

Reversible Watermarking Technique Based On a Timestamp in Relational Database

M.Amudha¹ and Rajeswari.V²

¹Associate Professor,

Department of Computer Science and Engineering,

R.M.D Engineering College, Chennai

sruthiraji15@gmail.com, sruthiraji1215@gmail.com

²PG Scholar,

Department of Computer Science and Engineering,

R.M.D Engineering College, Chennai

January 5, 2018

Abstract

Watermarking method is to recognizable pattern used to identify authenticity. Intentionally introduced pattern in the data is hard to find and destroy, robust against malicious attack. WATERMARKING, without any exception, has been used for ownership protection of a number of data formats images, video, audio, software, XML documents, geographic information system (GIS) related data, text documents, relational databases and so on that are used in different application domains. Recently, intelligent mining techniques are being used on data, extracted from relational databases, to detect interesting patterns (generally hidden in the data) that provide significant support to decision makers in making effective, accurate, and relevant decisions; as a result, sharing of data between its owners and legitimate users.

The owner of the Relational Database embeds the watermark data, the distortions in the original data are kept within certain limits, which are defined by the usability constraints, to preserve the knowledge contained in the data. The proposed algorithm embeds every bit of a multi bit watermark (generated from date-time) in each selected row (in a numeric attribute) with the objective of having maximum robustness even if an attacker is somehow able to successfully corrupt the watermark in some selected part of the data set.

Key Words :water marking, privacy, security,mining techniques .

1 INTRODUCTION

Most companies and government agencies have a dire need for protecting sensitive information. Data mining is special technical term related with the discovery of new and interesting pattern of data from large data sets. The extraction of hidden information from large databases is a new emerging technology having the huge potential for the help of companies to focus on the important information in the data warehouse. The data mining automates the system to search the relevant information from the databases of data warehouse of the given enterprise which maintains the data warehouse. Most of the companies already collect and refine massive quantities of data. One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals .

Another issue is that of data integrity. Clearly, data analysis can only be as good as the data that is being analyzed. A key implementation challenge is integrating conflicting or redundant data from different sources. For example, a bank may maintain credit cards accounts on several different databases. The addresses (or even the names) of a single cardholder may be different in each. Software must translate data from one system to another and select the address most recently entered.

A technical issue is whether it is better to set up a relational database structure or a multidimensional one. In a relational structure, data is stored in tables, permitting ad hoc queries. In a multidimensional structure, on the other hand, sets of cubes are arranged in arrays, with subsets created according to category. While multidimensional structures facilitate multidimensional data mining, relational structures thus far have performed better in client/server environments. And, with the explosion of the Internet, the world is becoming one big client/server environment.

Finally, there is the issue of cost. While system hardware costs have dropped dramatically within the past five years, data mining and data warehousing tend to be self-reinforcing. The more powerful the data mining queries, the greater the utility of the information being gleaned from the data, and the greater the pressure to increase the amount of data being collected and maintained, which increases the pressure for faster, more powerful data mining queries. This increases pressure for larger, faster systems, which are more expensive.

2 DATA MINING SECURITY TECHNIQUES

2.1 Solving Constrained Nonlinear optimization problems with Particle Swarm Optimization

Particle Swarm optimization algorithm is mainly used for constrained nonlinear optimization problems. In PSO, the potential solutions, called particles are flown through the problem space by learning from the current optimal particle and its own memory. PSO is started with a group of feasible solutions and feasibility function is used to check if the new explored solutions satisfy all the constraints. All particles keep only those feasible solutions in memory. PSO is an evolutionary computation technique. It is initialized with the population of random solutions. It searches for the optimum by updating generations, and Population evolution is based on previous generations. Eleven test cases were tested and showed that PSO is an efficient and general solutions to solve most nonlinear inequality

constraints.

It is known that different constraint-handling techniques provide different quality results for different constrained non-linear optimization problems. We see that the PSO algorithm is consistent in locating the area of the global optimum in CNOP cases. Due to the random origin of evolutionary algorithms, it is difficult to deal with the equation constraints.

Two versions of PSO can be used in real problems. The local PSO algorithm is preserved for more accurate results, while the global version is a little bit faster. Higher iteration number and higher population size can be tried if the result is not satisfactory. PSO is an efficient and general method to solve most constrained parameter optimization problems. The issue is that it represents only the first step in the investigation of solving constrained parameter optimization problems using particle swarm optimization. The ability of particle swarms to solve more complex constrained optimization problems will need to be proven.

2.2 A public key authentication watermarking for binary images

A watermark is a hidden signal added to images that can be detected or extracted later to make some assertion about the host image. Watermarking techniques can be classified as either robust or fragile. Robust watermarks cannot be easily removed and should resist common image manipulation procedures. They are useful for copyright and ownership assertion purposes. Fragile watermarks (or authentication watermarks) are easily corrupted by any image processing procedure. Watermarks for checking image integrity can be fragile because if the watermark is removed, the watermark detection algorithm will correctly report a corruption of the image.

Authentication Watermarking is the process that inserts hidden data into an object in order to detect any fraudulent alteration perpetrated by a malicious hacker. 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.

There are many authentication watermarking technique for continuous tone images. In a typical cryptography-based AWT an authentication signature(AS) is computed from whole image data and inserted into image itself. It is mainly used to check integrity. Use of this is to send faxes and documents over uncontrolled networks and the internet. Among this another method is data hiding into binary images. A technique called template ranking is being used. The image is being divided into small blocks. The neighborhood of each pixel is analyzed via template ranking where 512 33 binary templates are ranked. Insert one bit in each block to have even or odd number of white pixels to insert 0 or 1 respectively. As different blocks have different quantities it should be shuffled before embedding data. Another technique used is authenticating binary images. The idea of computing AS of whole binary image and inserting into same image fails because the insertion will modify image fingerprint. Divide image into two regions first (small) region where AS will be inserted and second (large) region from where AS will be completed.

A new technique has being proposed for cryptographically secure authentication watermarking for binary images. It can be used with secret- or public/private-key cipher. We also derived a variant of the algorithm that is able to withstand parity attacks. The proposed technique is suitable to watermark most binary images except dispersed-dot halftones. The method using a public/private-key can be used in trusted FAX machines, i.e. to electronically sign binary documents. Further research is necessary to adapt the method to scanned documents.

2.3 Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing

A new method for watermarking electronic text document that contain justified paragraphs and irregular line spacing. This 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. 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. Most companies and government agencies need to protect many sensitive information or data. Encryption, access restriction and locking documents behind firewalls are some techniques for protecting sensitive information. Encryption is mainly used for preventing an unauthorized person from viewing the content of the document. Once the document is decrypted for viewing using the secret key, the authorized person can save, copy, print, or transmit the unencrypted document anywhere he or she wants without any major difficulty. Restricting access of a document is like providing information to trustworthy individuals. Identifying the untrustworthy person is difficult task to be done. Firewalls are an effective means to ban casual outsiders from accessing an organizations network. Firewalls also make it very difficult for a savvy computer hacker to break in. firewalls cannot prevent an insider from copying a sensitive document into a disk or emailing it to an outsider using a third-party Internet service provider to avoid tracking . Solution for ensuring the security of sensitive documents cannot rely on a single technique. Fingerprinting is also a technique which is used for protecting sensitive information. Fingerprinting refers to embedding an ID in the document, which identifies the owner or the recipient of the document. Fingerprinting does not mean cryptographically hashing the document into a signature that can be used to identify the document uniquely. The embedded ID can be detected and decoded from a fingerprinted document whenever and wherever the document is encountered, even after DA/AD conversion (print and scan). This process can be achieved through the use of digital watermarking techniques. Image watermarking techniques can be easily applied to a text document, but they introduce a very noticeable white noise in the text document. This noise is due to the binary (black and white) nature of the text document and its large white background. Image watermarking techniques are applicable only to image bitmaps, which implies that the text

document must first be rasterized.

To avoid problems several watermarking techniques have been developed. A method to reduce the watermarking noise by restricting the changes due to the process to the text boundary. Edge patterns are first isolated then each edge pattern is replaced with one of two predefined patterns to indicate the value of the embedded bit. Detecting the watermark is achieved easily without referring to the original document by inspecting each of these edge patterns to decide if it corresponds to a zero or one. The first method is called line-shift coding, in which each even line is slightly shifted up or down according to the value of a specific bit. If the bit is one, the corresponding line is shifted up; otherwise, the line is shifted down. The odd lines are considered control lines, hence they remain unmoved. They are used as references for measuring and comparing the distances between lines during the decoding stage. Decoding is achieved by comparing the distances between the bases of the lines, or the distances between the centroids of the lines. The bases of the lines in the original document are normally uniformly spaced. However, the centroids are not necessarily uniformly spaced; therefore, the original document must be used in the detection process that uses centroids. The second watermarking method is called word-shift coding. In this method, each line is first divided into groups of words. Each group has a sufficient number of characters. Then, each even group is shifted to the left or the right according to the value of a specific bit. The odd groups are used as references for measuring and comparing the distances between the groups during the decoding stage. A correlation and centroid-based method have been suggested for detecting the watermark. Both of these methods require the use of the original document. The third method is called feature coding. In this method, certain text features are altered in a specific way to encode the zeros and ones of the payload. Watermark detection is achieved by comparing the original document with the watermarked document.

A text watermarking technique based on the average inter-word distance in each line. They readjust these distances and the length of each word so that the average inter-word distance in each line represents a sample of a sine wave of specific phase and frequency. The ID (payload) is represented by the phase and the frequency of the sine wave. A correlation detector is used to decode the payload

from the watermarked document without referring to the original document.

we presented a new algorithm for fingerprinting a sensitive document to identify its original recipient whenever the document is found. The algorithm is based on adjusting the distances between the words or the distances between the lines of the text. This algorithm is more suitable for centered, left- and right-aligned text, and justified text as well as text with irregular spacing between lines. It uses spread spectrum and BCH error correction techniques, and it does not need the original document for detection. The spread spectrum technique guards against noise due to paragraph justification or irregular line spacing. The BCH error correction technique guards against printing and scanning noise. Simulation results showed that the algorithm is resilient to some document reformatting, such as changing the typeface and changing the margins. Fine tuning and additional investigation of the algorithm is underway to detect the watermark from a printed document.

2.4 Watermarking Relational Databases

The need for watermarking database relations to deter their piracy, identify the unique characteristics of relational data, and provide desirable properties of a watermarking system for relational data. A watermark can be applied to any database relation having attributes. The tuples, attributes within a tuple, bit positions in an attribute, and specific bit values are all algorithmically determined under the control of a private key known only to the owner of the data. Protection of these assets is usually based upon the insertion of digital watermarks into the data. The watermarking software introduces small errors into the object being watermarked. These intentional errors are called marks and all the marks together constitute the watermark. The marks must not have a significant impact on the usefulness of the data and they should be placed in such a way that a malicious user cannot destroy them without making the data less useful. Thus, watermarking does not prevent copying, but it deters illegal copying by providing a means for establishing the original ownership of a redistributed copy.

A multimedia object consists of a large number of bits, with considerable redundancy. A database relation consists of tuples,

each of which represents a separate object. The watermark needs to be spread over these separate objects. Tuples of a relation on the other hand constitute a set and there is no implied ordering between them. Portions of a multimedia object cannot be dropped or replaced arbitrarily without causing perceptual changes in the object. However, the pirate of a relation can simply drop some tuples or substitute them with tuples from other relations. This technique marks only numeric attributes and assumes that the marked attributes are such that small changes in some of their values are acceptable. All of the numeric attributes of a relation need not be marked. The data owner is responsible for deciding which attributes are suitable for marking. A one-way hash function operates on an input message of arbitrary length and returns a fixed length hash value. A message authenticated code (MAC) is a one-way hash function that depends on a key. Because of the use of MAC, only the owner who has the knowledge of the private key K can easily determine which tuples have been marked. We assume that it 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. The various techniques for watermarking images along the following dimensions: i) the method for selecting pixels where the watermark message will be hidden; ii) the choice of workspace to perform the hiding operation; iii) the strategy for formatting the message; iv) the method for merging the message and cover; and v) the operation needed for extracting the message. Our technique uses a private key and the primary key to select the bit positions. We do the hiding in the original space. We do not have a fixed message. The patchwork algorithm chooses random pairs of points of an image in a spectral domain and increases the brightness a by 1 unit and decreases the brightness b by 1 unit. Random changes to relational data can potentially introduce large errors. A technique proposed for watermarking a sequence of numbers. The basic idea is to modify the numbers to force them to be quadratic residues or nonresidues modulo a secret prime, according to the parity of the next bit of a user-provided message. The watermark is repeated many times throughout the data. The advantage of this technique is that we do not require data to be ordered and hence our technique is robust in the presence of updates. We also do not have a fixed message that

is encrypted and repeated in the data.

Another issue is the identification of the rights management of relational data through watermarking as an important and technically challenging problem for database research. Enunciation of the various malicious attacks from which the watermark inserted in a relation must be protected. we would like to extend the proposed watermarking technique to also mark non-numeric attributes. We would address the related problem of fingerprinting to be able to identify the culprit in cases where there can be multiple sources of piracy.

2.5 Content based Zero-Watermarking Algorithm for Authentication of Text Documents

Copyright protection and authentication of digital contents has become a significant issue in the current issue with efficient communication mediums such as internet. Plain text is 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. 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. Digital watermarking is one of the solutions which is used to authenticate and to protect digital contents A digital watermark is either visible or invisible and the identification code that is embedded with data will remain even after the decryption process.

The extraction algorithm takes the plain text and keyword as input. The text may be attacked or un attacked. The watermark is generated from the text by the extraction algorithm and then, compared with the original watermark registered with the CA (certifying authority). We have also recorded author name, current date and time with the CA. Multiple watermark registration conflicts with CA can be resolved by keeping record of time and date. The author having former registration entry will be regarded as the original author. The watermark will be accurately detected by this algorithm in the absence of attack on text, and text document will be called authentic text without tampering. The watermark will get distorted in the presence of tampering attacks with text.

Tampering can be insertion, deletion, paraphrasing or reordering of words and sentences in text.

The existing text watermarking solutions are not applicable under random tampering attacks and on all types of text. With the small volume of attack, it becomes impossible to identify the existence of attack and to prove authenticity of information. Using zero-text watermarking algorithm, the contents of text to generate a watermark and this watermark is later extracted to prove the authenticity of text document.

3 CONCLUSION:

Security is one of the major issue which is being faced in real world. Day by Day millions of data are being generated which cause an issue from keeping them secured. While these techniques provides a starting point for securing the data from the intruders from accessing it, further research is needed to turn them into practical solutions that can achieve privacy and security in the real world.

References

- [1] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*, Intel. Technol.Theory Appl. New Trends Intell. Technol., vol. 76, pp. 214 220, 2002
- [2] R. Agrawal99 and J. Kiernan, *Watermarking relational databases*, in Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 155166.
- [3] Y. Zhang, B. Yang, and X.-M. Niu, *Reversible watermarking for relational database authentication*, J. Comput., vol. 17, no. 2, pp. 5966, 2006.
- [4] E. Sonnleitner, *A robust watermarking approach for large databases*, in Proc. IEEE First AESS Eur. Conf. Satellite Telecommun., 2012, pp. 16

- [5] K. Jawad and A. Khan, *Genetic algorithm and difference expansion based reversible watermarking for relational databases*, J. Syst. Softw., vol. 86, no. 11, pp. 27422753, 2013.

