A Novel Multiple Watch marking for Relational Databases using Multi-Media

Xuemin Zhao, Lingling Li, Qingtao Wu

Department of computer, Zhengzhou Institute of Aeronautical Industry Management
Zhengzhou 450015, P.R. China

Abstract

As a tool for interests in both storing and managing data, relational database is widely used in many information systems. In this paper, a novel method of watermarking is presented to protect relational databases copyright. We propose a novel multiple watermarking scheme, which embeds multi-media watermarks into relational database. The watermark embedding and extraction algorithms are specified, the results of the corresponding watermark experiments and the attack experiments verify that the proposed method is correct, feasible and robust.

© 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Garry Lee

Keywords: Watermarking relational databases; Copyright protection; Information hiding.

1. Introduction

With the rapid development of Internet techniques, the copy and distribution of electron data become easier and easier. Simultaneity, the illegal copy and tampering of electron data seriously increase. This possibility has greatly expanded by the advent of the Internet. However, the ease of copying and reproducing digital data is likely to encourage Intellectual Property Rights violation. The digital watermark technique represents a valid solution to the above problem, since it makes possible to identify the source, author, creator, owner, distributor or authorized consumer of digital image, video, audio and text, etc[1].

Improve the robustness and database security database watermarking has been one of the main objectives. People in actively explore how to make the hidden information can avoid intentionally or unintentionally attack, seeking for as many attack algorithm has immunity in database watermark embedding algorithm is proposed. Francesc Sebe studied keep watermark database data, and add noise attacks with good watermarking algorithm robust database[6]. Xiamu Niu forward in watermark before
introducing numerical can change range of restrictions, to satisfy the restrictions of attribute values of the
least significant bit embedded information. Zhang yong combining cloud model theory realized in
relational database watermark signal embedding and extracting.

For fragile watermark On the database research. Relative to research more robustness watermarkings
are concerned, fragile watermark in has stronger ability against the attack at the same time, it also requires
a strong sensitivity that is allow some degree of distortion, and can detect out situation distortion. Yingjiu
Li explored for the first relational database fragile watermark algorithm and applies it to the database
integrity verification and tamper with the judgement. Xiao Xiang-rong use numeric data in a certain
accuracy range allows deformation characteristics to add watermarks, vulnerability and attacks of the
watermark and the experimental data recovery.

In this paper, we study a novel multiple watermark for relational databases using multi-media. An
identification image is embedded into the relational data for representing the copyright information which
is for robustness. Meanwhile, we all use speech as the original watermark. As a biometric of human being,
speech is inherent and not changes along with time. So using speech as watermark can make up the
limitations mentioned above. This can be for fragile. To be effective and convictive the watermark
embedding process is imperceptible, secure and reliable, the watermark extraction process is a blind
detection process and the watermark scheme is resistant to some malicious attacks[2].

2. Our algorithms

We now propose a new scheme for watermarking relational databases. We use speech and information
of the copyright holder to generate watermark by watermark generation algorithm, then present the
corresponding insertion algorithm and detection algorithm (see Fig.1). The problem of watermarking
relational databases can be modeled as follows. Suppose relation R contains primary key P and numerical
attributes A0, A1, ..., Av-1. Assume that it is acceptable to change one of Av least significant bits (LSB). A
character image which will convert as a sequence of 0 and 1 is to be embedded into relation R for the
purpose of copyright protection[3].

2.1 Watermark generation

We assume that some minor changes of some attributes values can be tolerated. And we will embed
copyright information into these attributes. We consider the character image (copyright information) as a
sequence of 0 and 1, the marks of 0 and 1 are small errors in the relational data. All the marks of 0 and 1
represent integrated copyright information.

The detailed process of watermark generation algorithm is as follows: 1) compress the media signal.
Watermarking relational databases introduces small errors into the relations by inserting watermark into
them. The marks must not have a significant impact on the usefulness of the data, so the watermark
should be small. Thus, the compression of the image signal is necessary because of its large information
capacity. We use wavelet to compress it; 2) media signal convert. Convert the signal waveforms to the 8-bit A-law ;The pixels of the copyright image are arranged by the order from left to right and then top to bottom and we can get a set $S = \{S_1, S_2, ..., S_{N \times N}\}$, in which each element is the decimal grey value of the pixel.

2. Insertion Algorithm

We use a one-way hash function result decided by the primary key $P$ and the secret key $K$ to choose where to mark and what to mark[4]. For the embedded watermark in relational database is doing two hypotheses: watermarking embedding object relational database of only numeric attributes, A relational database of numeric attribute value can tolerate small scope changes. Based on the above two hypotheses, in data minimum significance realizing embedding throne. Embedding procedure is as follows:

```plaintext
// Weighted watermark algorithm, return marked R
// Hi is one-way hash functions, L is the length of EMC
// Parameters k, L, α, ζ and υ are private to the owner.
1  E[L]=H(k concatenate M)    // calculate L-bit EMC
2  foreach tuple r ∈ R do
3    t= H1( k concatenate r.P)
4    if ( t mod υ equals 0) then    // mark this tuple
5      i = select_attribute()    // mark i-th attribute
6      j = t mod ζ       // mark j-th bit
7      k= t mod L       // use the k-th bit of EMC
8      m=Ek XOR (k mod 2)  // value of marked bit
9      set the j-th LSB of r.Ai to m
10  return  R
11  procedure  select_attribute()
12    u= H2( k concatenate r.P)
13    d=(1/ Wi)   i ∈[0,i-1]
14    if (u mod d) ∈ (Σ(1/ Wi-1),Σ(1/ Wi] then return i    // W-1=0
```

2.2 Watermark Extraction Algorithm

If the owner of the dataset suspects that some datasets are illegal copied or tampered from his relational database R. He or the third party can use watermark extraction algorithm to verify the ownership of the suspicious database. Now we introduce our watermark extraction. Because of the identical distribution of the hash function when seeded by the same key, the selected tuples in each group is of the same order as in the insertion algorithm. For each marked bit, we count the numbers of its value to be zeroes or ones respectively, and then a majority voting mechanism is to decide the final value of this bit. The detected result is a binary sequence which includes the copyright information of the dataset[5]. Thus, we transform the binary sequence return back to the Beijing 2008 Emblem which is the Watermark image. The copyright holder can make use of the Watermark image to prove the copyright. The watermark detection algorithm was given as follows:

```plaintext
// Algorithm to return a watermark M[ ] from relation R
// parameters k, L, α, ζ and υ are also private to the owner.
1  for s=0 to L-1 do
```
2 DM[s] =”” // initialize detected mark code
3 count[s][0]=0, count[s][1]=0 // initialize counter
4 for each tuple r∈R do
5 t = H1(k concatenate r.P)
6 if (t mod γ equals 0) then // select this tuple
7 i = select_attribute() // mark i-th attribute
8 j = t mod ξ // select j-th bit
9 k = t mod L // mark the k-th bit of EMC
10 m = (j-th LSB of r.Ai) XOR (k mod 2)
11 count[k][m] = count[k][m]+1 // add the counter
12 for s=0 to L-1 // get the watermark
13 if (count[s][0]>=count[s][1]) // majority voting
14 then M[s]=0 else M[s]=1 // the final bit value

3. Experiment and result

To test the validity and robustness of this algorithm, we perform experiments on a computer running Windows XP Professional with 2.4 GHz CPU and 256MB RAM. Algorithms are implemented on Java Eclipse Platform Version 3.0 using JDBC to visit Microsoft Office Access 2003. We applied our algorithms to generated synthetic data with 8 attributes. The size of the generated set was 30,000. We choose MD5 as the one-way hash function, and significance level is 0.01. Multiple watermarking relational databases using image, every watermarking image has its own parameters that used in insertion algorithms and the corresponding extraction algorithms. The testing results of the various subset attacks on relational database and the necessary analysis were given by following.

Subset Selection: In this kind of attack, violators just select a subset from the relation R. We select different ratio of the original data to simulate such attacks[6]. We can see in Fig 2 it has no effect at all to our watermark by selecting 50% of the watermarked relation. This shows a great of robustness, Even when 25% of the data is selected, outline drawing of the watermark image can be successfully recovered. We can see that even a small part of the marked relation is enough for a successful detection.

Fig. 2 Result of subset selection attacks

Subset addition attack. The attacker randomly selects out part of the watermarked relation and mixes them with similar tuples probably without watermarks to form a new relation of approximately the same size of the original one[7]. Fig 3 shows the result. When (ω/L) = 10, 100% of the watermark can be recovered when we randomly select 70% tuples from the watermarked relation and mix them with 30% tuples from the original unmarked relation. But when we select 50% or less of the watermarked relation,
we fail to detect the watermark based on statistics at the significance level of 0.01. When we enlarge \((\omega/L)\) to 20, we get better results.

![Fig.3 Result of subset addition attacks](image)

Subset Alteration: Violators try to alter some attribute values of relational database which had watermarked, and stochastically alter the attribute values. We randomly altered a portion of watermarked tuples by resetting one bit oppositely among the 8 candidate bits in each tuple, i.e., change the value 0 of a bit into 1 and value 1 into 0. We can see in table 4 that we can successfully recover the entire watermark image when we altered 50% or less tuples. But fail to detect the watermark when 60% or more tuples are attacked. See table IV[8], see fig 4.

The algorithm is “blind” in that it requires neither the original data nor the watermark in order to detect a watermark in an object. Thus the watermark can be detected in a copy of the data base relation, irrespective of later updates to the original relation. The watermark can be easily and efficiently maintained[9].

![Fig. 4 Result of subset alteration attacks](image)

### 4. Conclusion

In this paper, we propose a novel method for watermarking relational database, which uses media as the watermark. These experiments show that our watermark detection algorithm is robust even when an attacker drops some of the tuples or the watermarked attributes from the relation. Moreover, depending upon the number of attributes omitted and the number of tuples dropped, we can estimate the size of the sample needed for detecting the watermark. Our approach is more intuitive, and it support easy watermark identification.
Acknowledgment

The research work is supported by Program for New Century Excellent Talents in University; by Excellent Youth program for science & technology innovation talent of Henan Province; by Aeronautical Science Foundation supported by Key Laboratory of Science and Technology for National Defense of fire control (No. 20095155008); by key project of the Department of Science & Technology of Henan Province (No. 102102210453); by Foundation for University Youths Key Teacher by the Education Department of Henan Province; and by Technological Innovation Cultivate Program of Zhengzhou City (No. 10PTGG342-1).

References

[1] Rakesh Agrawal, Jerry Kieman. “Watermarking Relational Databases”. Proceedings of the 28th VLDB Conference. Hong Kong, China, 2002: 155—166.
[2] Radu, S., Mikhail, A., Sunil, P.: Rights Protection for Relational Data. Proceedings of ACM SIGMOD, (2003) 98–1094. Min, H., Cao, J. H., Peng, Z. Y., Fang, Y.: A New Watermark Mechanism for Relational.
[3] Xiamu Niu, Liang Zhao, Wenjun Huang, “Watermarking Relational Databases for Ownership Protection”. Chinese of Journal Electronics. 2003, 31(12A): 2050–2053.
[4] Guo Fei, Wang Jianmin, Zhang Zhihao, et al. An Improved Algorithm to Watermark Numeric Relational Data, [C]/Proceedings of WISA 2005, Lecture Notes in Computer Science 3786, Beijing, China, 2006:138 – 149.
[5] Y. Li, V. Swarup, S. Jajodia. Fingerprinting Relational Databases: Schemes and Specialties, [J]. IEEE Transactions on Dependable and Secure Computing, Vol. 2, NO. 1, Jan.-Mar. 2005:34-45
[6] Zhang Yong, Zhao Dongning, Li Deyi. Digital Watermarking for Relational Databases. [J].Journal of PLA University of Science and Technology, Vol.4, No.5, 2003:1-4 (Ch)
[7] D. Gross-Amblard. Query-preserving Watermarking of Relational Databases and XML Documents, [C].PODS 2003, San Diego CA. 2003:191–201.
[8] Y. Li, V. Swarup, S. Jajodia: Constructing a Virtual Primary Key for Fingerprinting Relational Data, [C]. DRM’03, October 27, 2003, Washington DC, USA. (2003),133-141
[9] Schneier. Applied Cryptography. [M] John Wiley & Sons, second edition, New York, USA, 1996.