A Survey on Relational Database Watermarking Using Clustering Approach

Bhatt Nisha, Nishidh Chavda
Department Of Computer Enginerring,
Chandubhai S. Patel Institute Of Technology,Changa
Bhattnisha007@gmail.com,nishidhchavda.ce@charusat.ac.in

Abstract-Digital watermarking has been widely applied to relational database for ownership protection and information hiding. But robustness and reversibility are two key challenges due to the frequently database maintaining operators on those tuples. This paper proposes a novel relational database watermarking scheme based on a fast and stable clustering method on database tuples, which adopts Mahalanobis distance as the similarity measurement. Before the process of watermark embedding and detecting, the databases tuples are adaptively clustered into groups according to the length of binary watermark. Moreover the watermark segments are respectively embedded into or detected from those groups according to the numeric field's Lowest Significant Bit (LSB) and polar angle expansion. The majority decision strategy is used to determine the value of watermark bit in blind detection process.

Index Terms-Database watermarking; robustness; reversibility; blind detection; tuples clusteringpolar angle expansion.

I. INTRODUCTION

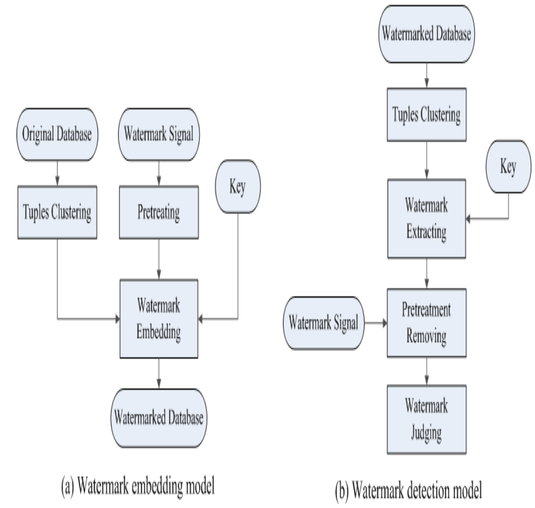
THE rapid growth of internet and related technologies has offered an unprecedented ability to access and redistribute digital contents. In such a context, enforcing data ownership is an important requirement which requires articulated solutions, encompassing technical, organizational and legal aspects[1].Though we are still far from such comprehensive solutions,in the last years watermarking techniques have emerged as an important building block which plays a crucial role in addressing the ownership problem[2]. Such techniques allow the owner of the data to embed an imperceptible watermark into the data. A watermark describes information that can be used to prove theownership of data, such as the owner, origin, or receiptof the content [3]. Secure embedding requires that the embedded watermark must not be easily tamperedwith, forged, or removed from the watermarked data [4]. Imperceptible embedding means that the presence of the watermark is unnoticeable in the data.Furthermore, the watermark detection is blinded,that is, it neither requires the knowledge of the original data nor the watermark. Watermarking techniques have been developed for video, images, audio, and text data [5] and also for softwareand natural language text [6].

There is a rich body of literature on watermarking multimedia data. Most of the work is developed for still images and then video and audio sources. There is a many differences between multimedia data and relational database. So watermarking of multimedia data cannot be directly used for relational database[7].

Digital watermarking technique has been successively applied to protect the multimedia works and software products. Similarly, database watermarking has been proposed on large database security-control. However, there are some differences between relational database and multimedia data [8]. Firstly, a relational database table consists of many attributes and tuples, but there is no certain ordering between tuples or attributes of a relation table. Secondly, database maintaining operators could frequently change those tuples unlikely other type of multimedia object. Moreover, database tuples processing rely on logical set operational language such as SQL. So database watermarking should also have the ability of real-time update and blind detection and cannot directly adopt those multimedia watermarking method[9].It is more difficult to ensure the robustness and reversibility of database watermarking.

II.FRAMEWORK AND STRATEGY

Allowing for the disorderliness of tuples and attributes, insufficient redundant space of database, along with weak robustness of the general database watermarking algorithm, it is practicable to realize the database watermarking embedding and robust detection with the stable[10], high-efficiency and large-capacity database tuples clustering method, which is regarded as the basis of database watermarking algorithm in this paper. Thereinto the similarity among databases tuples is measured by Mahalanobis distance since it can effectively eliminate the influence of dimension and correlation interference. Meanwhile there are frequently database maintaining operators on tuples and attributes which would affect the robustness of database watermarking, and we use majority decision method to solve the problem when extracting watermark. Based on the above tuples clustering and majority decision strategy, we present a robust database watermarking framework shown in Figure 1(a) and (b)[11].



(a) Watermark embedding model

(b) Watermark detection model

Figure 1. Robust Database Watermarking Framework based on Tuples Clustering and Majority Decision Strategy.

III.TYPES OF DIGITAL WATERMARKING

A. Spatial Domain Method

The spatial domain is the normal image space, in which a change in position in I directly projects to a change in position in space. Distances in I (in pixels) correspond to real distances(e.g. in meters) in space. This concept is used most often when discussing the frequency with which image values change, that is, over how many pixels does a cycle of periodicallyrepeating intensity variations occur[12]. One would refer to the number of pixels over which a pattern repeats (its periodicity) in the spatial domain. Here we use Least Significant bit(LSB)method.

B. Transform Domain Method

The produce of high quality watermarked image is by first transforming the original image into the frequency domain by the use of Fourier, Discrete Cosine Transform (DCT) or Discrete Wavelet transforms (DWT) for example. With this technique, the marks are not added to the intensities of the image but to the values of its transform coefficients. Then inverse transforming the marked coefficients forms the watermarked image. The use of frequency based transforms allows the direct understanding of the content of the image; therefore, characteristics of the human visual system (HVS) can be taken into account more easily when it is time to decide the intensity and position of the watermarks to be applied to a given image.

C. Data Security Through Watermarking

Invisible digital watermarks are a new technology which couldsolve the “problem” of enforcing the copyright of contenttransmitted across shared networks. They allow a copyright holderto insert a hidden message (invisible watermark) within images,moving pictures, sound files, and even raw text. Furthermore, theauthor can monitor traffic on the shared network for the presenceof his or her watermark via network system. Because this methodconceals both the content of the message (cryptography) and thepresence of the message (steganography) an invisible watermarkis very difficult to remove[13] Thereby, this technology could greatlystrengthen the enforcement of copyright law on the Internet.

Data hiding is defined as the process by which a message or imageis imperceptibly embedded into a host or cover to get a compositesignal. Generally, in encryption, the actual information is notmaintained in its original format and thereby it is converted intoan alternative equivalent multimedia file like image, video oraudio, which in turn is being hidden within another object. Thisapparent message is sent through the network to the recipient,where the actual message is separated from it .

IV.LEAST SIGNIFICANT BIT

One of the simplest technique in digital watermarking is inspatial domain using the two dimensional array of pixels in thecontainer image to hold hidden data using the least significantbits (LSB) method. Note that the human eyes are not veryattuned to small variance in color and therefore processing ofsmall difference in the LSB will not noticeable. The steps toembed watermark image are given below.

A. Steps of Least Significant bit

1) Convert RGB image to gray scale image.
2) Make double precision for image.
3) Shift most significant bits to low significant bits of watermark image
4) Make least significant bits of host image to zero
5) Add shifted version (step 3) of watermarked image to modified (step 4) host image.

B. Limitations of Spatial Domain Watermarking

This method is comparatively simple, lacks the basic robustness that may be expected in any watermarking applications. It can survive simple operation such as cropping, any addition of noise. However lossy compression is going to defeat the watermark. An even better attack is to set all the LSB bits to ‘1’ fully defeating the watermark at the cost of negligible perceptual impact on the cover object. Furthermore, once the algorithm was discovered, it would be very easy for an intermediate party to alter the watermark[14].

V. DISCRETE COSINE TRANSFORM WATERMARKING

The DCT allows an image to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that they have minimized they avoid the most visual important parts of the image (low frequency) without over-exposing themselves to removal through compression and noise attacks[1].

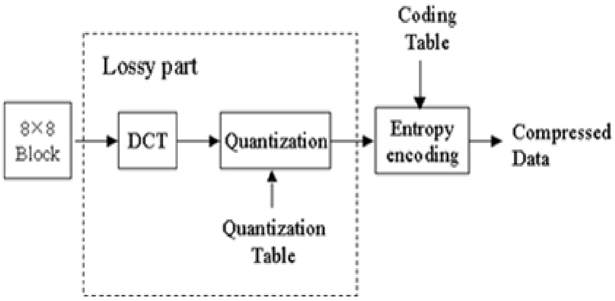


Fig. 2. DCT Watermarking

V.CLUSTERING APPROACH

Here we apply the fast clustering method to the classification of database tuples, whichbegins with classifying samples roughly, then uses certain regulations to adjust the categories gradually based on the distance between samples. It is suitable for clustering analysis of large data sets. The similarity of samples is measured by distance[15]. Due to the disunity of various attributes units in database, in order to eliminate the influence of dimension, this paper adopts k-Mean distance to cluster the tuples.which discards the traditional method to partition subsets. we also introduce an odd-even modifying method to embed watermarking information shows it is effective.

VI. CONCLUSIONS

In this paper, we proposed a fragile watermarking scheme for database relations.The watermarks are embedded into a database relation on the group basis under the control of a secure embedding key. The embedded watermarks form a watermark grid which cannot only detect, but also localize and characterize any modifications made to the database. Security analysis showed that it is very difficult for an attacker to modify the database without affecting the embedded watermarks, and the security upper bound was given. Future work will focus on designing a semi-fragile watermarking scheme so that the embedded watermarks can be robust to small modifications and fragile to severe modifications and watermarking scheme that can embed watermarks to non-numeric attributes.

REFERENCES

[1]R. Agrawal, J. Kiernan, Watermark relational databases, in: Proc. of the 28th Int. Conf. on Very Large Data Bases, 2002.

[2] D. Barbara, R. Goel, S. Jajodia, A checksum-based corruption detection techniques, J.Comput. Security 11 (3) (2003) 315–329.

[3] M. Chen, Y. He, R. Lagendijk, A fragile watermark error detection scheme for wireless videocommunications, IEEE Trans. Multimedia (August) (2003) 315–329.

[4] I.J. Cox, J. Kilian, T. Leighton, T.G. Shamoon, Secure spread spectrum watermarking formultimedia, IEEE Trans. Image Process. 6 (12) (1997) 1673–1687.

[5] P. Devanbu, M. Gertz, C. Martel, S. Stubblebine, Authentic data publication over the internet,in: Proc. 14th IFIP 11.3 Working Conf. in Database Security, August 21–23, 2000, pp. 102–104

[6] J. Fridrich, M. Du, Images with self-correcting capabilities, in: Proc. of the IEEE Int. Conf. onImage Processing, 1999, pp. 792–796.[10]Ching-Ming, Po-Zung& Chu-Hao," Privacy Preserving Clustering of Data streams", Tamkang Journal of Sc. & Engg,Vol.13 no. 3 pp.349-358

[7]R. G. van Schyndel, A. Z. Tirkel and C. F. Osborne, "A Digital Watermark", Proc. ICIP'94, vol. 2, **(1994)**, pp. 86-90.

[8] I. Cox, M. Miller, J. Bloom and C. Honsinger, "Digital Watermarking", Academic Press, USA **(2002)**.

[9] R. Sion, M. Atallah and S.Prabhakar, "Rights Protection for Relational Data", IEEE Transactions on Knowledge and Data Engineering, vol. 16, no.12, **(2004)**, pp. 1509-1525.

[10]R. Agrawal and J. Kiernan, "Watermarking Relational Databases", Proc. VLDB'02, **(2002)**, pp. 155-166.

[11]R. Sion, M. Atallah and S. Prabhakar, "On Watermarking Numeric Sets", Proc. IWDW, **(2002)**, pp. 12-15.

[12]X. Niu, *et al.*, "Watermarking Relational Databases for Ownership Protection", Chinese Journal of Electronics (in Chinese), vol. 31, no. 12A, **(2003)**, pp. 2050-2053.

[13].R. Sion, M. Atallah, S. Prabhakar, Rights protection for relational data, in: Proc. of ACMSIGMOD 2003, pp. 98–109.

[14] P. Samarati, Protecting respondents_ identities in microdata release, IEEE Trans. KnowledgeData Eng. 13 (6) (2001) 1010–1027.

[15].L. Chen, A. Min´e, and P. Cousot, "A sound floatingpointpolyhedra abstract domain," in APLAS '08: Proceedingsof the 6th Asian Symposium onProgrammingLanguages and Systems. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 3–18.