International Journal of Advance Engineering and Research Development
Scientific Journal of Impact Factor (SJIF): 4.72
Special Issue SIEICON-2017,April -2017
e-ISSN : 2348-4470 p-ISSN : 2348-6406

Speech Based Watermarking Relational Databases

Mayuree Rathva[1], Pratiksha Raval[2], Nishant Khatri[3], Nimit Modi[4], Pragna Makwana[5]

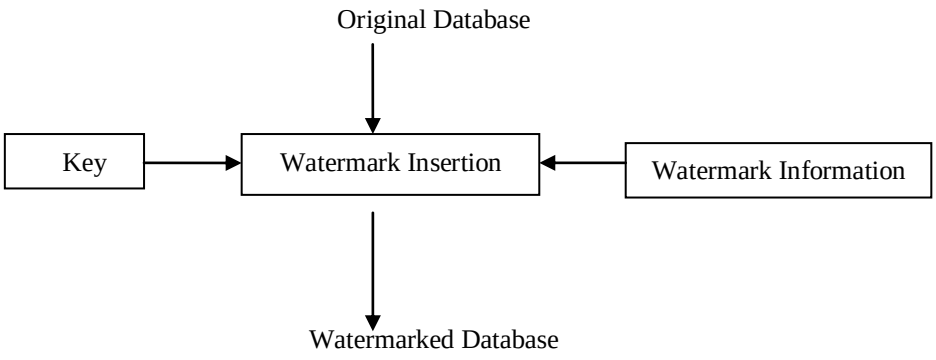
[1]*Computer Engineering, Sigma Institute of Engineering*
[2] *Computer Engineering, Sigma Institute of Engineering*
[3] *Computer Engineering, Sigma Institute of Engineering*
[4]*Computer Engineering, Sigma Institute of Engineering*
[5]*Information Technology, Sigma Institute of Engineering*

Abstract — *Digital watermarking is used for copyright protection. Because of increasing use of relational database systems in many real-life applications created an ever increasing need for watermarking database systems. As a result, watermarking relational database systems is now merging as a research area that deals with the legal issue of copyright protection of database systems. Here, we have discussed the basics of Digital Watermarking and different watermarking techniques. We have also discussed the Literature Survey on the work done in the field of the Digital Watermarking. We have proposed a new watermarking approach which is based on speech as watermark in which watermark is embedded at multiple places. Here, we have selected two attributes first is watermark is placed in numeric fields and second attribute is DATE field in which watermark is placed.*

Keywords-*Digital Watermarking, Relational Database Systems , Speech based, Numeric fields, DATE field.*

1. INTRODUCTION

The main reason for development of digital watermarking research is the endeavor for coming up with innovations to protect intellectual properties of the digital world. This is because the recent technological advancement in generation, storage, and communication of digital content has created/generated problems like copying the digital contents without any constraints, forgery, and editing without any prohibitive professional efforts. The absence of protecting techniques makes it doubtful to use the digital communication system in medical, business, and military applications. Watermarking is one of the most common solutions to make the data transferring secure from the illegal interference. Initially, most of the research on watermarking was done on the protection of still images, video, audio, VLSI (very large scale integration) design etc. However, in the recent years watermarking of database (DB) systems started to receive attention because of the increasing use of it in many real-life applications. There are two phases of database watermarking:- Watermark embedding and Watermark detection, as depicted in Figure 1. During watermark embedding phase, a secret key (known only to the owner) is used to embed the watermark W into the original database. The watermarked database is then made publicly available. To verify either the ownership or the integrity of a suspicious database, for the verification suspicious database is taken as input and by using the private key the same which is used during the embedding phase. The embedded watermark (if present) is extracted and compared with the original watermark information.



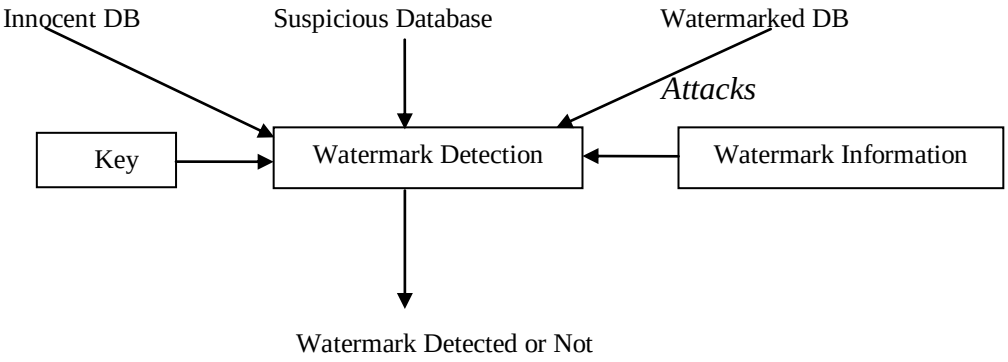


Figure 1.Basic Database Watermarking Process

1.1 Challenges of Database Watermarking

The watermarking of relational data has significant technical challenges and practical applications that deserve serious research effort. There is a rich body of literature on watermarking multimedia data. There are also new technical challenges because relational and multimedia data differ in a number of important respects:

1. Few Redundant Data: A multimedia object consists of a large no. of bits, there is a large amount of space available to hide the watermark. While Database consists of the tuples, each tuple represents a separate object. So, the watermark is spread over these separate objects.

2. Out-of-Order Relational Data: The change in a relative spatial/temporal positioning of multimedia object remains unchanged. While in case of database, updates in database may changes the tuples.

3. Frequent Updating: Multimedia objects typically remain intact; any portion of such an object is not dropped or replaced arbitrarily without causing perceptual changes in the object, whereas tuples may be inserted, deleted, or updated during normal database operations.

Due to these differences between relational and multimedia data, there exists no multimedia watermarking method which is suitable for watermarking of relational databases. These differences give rise to many technical challenges in database watermarking as well.

1.2 Different Types of Watermarking Techniques

Traditionally, database watermarking schemes can be classified to be robust or fragile according to their sensitivity to database attacks. Recently, a new class of watermarking called semi-fragile has emerged in the context of relational databases.

1. Robust watermarking: Robust watermarks are usually used for copyright protection, ownership proof, or traitor tracing. In a robust watermarking scheme, the embedded watermark should be robust against attacks to remove the watermark or making it undetectable, Attacks do not significantly degrade the data usefulness; otherwise there is no need to claim ownership over such data.

2. Fragile watermarking: Generally, digital watermarking for authentication is called fragile watermarking. The embedded watermark should be destroyed even if the change in the marked data is minor. As a result, In fragile watermarking, no legal update is allowed; It means that fragile scheme cannot differentiate between benign updates and malicious modifications.

3. Semi-fragile watermarking: Semi-fragile watermarking is a trade-off between robust watermarking and fragile watermarking: semi-fragile watermarks should be robust to benign updates and fragile to malicious modifications. Thus, a semi-fragile watermarking scheme aims to verify that the content of data has not been modified by illegitimate distortions, while allowing modifications by legitimate distortions. In other words, the purpose of semi-fragile watermarking is to prevent tampering and fraudulent of modified database relations.

1.3 Applications of Digital Watermark for Relational Databases

Digital Watermarks for relational databases are useful in many applications, including:

1. Ownership Assertion: Watermarks used for ownership assertion. To assert ownership of a relational database, Alice can embed a watermark into her database R using some private parameters which is known only to her. Then she can make the watermarked database publicly available. Suppose Alice suspects that the relation S published by Mallory has been pirated from her relation R. The set of tuples and attributes in S can be a subset of R. To defeat Mallory's ownership claiming, Alice can demonstrate the presence of her watermark in Mallory's relation. For such a scheme to work, the watermark has to survive intentional/unintentional data processing operations which may remove the watermark.

2. Fingerprinting: Fingerprinting aims to identify a traitor. In the applications where database content is publicly available over a network, the content owner would like to discourage unauthorized duplication and distribution by embedding a distinct watermark (or fingerprint) in each copy of the database content. If, at a later point in time, unauthorized copies of the database are found, then the origin of the copy can be determined by retrieving the fingerprint.

3. Fraud and Tamper Detection: When database content is used for very critical applications such as commercial transactions or medical applications, it is important to ensure that the content was originated from a specific source and that it had not been changed, manipulated or falsified. This can be achieved by embedding a watermark in the underlying data of the database. Subsequently, when the database is checked, the watermark is extracted conventionally, in cryptography literature; Mallory represents the malicious active attacker. Using a unique key associated with the source, and the integrity of the data is verified through the integrity of the extracted watermark.

1.4 Watermarking Threats

Watermarking is a game between Alice and malicious Mallory. In this game, Alice and Mallory play against each other within subtle trade-off rules aimed at keeping the quality of the result within acceptable bounds. It is as if there exists an impartial referee moderating each and every 'move'. Once outsourced, the watermarked data might be subjected to a set of attacks or transformations aimed at confusing the watermark detection. These attacks may be malicious .In a relational data framework; activities that can confuse watermark detection include the following:

1. Benign Update: The tuples or data of any watermarked relation are processed as usual. As a result, the marked tuples may be added, deleted or updated which may remove the embedded watermark or may cause the embedded watermark undetectable (for instance, during update operation some marked bits of marked data can be erroneously flipped). This type of processing are performed unintentionally.

2. Value modification attack:

Bit Attacks: The simplest malicious attack attempts to destroy the watermark by altering one or more bits, e.g. by flipping each bit to 0 or 1 according to the independent toss of a fair coin. Bit attack may be performed randomly which is known as Randomization Attack by assigning random values to certain bit positions; or by Zero Out Attack where the values in the bit positions are set to zero; or may be performed by inverting the values of the bit positions, known as Bit Flipping Attack.

Rounding Attack: Mallory may try to lose the marks contained in a numeric attribute by rounding all the values of the attribute. For this attack to be effective, Mallory needs to correctly guess the number of bit positions involved in the watermark. If he underestimates this number, his attack may not succeed; if he overestimates it, he has degraded the quality of his data more than necessary. Even if his guess is correct, his data will be inferior to Alice's data because his data values are less precise.

Transformations: An attack related to the rounding attack is one in which the numeric values are linearly transformed. For example, Mallory may convert the data to a different unit of measurement.

3. Subset Attack: Mallory may consider a subset of the tuples or attributes of a watermarked relation and by attacking (deleting or updating) on them he may hope that the watermark has been lost.

4. Superset Attack: Some new tuples or attributes are added to a watermarked database which can affect the correct detection of the watermark.

5. Collusion Attack: This attack requires the attacker to have access to multiple fingerprinted copies of the same relation.

• Mix-and-Match Attack: Mallory may create his relation by taking disjoint tuples from multiple relations containing similar information.

• Majority Attack: This attack creates a new relation with the same schema as the copies but with each bit value computed as the majority function of the corresponding bit values in all copies so that the owner cannot detect the watermark.

6. False Claim of Ownership: This type of attack seeks to provide a traitor or pirate with evidence that raises doubts about merchant's claim.

• Additive Attack: Mallory may simply add his watermark to Alice's watermarked relation and try to claim his ownership. An additive attack can be thwarted by raising the watermarking error to a predetermined threshold such that any additive attack would introduce more errors than the limit.

• Invertibility Attack: Mallory may launch an invertibility attack to claim his ownership if he can successfully discover a fictitious watermark which is in fact a random occurrence from a watermarked database.

7. Subset Reverse Order Attack: An attacker enjoys this attack by exchanging the order or positions of the tuples or attributes in relation which may erase or disturb the watermark. This attack is also called sorting attack.

8. Attribute Remapping: This attack is specific to categorical data. If data semantics allow it, re-mapping of relation attributes can amount to a powerful attack that should be carefully considered. The difficulty of this challenge is increased by the fact that there likely are many transformations available for a specific data domain. This is thus a hard task for the generic case. One special case is primary key re-mapping.

9. Brute Force Attack: In this case, Mallory tries to guess about the private parameters by traversing the possible search spaces of the parameters. This attack can be thwarted by assuming that the private parameters are long enough (e.g., 160 bits) in watermarking.

2. PROPOSED WORK

We propose a new method for speech as watermark and we embed the watermark at two places. One in our original database as proposed by H.wang and another in the time (minute) field of date attribute. So, We have multiplied watermarking so the security will also increase and we can apply to any type of database because we are using the modified version of the Brijesh B. Mehta and H.wang's method. It is comparatively easy to extract the original watermark because of the watermark will be embedded at multiple places. We can have one correctly extracted watermark from that multiple places.

We use java as development platform for implementation of the proposed method then analysis of implemented method for its robustness and impersistantness will be made. We also compare our method with the other existing methods to get idea that how better our method is over other methods so result comparison will be done.

2.1 Watermark Generation:

Here, we are using speech as watermark. Speech is in the analog form so; we need to convert into digital form. The steps are as follows.

Compress the speech 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 speech signal is necessary because of its large information capacity. Here wavelet transform is used to compress it.
Speech signal enhancement: Use spectral noise subtraction to reach it. This technique operates in the frequency domain and assumes that the spectrum of the input signal can be expressed as the sum of the speech spectrum and the noise spectrum. The procedure contains two tricky parts: Estimate the spectrum of the background noise; Subtract the noise spectrum from the speech [2].

Speech signal convert: Convert the signal waveforms to the 8-bit A-law.
Generate the watermark: Use the message of the copyright of the holder and the result of the speech signal convert to generate the watermark. This procedure is shown in figure 5.
So we can also apply our watermark using that same algorithm stated in [1] but the main difference is that this algorithm is they have used Binary Image but here we are using Speech.

2.2 Watermark Insertion:

Suppose database Relation R(P,A0,A1,……Av),where P is the primary key .Here,we assume that all v attributes are numeric.The parameters v, n, k1, k2 are known to the owner only. K1 is of 5bits and k2 is of 4bits.Here, binary image is used as watermark.

1) for each tuple s ϵ S do
2) if ((r.P) mod k1 equals 0) then
3) for each of attribute of tuple
4) if (attribute ϵ v) then
5) find LSB of that attribute and replace it with watermark bit
6) end if
7) end loop
8) end if
9) end loop

This way watermark inserted in numeric fields. We are inserting only 1bit of speech which is concatenated with the 4bit key. Now next the watermark will be inserted in the "MM" field Of "Time" field which is given below.

1) Concatenation of the value watermark bit and k2
2) Find the decimal equivalent of the string
3) Embed the decimal number in tuples selected by the pre-defined key k1 as follows:
for each selected tuple do
for each selected "Time" attribute do
if the "SS" field of the "Time" mode k1 = 0
Embed the decimal number in "MM" field
else Next attribute
end if
end loop
end loop.

This way watermark will be inserted at two different places.

2.3 Watermark Detection:

Watermark detection algorithm is used to check whether watermark is present or not.

1) for each tuple s ϵ S do
2) if ((s.P) mod k1 equals 0) then
3) for each attribute of tuple
4) if (attribute ϵ v) then
5) Find LSB of that attribute and extract it as our watermark
6) end if
7) end loop
8) end if
9) end loop

This way watermark from numerical fields will be extracted. Now the watermark from time field will be extracted by given below algorithm.

1) Extract the decimal number in tuple selected by the pre-defined key k1 as follows:
for each selected tuple do
for each selected "Time" attribute do
if the "SS" field of the "Time" mode k1 = 0
extract the decimal number from MM field
else Next attribute
end if
end loop
end loop
2) find the binary equivalent of the extracted decimal number
3) extract last bit (LSB) from it which indicate our watermark.

This way we will get watermark from two places and we will compare with original watermark.

Matchcount=0, totalcount=0


```
for each bit of watermark
    if( WM = WM1 or WM=WM2)
            then matchcount=matchcount+1
      end if totalcount=totalcount+1
end for
```

One more change in our algorithm is that we are applying same watermark in two different places so that it can be detected and compared.

3. IMPLEMENTATION METHODOLOGY AND RESULTS

3.1 Experimental Setup:

We have implemented this algorithm in Java. Our database is Microsoft Access. We have taken the dataset of forest cover type. In which there are Seven fields in which six numeric fields and one non numeric field and 1000 tuples we have taken. On which the watermark insertion algorithm is applied. We have taken auto increasing number as the primary key.here,sr. no is auto increasing number which we have considered as primary key.
Our dataset has attributes like,

Attribute Name	Attribute Type	Attribute value in Range
Sr. No.	Auto-increment number	Max dataset value
Elevation	Numeric	Meters
Aspect	Numeric	Meters
Slope	Numeric	Degrees
Horizontal_dist_Hydrology	Numeric	Meters
Vertical_dist_Hydrology	Numeric	Meters
Time(HH:MM:SS)	Non-Numeric	Time

Table 1. Dataset of Forest Cover set

3.2 Performance Evaluation:

We have tested robustness by applying the different attacks as given below.

1) Bit Attack
2) Subset addition Attack
3) Subset Deletion Attack
4) Subset Selection Attack
5) Subset Alteration Attack

3.2.1 Bit Attack
This attack attempt to destroy watermark by altering one or more Bits in the Watermark data.[1]By doing this attacker thinks that he/she will get success in removing watermark from database.but by using this approach even changing the bits of tuples we still can extract the original watermark.

3.2.2 Subset addition Attack
This attack attempts to add a set of tuples to original database.[1]By doing this attacker thinks that he/she will get success in removing watermark from database.but this kind of attacks have no effect on database it will simply ignored and we will get 100% extraction after 100% new tuples.

3.2.3 Subset Deletion Attack
This attack attempts to delete a set of tuples to original database. [1]By doing this attacker thinks that he/she will get success in removing watermark from database.
But even if 4% tuples are there then also we can extract a watermark from the database.

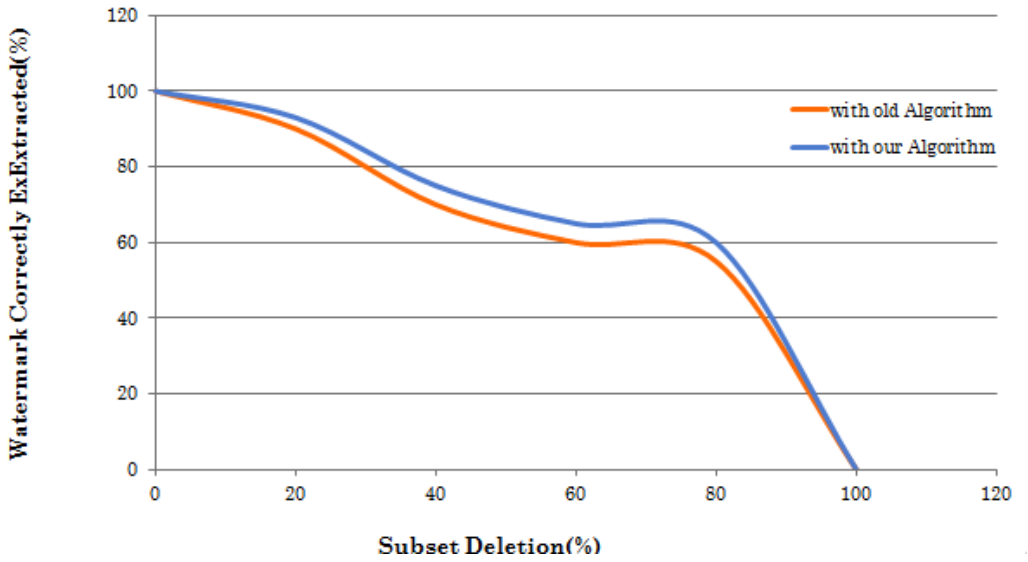


Figure 2. Subset Deletion

3.2.4 Subset Selection Attack

This attack attempts to select a set of tuples to original database. [1]By doing this attacker thinks that he/she will get success in selecting watermark from database.

But even if 20% tuples are selected then also we can extract a watermark from the database.

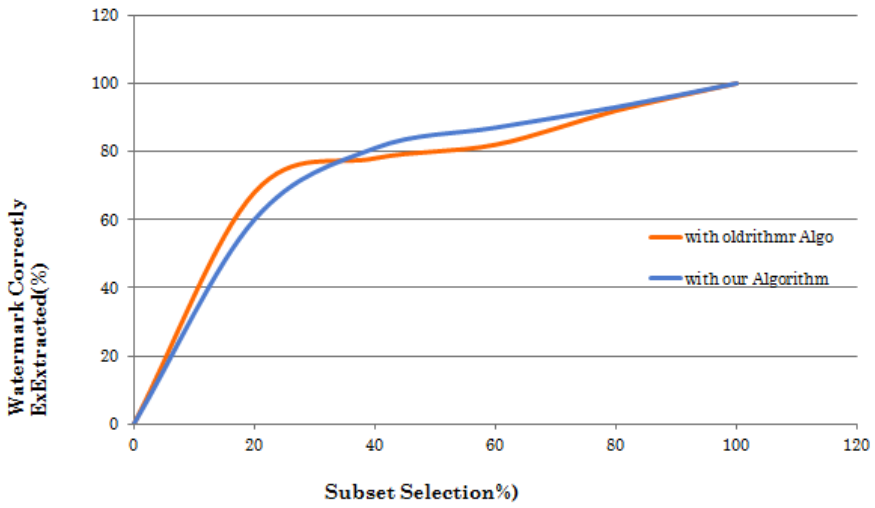


Figure 3. Subset Selection

3.2.5 Subset Alteration Attack

This attack attempts to alter the tuples of database by using linear transform. [1]By doing this attacker thinks that he/she will get success in removing watermark from database. But we can still extract the watermark after alteration.

By applying the above attacks we are getting good results as compare to others. And by inserting watermark at different places errors may be occur in database but we have to pay this price if we want to get robust result.

3.3 Results

Watermark data=101011011011001
K1=00100=4, k2=0010=2

1) Original Database

ORIGINAL_DATASET

SRNO	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical _Distance_To_Hydrology	time
1	5	8	9	2	3	5:06:15 AM
2	4	7	1	3	4	7:07:30 AM
3	5	6	2	3	8	11:21:35 AM
4	7	9	3	0	1	12:30:20 PM
5	6	8	2	1	8	1:20:10 AM
6	5	8	12	4	9	2:34:08 AM
7	12	10	4	7	8	6:30:21 AM
8	2	14	12	16	3	7:07:05 AM
9	12	5	6	8	3	4:10:17 AM
10	2	7	8	13	12	2:09:15 AM
11	3	4	7	8	12	9:17:19 AM
12	23	12	3	4	7	3:10:45 AM
13	3	5	12	6	8	12:00:30 PM
14	4	8	9	12	13	6:20:25 AM
15	3	6	8	9	12	7:38:10 AM
16	12	14	4	6	8	5:09:10 AM
17	11	3	7	9	10	7:23:41 AM
18	10	12	15	7	11	4:01:10 AM
19	2	7	11	13	11	9:10:20 AM
20	12	13	14	15	12	1:01:10 AM
21	6	8	1	9	3	11:17:10 AM
22	7	9	3	0	1	12:30:20 PM
23	2	14	12	16	3	7:07:05 AM

Figure 4. Original Database

2) Watermark Insertion(Numeric Field) **3)Watermark Insertion(Time)**

```
4,7,9,3,1,1
8,3,15,13,17,3
12,23,13,3,5,7
16,13,15,5,7,9
20,13,13,15,15,13
24,5,7,3,3,9
28,13,11,5,7,9
32,3,5,7,9,13
36,23,13,3,5,7
40,7,9,3,1,1
44,5,9,13,5,9
48,1,5,7,3,3
52,5,7,1,3,5
56,5,9,13,5,9
60,13,5,7,9,3
64,3,5,13,7,9
68,13,19,3,1,7
72,5,3,9,13,13
76,11,3,3,3,11

4,6,8,2,0,0
8,2,14,12,16,2
12,22,12,2,4,6
16,12,14,4,6,8
20,12,12,14,14,12
24,4,6,2,2,8
28,12,10,4,6,8
32,2,4,6,8,12
36,22,12,2,4,6
40,6,8,2,0,0
```

```
12:5:20
02:5:8
09:5:20
12:5:20
12:5:20
02:5:8
12:5:20
02:5:8
01:5:20
07:5:20
12:5:20
02:5:8
19:5:32
06:5:56
08:5:4
,,,,,,,
12:4:20
02:4:8
09:4:20
12:4:20
12:4:20
02:4:8
12:4:20
02:4:8
01:4:20
07:4:20
12:4:20
02:4:8
19:4:32
06:4:56
```

Figure 5. Watermark Insertion (Numeric & Time)

4) Watermark Extraction(Numeric) **5) Watermark Extraction(Time)**

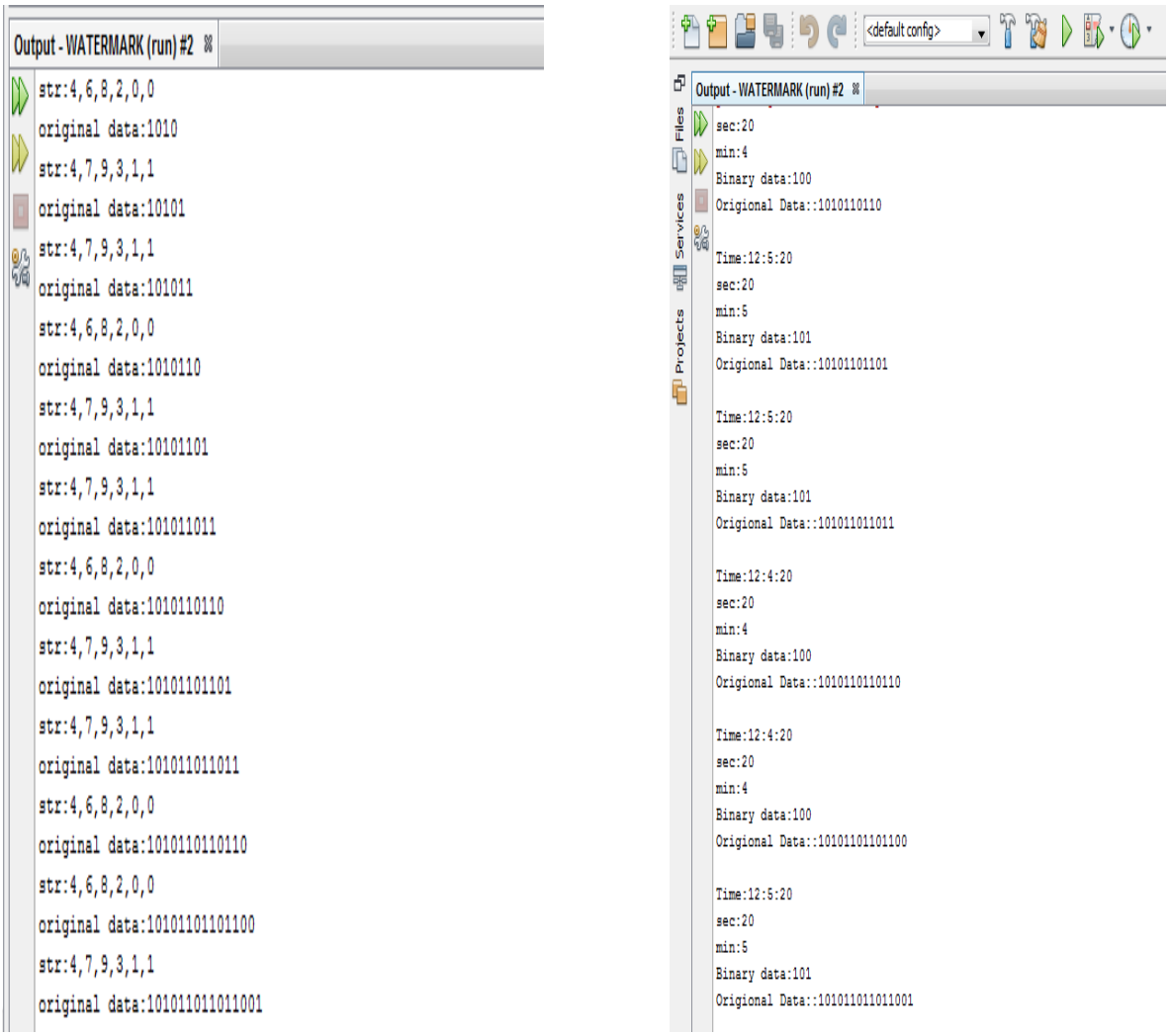


Figure 6. Watermark Extraction (Numeric & Time)

4. CONCLUSION

Database watermarking is link between three areas, database, security, and image processing. As watermarking is a basic concept introduced in image processing then it is used for security purposes like copy right protection and now days it is used in database as well because now a days with the development of world wide web databases are shared so easily and they required to be copy right protected to stop its misuse.

Here, we have proposed the new approach for the multiplace watermarking. We have used the speech as watermark. Since the watermark will be applied multiple places hence there is less chances to get attacked.it is comparatively easy to extract original watermark.

REFERENCES

[1] Brijesh B. Mehta, Udai Pratap Rao*," A Novel approach as Multi-place Watermarking For Security in Database".* in proceedings of SAM'11 - 10th International Conference on Security and Management at Las Vegas, Nevada, USA, pp. 703-707, July-2011

[2] H. Wang, X. Cui and Z. Cao, "A speech based Algorithm for Watermarking Relational Database", IEEE, ISIP, pp. 603-606, 2008.

[3] R. Agrawal, P.J. Haas, and J. Kiernan. Watermarking Relational Data: Framework, Algorithms and Analysis. The VLDB Journal. 2003, 12(2): 157~169

[4] Z. Qin, Y. Ying, L. Jia-jin, and L. Yi-shu."Watermark Based Copyright Protection of Outsourced Database". In Proceedings of the 10th International Database Engineering and Applications Symposium (IDEAS'06). Delhi, India. IEEE Computer Society. 2006, 301~308.

[5] Y. Zhang, X.M. Niu and D.N. Zhao."A Method of Protecting Relational Databases Copyright with Cloud Watermark". International Journal of Information Technology. 2004,1(4):206~210.

