Voice Based Watermarking Technique for Relational Databases

Sachin Balawant Takmare, Ravindra Kumar Gupta, Gajendra Singh Chandel

Abstract— This paper is about the technique of watermarking relational databases for unauthorized access. Signals of voice are introduced as watermark to be embedded into the relations; associated novel watermark insertion algorithm and detection algorithm are proposed. Detecting the watermark neither requires access to the original data nor the watermark. Thus, the watermark signal in this method is expected to be more meaningful and has closer correlative to the unauthorized user. The proposed technique is feasible, and robust against various forms of malicious attacks.

Index Terms— watermarking relational databases, unauthorized access protection, analog signal, biometric technology, malicious attacks.

1 Introduction

THE piracy of the data is increasing now a days watermarking relational database has become a hotspot in recent years. Yet most of them use a private key of a copyright holder as watermark but still have some limitations. Thus we propose a new idea for watermarking relational database, which uses voice as the original watermark as far as we know voice of a human being is inherent and not changes along with time. There are some features which we can use as meaning of the voice signal to prove the copyright[1][4]. The main contribution of this work is the voice signal is embedded into relations. Voice of copyright holder is used to generate watermark by watermark generation algorithm. Thus relational database watermarking is in fact, a process challenged by many factors such as data redundancy fewness, relational data, out of order and frequency updating

2 A Voice Based Technique For Watermarking Relational Databases

We have a new scheme for watermarking relational databases. Voice of database holder is used to generate watermark by watermark generation algorithm, then corresponding insertion and deletion algorithm are being presented in figure 1. Suppose totally we have η tuples in R out of which we will mark ω tuples. It is acceptable to change one of ξ least significant bits(LSB) in a small number of (1/ α) numeric values, relational R contains η tuples, and a fraction ω=1/α then can be used for watermark. Table 1 summaries the important parameters used in our algorithms.

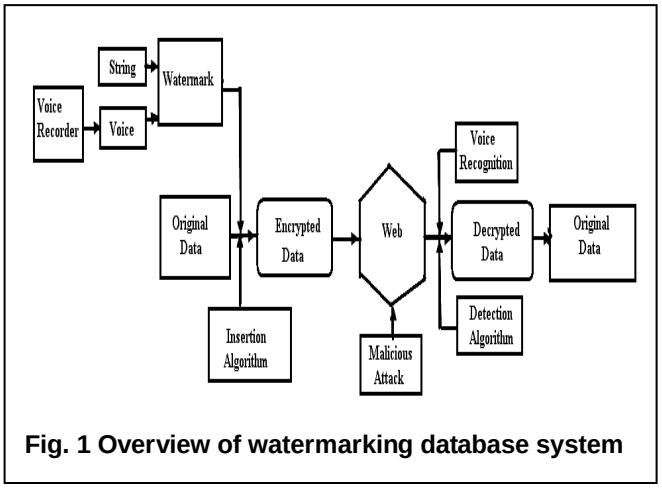

Fig. 1 Overview of watermarking database system

2.1 Use of voice as a watermark

Voice of copyright holder is used for watermark because of its following characteristics.

1) No two persons can have same fetures in their voice i. e. voice of every persone has a unique fetures.
2) Feture of voice don't get change along with time, so we can easily use our voice for inserting watermark as our voice may not change little since we grow up.
3) everyone has speech so universality is again one of the characteristic of voice.
4) Measurable, speech can easily be converted into bit format thus use of voice content large data. So it is challenging to use voice as a watermark.

3 Watermark Generation Algorithm

Procedure of this algorithm is divided into two parts.

1) Voice of a copyright holder is taken from microphone and it is being converted into bit format.
2) Bits of voice along with one string are being used to generate watermark.

The detail procedure of watermark generation algorithm is as follows.

1) Watermark insert small errors into relations by inserting watermark into them. Watermark should be small as it must not effect on the usefulness of data. Thus the voice of the copyright holder must be compressed as it has large information capacity thus the compression of a voice signal must be done.
2) Input signal taken from the microphone is sum of speech spectrum and the noise spectrum. Thus to remove the background voice the noise spectrum must be removed from the speech.
3) Voice of the copyright holder taken from microphone is then converted into 8-bit-a-low.
4) Thus after conversion of voice into bits those bits are then used along with the insertion algorithm to generate watermark [4].

4 Watermark Insertion Algorithm

First we convert the watermark w into flow(Encrypted mark code) of certain length. We use the one way Hash function to decide which tuple and which bit to be marked. We can divide the relation into groups of varied but similar sizes. The ith bit selected for the watermark is being replaced by the 1st bit of the voice file and so on Figure 2 gives the watermark insertion

algorithm[1][2].

//Hi is one way hash function is the length of EMC
//Parameters k,L, α, ξ and v are private to the owner.
1. E[L]=H(k concatenate M)//calculate L-bit EMC
2. For each tuple r∈R do
3. t=H1(k concatenate r.P)
4. if (t mod γ equals 0) then // mark this touple
5. i=t mod ξ // mark i^{th} bit
6. j=t mod L // use the j^{th} bit of EMC
7. m=Ek XOR (j mod 2) // value of marked bit
8. set the i^{th} LSB of r.Ai to m
9. return R

Fig. 2 Watermark Insertion Algorithm

5 WATERMARK DETECTION ALGORITHM

The watermark detection algorithm is used to recover the watermark from the suspect relation.

1) The majority voting mechanism is used to find the final watermark.
2) We make voice denoising to eliminate the impact of attacks after the final watermark changed to the voice by the inverse process oof the watermark generation. Thus we compare the designated features from the recovered voice signal and compare it with same feature of the copyright holder thus our scheme is proved to robust[1][2].

//Parameters k,L, α, ξ and v are also private to the owner.
1. for s=0 to L-1 do
2. DM[s]=”// initialize detection 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 touple
7. i=t mod ξ // select i^{th} bit
8. j=t mod L // mark the j^{th} bit of EMC
9. m=(j^{th} LSB of r.Ai) XOR (j mod 2)
10 count[j][m]=count[j][m]+1 //add the counter
11. for s=0 to L-1 //get the watermark
12. if(count [s][0]>= count [s][1]) // majority voting
13.then DM[s]=0 else DM[s]=1// the final bit value
14.return DM[]

Fig 3 : Watermark Detection Algorithm

6 EXPERIMENTS AND RESULT

The algorithms which we have implemented are being tested on the databases which we have constructed. The database consists of 1000 tuples and runs under the Mysql platform. We concentrate our performance evaluation on the robustness of the proposed algorithm by virtue of the fact that, database watermarking algorithms must be developed in sach a way to make it difficult for an adversary to remove or alter the watermark beyond detection without destroying the value of the object[5]. The algorithms should make the watermarked databases robust against the following types of attacks:

1. Subset addition attack
2. Subset deletion attack
3. Subset selection attack

6.1 Subset addition attack

The attacker may add a set of numbers to the original set. This addition is not to significantly alter the properties of the initial set versus the result set Figure 4 shows the subset addition attack 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 (ω/L) to 20, we get better results.

6.2 Subset selection attack

The attacker can randomly select and use a subset of the original data set that might still provide value for the intended purpose. We select different ratio of the original data to simulate sach attacks and the mark length is 256 with a marking frequency (ω/L) of 10. Figure 5 shows the detection ratio of watermarks when subset selection attack occurs.

6.3 Subset alteration attack

The attacker alters a subset of the items in the original data set such that there is still value associated with the resulting set. The figure 6 shows the result. When (ω/L) = 10, we can detect 100% watermark from the watermarked set if alter 40%or less items, but fail to detect when 60% or more items. When (ω/L) = 20, we can detect 85% even when alter 80% of the items

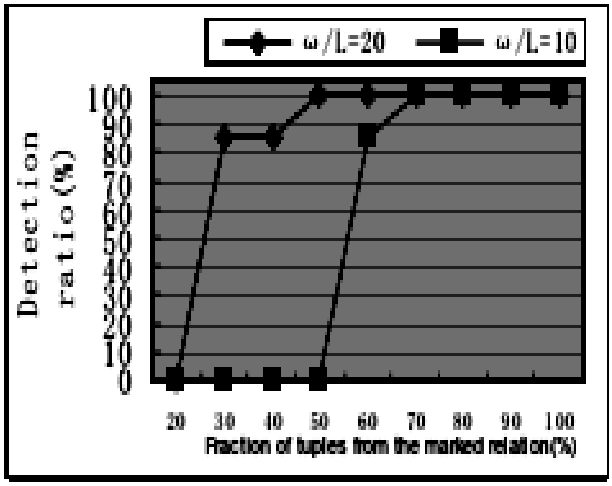

Fig. 4 Result of subset addition attack

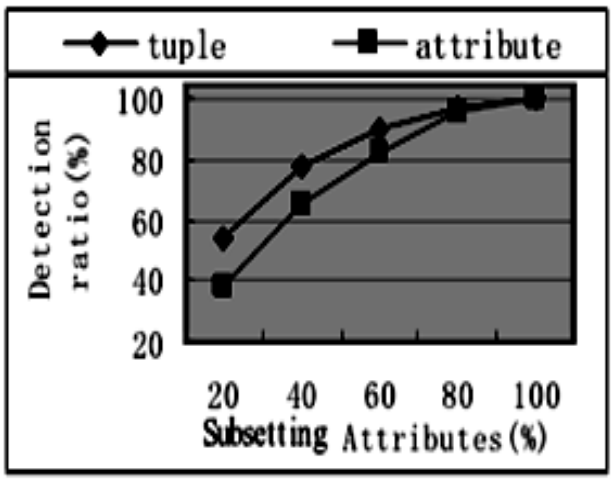

Fig. 5 Result of subset selection attack

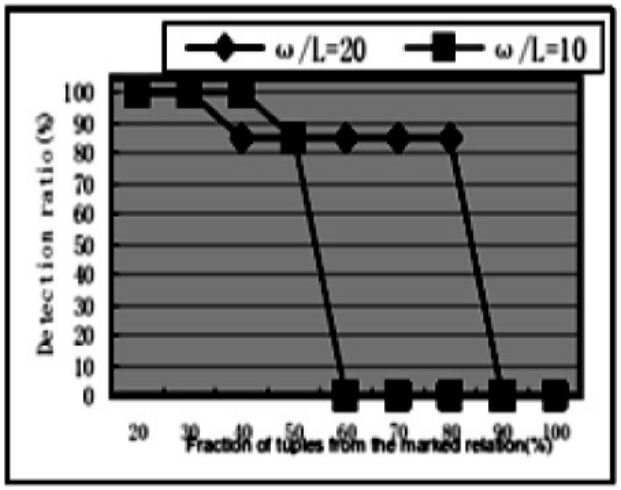

Fig. 6 Result of alteration attack

7 CONCLUSION

In this paper we propose a novel method for watermarking relational database, which use voice of a owner as a watermark. The robustness of the proposed algorithm was verified against a number of database attacks such as subset alternation, subset addition, subset alteration. Thus the copyright holder can make use of speech to prove the copyright. In future, we intend to conduct research on improving the robust of the scheme by the multiple watermarking technologies.

ACKNOWLEDGMENT

We would like to thank our family members and friends who have been a constant source of inspiration throughout our paper presentation work.

8 REFERENCES

1) R.Agrawal, J. Kiernan, "Watermarking relational databases", proceedings of the 28th International Conference on VLDB, 2002, pp.155-166.

2) R.Agrawal, P. J. Hass, J. Kiernan "Watermarking relational data : Framework, Algorithms and Analysis", VLDB Journal, 2003, pp.157-169.

3) Ali-Haj, Ashraf Odeh, "Robust and Blind Watermarking of relational Database system",Journal of computer science4(12); 1024-1029,2008, ISSN 1549-3636.

4) V. S. Inamdar, PP. Rege, Aarti Bang "Speech Based watermarking for Digital Images", TENCON 2009, 978-1-4244-4257/9/09.

5) Arti Deshpande, Jayant Ghatage ,"New watermarking Technique for relational databases", Second International Conference on emerging trends in engineering and technology, ICETET-09.

6) Meng-Hsium Tsai, Hsaio-Yun Tseng, Chen-Ying Lai, "A database technique for temper detection".

7) Sonia Jain, "Digital watermarking techniques: A case study in fingerprints and faces".

8) Mailing Meng, Xinchun Cui, haiting Cui. "The Approch for optimization in watermark signal of database by using generic algorithms". International conference on computer science and information technology 2008.