An Effective Watermarking Technique for Relational Data

Sujith A.V , KCG College of Technology, IVth year CSE , sujithavps94@gmail.com

Suresh G, KCG College of Technology, IV th year CSE, sureshg8691@gmail.com

Yohith R.R, KCG College of Technology, IVth year CSE, yohithrajendran@gmail.com

Mrs. Dhanya Anand, Assistant Professor,KCG College of Technology ,anand.dhanya@gmail.com

***Abstract*–Improvement in the IT industry is playing an important role in the use of information systems constitutes relational databases. These databases are utilized effectively in collaborative environments for data extraction; sequentially, they are prone to security threats concerning ownership rights and information tampering. Watermarking is employed here to enforce ownership rights over the shared relational data and for providing a means for analyzing and tracking data tampering. Here the ownership rights are employed using reversible watermarking technique; the stored data undergoes certain alterations such that the data quality is preserved. Reversible watermarking technique is used to ensure that both the quality of data and recovery is achieved. Even though, these technique is usually not vulnerable against data tampering attacks and no mechanisms is provided to watermark a selected or particular attribute by taking into account its role in knowledge discovery. Hence reversible watermarking is needed to ensure the data quality and recovery. The watermark encoding and decoding is achieved discovery of its knowledge (i), (ii) raw data is recovered in spite of the data tampering attacks. In this, a reversible watermarking technique for numerical relational data has been postponed that addresses the above objectives. Experimental studies prove the effectiveness of Reversible watermarking technique against data tampering attacks and show that the proposed technique dominates the existing ones.**

Index terms-Reversible watermarking, genetic algorithm, data recovery, data quality, robustness, and numerical data

I. INTRODUCTION

Watermarking techniques have been used in earlier ages for the protection in terms of ownership rights and corrupt proofing for a different of formats of data. This contains files of videos, audios and pictures, and relational databases and more. Watermarking techniques can give recovery of data along with ownership security. Fin-reprinting, hashing of data, classification codes are the usual techniques for protecting the owners. .Fingerprints also called watermarks and are used to examine and monitor digital rights of owners by watermarking all the data with variety of watermarks for variety of users. Initially this type of watermarking of digital data tries to inspect the roots of data leaks by tracing an intruder. In hash functions, contents of digital data can be protected by doing a hash function technique thereby the contents and records is not altered. If the hash of the raw and corrupted data is the similar, authentication of data can be proved but ownership cannot be verified as east as possible. Classification or the serial codes are used for reducing the unnecessary contents over the World Wide Web and are mainly implacable to pictures, videos and audios. Watermarking has the ability that it can give ownership security over the content of digital data by embedding the data with a watermark distinct to the owners. The encoded watermark can be used afterwards.

This Proposed system which we implement is a different approach to generate the watermark bits from UCT (Coordinated Universal Time) TIME AND DATE which is the major standard time used to synchronize the world's time. An effective watermark algorithm is applied to encode watermark bits into relational data of Owner of the database. The watermark encoding algorithm takes a private key and the watermark containing bits as an input and establishes a data set into watermarked data content. A MD5 which is hash function algorithm is implied on the data selected for only those tuples which has an even hash value. The Process of Watermark embedding includes decoding and encoding sections. The watermark encoding section has Data partitioning, Tuple selection for watermarking process and finally watermark embedding process. Decoding section contributes of decoding the watermarked data.

A MD5 which is cryptographic hash function is implied on the DATASET and to encode even hash value tuples. This section has two motives: 1) it improvises the watermark security by hiding the identity of the watermarked tuples from an intruder; and 2) it reduces the number of to-be-watermarked tuples to limit distortions in the data set .If the Hash Value Computation Is Satisfied Select the tuples for Watermarking bits from selected tuples for Encoding process

III PROPOSED SYSTEM

Authenticity is achieved using the watermarking technique. Initially used techniques for not strong against the attacks on the data so watermarking , without any exception is used for securing the ownership of numerous amount of data format which includes audio, videos, software and others like images, documents, geographic information system (GIS) related data and so on. It mainly concentrates on relational databases are used in different application domains. Nowadays, excellent data mining techniques are being used on data, extracted from relational databases, to detect intelligent patterns like hidden data formats that provide essential support to decision makers in making robust, accurate, and relevant decisions; as that of data sharing between its owners and database users.

The admin or the owner of the Relational Database encodes the watermarked data, such that it minimizes the distortions in the original relational data which are known by its constraints, to secure the important knowledge content of the data. So the algorithm that we propose encodes every bit of a data watermarked i.e. generated from the UCT date and time in evert row of number attribute with the goal of having maximum effectiveness even if an intruder is somehow able to tamper or corrupt the watermarked set of data.

II. RELATED WORK

In the paper " Content based Zero-Watermarking Algorithm for Authentication of Text Documents" published by the authors ZuneraJalil, Anwar M. Mirza,Maria Sabir2 we retrieved the following ideas. Copyright protection and authentication of digital contents has become a significant issue in the current digital epoch with efficient communication mediums such as internet. Plain text is the rampantly used medium used over the internet for information exchange and it is very crucial to verify the authenticity of information. There are very limited techniques available for plain text watermarking and authentication. This paper presents a novel zero-watermarking algorithm for authentication of plain text. The algorithm generates a watermark based on the text contents and this watermark can later be extracted using extraction algorithm to prove the authenticity of text document. Experimental results demonstrate the effectiveness of the algorithm against tampering attacks identifying watermark accuracy and distortion rate on 10 different text samples of varying length and attacks.

In this paper "Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing" by the authors Adnan M. Alattar and Osama M. Alattar**.** In this paper, we retrieved a new method for watermarking electronic text documents that contain justified paragraphs and irregular line spacing. The proposed method uses a spread-spectrum technique to combat the effects of irregular word or line spacing. It also uses a BCH (Bose-Chaudhuri-Hocquenghem) error coding technique to protect the payload from the noise resulting from the printing and scanning process. Watermark embedding in a justified paragraph is achieved by slightly increasing or decreasing the spaces between words according to the value of the corresponding watermark bit. Similarly, watermark embedding in a text document with variable line spacing is achieved by slightly increasing or decreasing the distance between any two adjacent lines according to the value of the watermark bit. Detecting the watermark is achieved by measuring the spaces between the words or the lines and correlating them with the spreading sequence. In this paper, we present an implementation of the proposed algorithm and discuss its simulation results.

In this paper "A Public-Key Authentication Watermarking For Binary Images by the authors Hae Yong Kim, Ricardo L. de Queiroz**."** We retrieved the following idea. Authentication watermarking is the process that inserts hidden data into an object (image) in order to detect any fraudulent alteration perpetrated by a malicious hacker. In the literature, quite a small number of secure authentication methods are available for binary images. This paper proposes a new secure authentication watermarking method for binary images. It can detect any visually significant alteration while maintaining good visual quality. As usual, the security of the algorithm lies on the secrecy of a private-key. Only its owner can insert the correct watermark while anyone may verify the authenticity through the corresponding public-key. A possible application of the proposed technique is in internet fax transmission, i.e. for legal authentication of documents routed outside the phone network.

IV EXPERIMENT

The following modules are identified in the system,

a. Data Partitioning

b. Tuple Selection for Watermarking

c. Watermark Encoding

d. Authentication via edge detection and watermark decoding

Data Group Partitioning

In the above mentioned module, the Relational data are partitioned into logical groups and this is known as watermark encoding phase. This process is carried out by the admin or the owner of the database. The partitioning process is carried out data-partitioning algorithm and it works in the following way.

$partition(t)=H(Ks\|H(t.Pk\|Ks))mod\ l$

Where t is each tuple, Pk is the primary key of the tuple t, H () is a MD5 hash function, || is the concatenation operator and Ks is a security key. The data groups or the partitions has been done after applying this algorithm. The group length is decided by the administrator or the owner.

Tuples Selection for Watermarking

A Tuple is known as the data unit in a Relational Database. It is row of an database. In this section inorder to watermarking the tuple is done. Threshold Value Computation is a process for computing each attribute. Threshold value is set for the tuple and if it is higher than the respected value it is taken for Watermark encoding Process. The selection of the data by threshold value computing process is given by

$$T=c^{*}\ Mean+\ S.D$$

c is the confident factor with a numeric value between 0 and 1. S.D is standard deviation. The confident factor c is always kept secret such that it is difficult for the intruder to analyse on which tuple the watermark inserted. After this, the tuples having above the computed threshold value will only be selected.

In this section, MD5 hash value computation is applied on the data set which are having the even hash value. This section comprises of two objectives: 1) it then improvises the watermark security by hiding the indication of the watermarked tuples from an attacker; and 2) it then reduction in the distortion of the data set by minimization. So from the hash value computation it is selected to the next step called watermark encoding process.

Watermark Encoding

The watermark generating function takes UCT generated time-date as an input like a stamp and then produces the watermark embedding bits b1b2 b3 . . . bn from this time-date stamp. These embedding bits are given as input to the watermark encoding process. To construct a watermarked data set, these bits are encoded in the original data by using an embedding algorithm. The algorithm proposed encodes every bit of a multibit watermark generated from time-date in each selected row. Thus the watermark bits are embedded into the tuples selected using a watermarking function. This technique encodes each bit of the watermark in every selected tuple of each group or the partition.

Authentication via edge detection and Watermark Decoding

Authentication via the edge detector is used as an alternative technique to text based solution. Instead of alphanumerical passwords the image is used. The main conspiracy here is that key-images from the challenge area is selected and then he/she will be authenticated since people are better at image recognition and memorizing it. During User Registration section, Owner or Administrator has to provide some pictures to the user. The user is supposed to choose the key-images for the verification section. That picture has to be saved in the the Server for that particular User. During Login section, Administrator has to convert the original image to a grey scale and then manipulated to Edge detecting image. The concept here is the user has a challenge area which contains tricky and key-images. The tricky images are randomly generated images by the administrator. Key-image will be the user chosen image. Usually authentication is quite easy; a legal user needs to identify key-image correctly from the challenge set and then he or she will be given permission.

Watermark Decoding process in the extraction of watermark from the data. The Watermarked data Content has to be decoded only by legal user to give the database a proper ownership. If both the admin generated and user's proper ownership is same then the decoding phase is done. Otherwise it is rejected.

V IMPLEMENTATION

To implement the proposed application various algorithms being used which is described in the following sections. Message Digest 5 is known as cryptographic hash function algorithm for hash function in data group partitioning process and canny test (edge detector) is used as an algorithm for authentication.

MESSAGE DIGEST 5

STEP 1: The main MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted A, B, C, and D. These are initialized to certain fixed constants

STEP 2: The main algorithm then uses each 512-bit message block in turn to modify the state. The processing of a message block consists of four similar stages, termed rounds

STEP 3: Each round is composed of 16 similar operations based on a non-linear function F, modular addition, and left rotation. Figure 1 illustrates one operation within a round. There are four possible functions F; a different one is used in each round

STEP 4: One MD5 operation. MD5 consists of 64 of these operations, grouped in four rounds of 16 operations. F is a nonlinear function; one function is used in each round. Mi denotes a 32-bit block of the message input, and Ki denotes a 32-bit constant, different for each operation. left shifts denotes a left bit rotation by s places; s varies for each operation. Addition denotes addition modulo 232.

EDGE DETECTION AUTHENTICATION

STEP 1: Apply Gaussian filter to smooth the image in order to remove the noise. It can be determined using the equation for a Gaussian filter kernel of size (2k+1)*(2k+1) is given by:

$$\mathbf{H_{ij} = 1/2\pi\sigma^2 \exp\left(- \frac{(i-(k+1))^2+(j-(k+1))^2}{2\square^2}\right) i, j = 1....(2k+1)}$$

STEP 2: Find the intensity gradients of the image. The edge detection operator returns a value for the first derivative in the horizontal direction (G_x) and the vertical direction (G_y)

$\theta = \mathrm{atan2}(G_y, G_x)$ **G=** $\sqrt{G_x^{2} + G_y^{2}}$

STEP 3: Apply non-maximum suppression technique. There are still some edge pixels at this point caused by noise and color variation. To get rid of the spurious responses from these bothering factors, it is essential to filter out the edge pixel with the weak gradient value and preserve the edge with the high gradient value.

STEP 4: Apply double threshold technique to determine potential edges

STEP 5: Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges. To track the edge connection, blob analysis is applied by looking at a weak edge pixel and its 8-connected neighborhood pixels.

```
//create the detector

CannyEdgeDetector detector = new CannyEdgeDetector();

//adjust its parameters as desired

detector.setLowThreshold(0.5f);

detector.setHighThreshold(1f);

//apply it to an image

detector.setSourceImage(frame);

detector.process();

BufferedImage edges = detector.getEdgesImage();
```

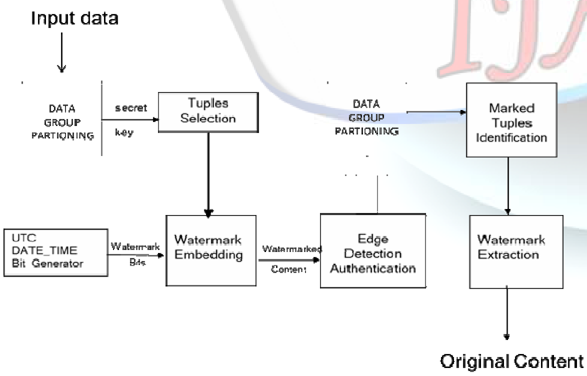

Figure 1.1 Architecture Diagram

EXPLANATION:

Reversible watermarking technique is based on securing the relational database rights of the owners. This helps in protecting the data from data tampering attacks by the intruders. Patterns by Watermarking process are quite tough to spot them and this became popular usage all over the countries for securing the relational data. This technique provides the effectiveness against the attack and reduction in data distortion. In this paper, data partitioning is done initially from the database records. Then tuples selection for watermarking is done. Tuple selection includes computation of threshold value such that stronger tuples for watermark embedding process is done and weaker tuples are rejected. Before embedding process all the records are converted to binary values. Then watermark encoding process is done on binary converted data by generating by the UCT time-date bit generator.

Next, edge detector algorithm is used for user's authentication instead of general password. Then after the login phase the user have to prove the ownership property rights by the time-date passcode. Once the UCT date-time is entered for specific user then the watermark decoding is done and the data is decoded for user's access.

VI CONCLUSION

Thus the project maintains the ownership of Relational Database and also minimizing distortion in the watermarked content proves the effectiveness of this technique against malicious attacks.

By applying the edge detection algorithm the user's password is kept secret. By applying the hash function the most records are protected from the malicious attacks. Thus it provides a security to the users content. Since the watermark embedding technique is reversible it minimizes the distortion of the data and also provides the record secure way of access.

V REFERENCE

[1] P. W. Wong, "A public key watermark for image verification and authentication," in Image Processing, 1998. ICIP 98.Proceedings.1998 International Conference on, vol. 1. IEEE, 1998, pp. 455–459.

[2] P. W. Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," Image Processing, IEEE Transactions on, vol. 10, no. 10,pp. 1593–1601, 2001.

[3] F. A. Petitcolas, “Watermarking schemes evaluation,” Signal Processing Magazine, IEEE, vol. 17, no. 5, pp. 58– 64, 2000.

[4] R. Agrawal and J. Kiernan, “Watermarking relational databases,”in Proceedings of the 28th international conference on Very Large DataBases. VLDB Endowment, 2002, pp. 155–166.

[5] R. Sion, M. Atallah, and S. Prabhakar, “Rights protection for categoricaldata,” Knowledge and Data Engineering, IEEE Transactionson, vol. 17, no. 7, pp. 912–
926, 2005.

[6] S. Subramanya and B. K. Yi, “Digital rights management,” Potentials,IEEE, vol. 25, no. 2, pp. 31–34,
2006.

[7] Christo Ananth, H. Anusuya Baby, “S-Box using AES Technique”, International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 3, March – 2014, pp 285-290

[8] K. E. Parsopoulos and M. N. Vrahatis, “Particle swarm optimization method for constrained optimization problems,” IntelligentTechnologies–Theory and Application: New Trends in Intelligent Technologies,vol. 76, pp. 214–220, 2002.

[9] R. Hassan, B. Cohanim, O. De Weck, and G. Venter, “A Comparison Of Particle Swarm Optimization And The Genetic Algorithm,”in46th AIAA/ASME/ASCE/AHS/ASC Structures, StructuralDynamics, and Materials Conference. American Institute of Aeronauticsand Astronautics, 2005, pp. 1–13.

[10] J. T. Brassil, S. Low, and N. F. Maxemchuk, “Copyright protectionfor the electronic distribution of text documents,” Proceedings ofthe IEEE, vol. 87, no. 7, pp. 1181–1196, 1999.