content and keyword in a on-line health-catalogue context, Proc. Mirage 2005: 229-36.
[4] J. J. K. Ó Ruanaidh, W.J. Dowling and F.M. Boland, "Phase watermarking of digital images," Proceedings of the
[5] International Conference on Image Processing, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 239-242.
[6] J. Tian, "High capacity reversible data embedding and content authentication," in IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. III–517–20, Hong Kong, Apr. 2003.
[7] J. Tian, "Reversible data embedding using a difference expansion," IEEE Transactions on Circuits Systems and Video Technology, vol. 13, no. 8, pp. 890–896, Aug. 2003.
[8] Le ThiLan (Marteren 2004). Indexation etrecherche d' images par le contenu, 114 pages.
[9] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Reversible data hiding," in Proceedings of the International Conference on Image Processing, pp. 157–160, NY, USA, Sept. 2002.