RRW: A Novel Watermarking Technique for Relational Data

Abhilash Shukla
B.E (I.T) Student
BVCOE&RI
Anjaneri, Nashik.

Swapnil Bhambar
B.E (I.T) Student
BVCOE&RI
Anjaneri, Nashik

Harshala Patil
B.E (I.T) Student
BVCOE&RI
Anjaneri, Nashik

K. S. Kumavat
HOD (IT)
BVCOE&RI
Anjaneri, Nashik

Harshada Ghyar
B.E (I.T) Student
BVCOE&RI
Anjaneri, Nashik.

Seema Desale
B.E (I.T) Student
BVCOE&RI
Anjaneri, Nashik

ABSTRACT

In real world, a huge amount of data of various forms such as audio, video, text, etc. is transmitted over internet. In relational database, it is easy to recover structured data but difficult to recover unstructured data. These databases are used in collaboration for extracting information. However they are vulnerable to security threats and malicious attacks. Watermark Technique is used to recognize pattern and identify authenticity of data[1]. Watermarking techniques provides ownership protection over relational database and prevents the data from getting corrupted, but such methods are not resilient and hence the system uses a robust and reversible watermarking technique which provides watermark encoding and decoding. Robust means resilient to any attacks. Reversible watermarking technique ensures data quality with data recovery and also prevents the data from being tampered. Though any changes are made to the data by the attacker (like insertion, deletion, alteration), data is fully recovered with the use of robust and reversible watermarking techniques. Robust and reversible Watermarking Techniques has provided security to the digital data by marking the data which is unique and can be used for claiming ownership of data. Using Genetic algorithm feature analysis and selection is done and then a watermark is created. After that the data is passed to the attacker channel where attacks take place, but due to Robust and Reversible Watermarking technique, data is recovered completely without any loss.

Keywords

RRW, Watermark, Robust, Reversible, Relational Database, Recovery.

1. INTRODUCTION

In recent times, a large amount of data is generated because of growth of internet and cloud computing[1]. Availability of data is in various formats. Reversible Watermarking techniques allows data recovery and provides ownership protection.it provides the ownership protection by marking format such as images, audio, and relational databases .A large number of organizations today have relational database and their security is of utmost importance. Reversible Watermarking techniques allows enforcement of ownership rights and prevents data from being tampered. As data is available in various formats out of which relational data is structured which is difficult to retrieve as compared to multimedia data. Some primitive techniques were use such as Cryptography, Fingerprinting, and Steganography. These techniques however are not robust however achieving robustness is a very difficult task for these reversible watermarking technique is used .Some of the earlier watermarking techniques are as follows:-Histogram Techniques Problem:-In that system firstly Histogram technique is used. But at time of heavy attack this technique is fully exposed. In histogram, by considering a method of distribution of error between two distributed variables and selected some initial nonzero digits of errors to form histograms. For authenticating data quality, Histogram technique is keep track of overhead information. Histogram technique is not robust against heavy attacks.. Reversible watermarking technique prevents data quality from getting degraded. Difference Expansion Watermarking Technique. : This technique is better than Histogram technique, but also having some drawbacks. This technique exploits methods of arithmetic operations on numeric features and performs transformations[3]. The watermark information is normally grouped in the Least Significant Bite of features of relational databases to minimize distortions. . But, in RRW, a GA based optimum value is embedded in the selected feature of the dataset with the objective of preserving the information and data quality while reducing the data distortions as a result of watermark embedding. Another reversible watermarking technique considered is depend on difference expansion and support vector regression (SVR) prediction to protect the database from being tampered. This technique is similar only exception as Difference Expansion, only difference is that it uses support vector regression. The design of these techniques is to provide and ensure ownership proof. Such watermark techniques are vulnerable to modification attacks as any change or modification in the expanded value will not able to detect watermark information and the original data. This technique is not able to recover original data. Technique used to solve problem:- System is not able to work correctly in heavy attacks. Also fail to detect watermark information and the original data. In order to overcome these problems, A difference expansion watermarking technique is used which is based upon genetic algorithm. This is proposed reversible and robust solution for database. This technique improves upon the drawbacks mentioned above by minimizing distortions in the data,and increasing watermark capacity .

2. SYSTEM ARCHITECTURE

Following Fig shows RRW Architect;

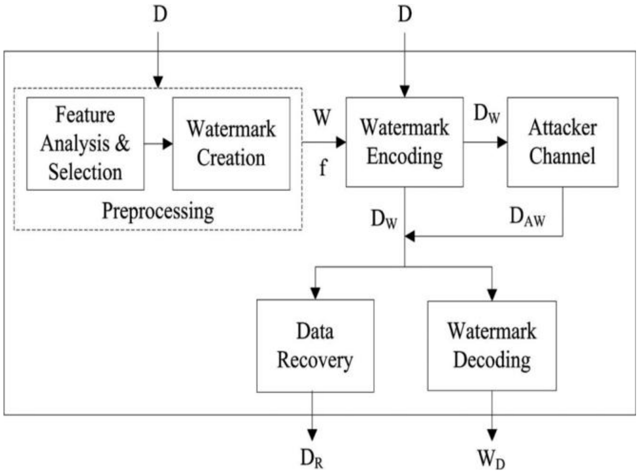


Figure 1: RRW Architecture.

In this phase, the study of RRW for reversible watermarking of relational databases that improves data recovery ratio. The main architecture of RRW is presented above RRW contains the four major phases: (i)Watermark preprocessing (ii) Watermark encoding (iii) Watermark decoding and (iv) Data recovery. For calculation of an optimal watermark computes different parameter in watermark preprocessing phase. These parameters or values are used for both encoding and decoding of watermark. The main hot spot of watermark encoding .phase is to grouped watermark information and data in such a way that it does not affect the quality of data. In watermark embedding phase , data gets changed with respect to the allocated bandwidth or capacity of the watermark information. Watermark bandwidth should be sufficiently large to ensure robustness but also it not having that large which destroys the data quality. The data owner decides the amount of data modification and quality is not compose for a particular database application before-hand and therefore defines usability constraints _ to introduce tolerable distortion into the data. In paper, the datasets are used which are taken from UCI machine learning repository, which include, Cleveland Heart Disease dataset, MAGIC Gamma Telescope dataset, and PAMAP2 Physical Activity Monitoring dataset[3]. The purpose of using datasets for evolution and that RRW is not dependent on any particular dataset or feature. Determination of Suitable feature is done to embed watermark on the basis of mutual information. The data is released to the intended or original receiver over a internet that is consider not secure and termed as the "attacker channel". In attacker channel data undergo several malicious attacks. The ability and capability of RRW is to resist subset insertion, alteration and deletion attacks. Watermark decoding is responsible to recover information.

It consists of following features:

2.1 Watermark Preprocessing Phase

Feature Analysis and Selection:-In this, analyzing feature related to data and select appropriate feature for watermarking creation. For creating an information model for various features of the dataset, all the features are numbered according to their priority in information extraction on other features. Mutual information (MI) is used for only this purpose. For computations of mutual dependence of two variables, it is an important statistical measure. Features with the lowest MI are predicted by attacker, in an attempt to identify which feature has been watermarked. For choosing the feature, a secret threshold can be used for watermark embedding to prevent the attacker in this particular scenario. In this context, the data owner can define a secret threshold which is based on Mutual Information which is MI of all the features in the database. Features selected for watermarking are having lower MI than threshold. Attacker never attacks the features which is having large MI. If the attacker attack on large feature, the data will adjusted. Therefore, attacker will be forced to attack the lower MI features and it not uses secret threshold for watermarked the features. Watermark Creation through Genetic Algorithm (GA)For the creation of peak watermark information, it needs to be enclosed in the original data, we use an evolutionary technique; GA inspired from genetic evolution. To an optimization problem Genetic Algorithm is a potential solution by searching the possible solution space. In the search of optimal solution, to evolve a population of chromosomes Genetic Algorithm follows constant mechanism. Selection, crossover, mutation and replacement are the application for essential information which is preserves by GA. To evaluate the quality of each candidate chromosomes GA employing fitness function.

2.2 Watermark Encoding:

To meet the data quality constraint of the data owner, watermark information calculation is formulated as a CO problem. A Genetic Algorithm is used to create optimal watermark information that having: (1) Optimal chromosomal string (length l) and (2) value b is a parameter that is computed using GA and a tolerable amount of change is enclosed in the feature values. Once the optimum value of b for each candidate features A is found, it is saved for watermark encoding and decoding. The length l of watermark bit and an optimum value b is used to operate the data provided it satisfies the usability constraints _. The value b is added into every tuple which is related to selected features. If a selected features given bit is not 0, then its value is subtracted from the original value of the features. It is ensured that the mutual information remain of a feature. At the time of inserting watermark into the database, information of features want to unchanged. Also the watermark is added into each and every tuple of the dataset for selected features. Data owner can select any number of features for watermark embedding using secret threshold.

2.3 Watermark Decoding:

In the watermark decoding, firstly, marked feature have been locate. Optimization process through Genetic Algorithm is not required in Watermark Decoding. Watermark decoder z is used to calculate the amount of change in the value of a feature that doesn't affect the data quality. In the decoding phase, hdr is calculated using Equation and represents the percent change detected in the watermarked data. The watermark decoder is used to decode the watermark by working on only one bit at a time. The value of hdr, hDr is calculated using the values of tuple r and therefore might be different for every r. The parameters hDr is computed by calculating the difference between the original data change amount hDr and the watermark detected change amount is calculated.

2.4 Data Recovery

After identifying the watermark code, error correction and data recovery are carried out using some steps. Computing optimized value of b through the Genetic Algorithm is used for regeneration of original information and data. The data recovery algorithm is used for recover data. The value of a numeric feature is recovered using Equation .After data preprocessing phase, it goes to the data encoding phase. In data encoding phase data is encoded. After encoding of data attacker tries to access data and modified it. The attacker channels try to alteration, deletion and insertion of subset, this attacks generated by the adversary. These attacks change or modify the original data and try to poor its quality. In the watermark decoding the grouped or embedded watermark is decoded from the suspicious data. In order to achieve this, feature selection and watermark creation step is performed again, and strategies (feature selection on the basis of MI, b the optimized value from the Genetic Algorithm and hr the change matrix) are decoded to recover the watermark. Semi-blind nature of RRW is used for data reversibility[4]. Also this technique used in heavy attacks (attacks that may target large number of tuples). Data recovery phase main focus is to recover original data, using post processing steps for error correction and recovery.

3. ALGORITHMS

Algorithmic strategy contains 3 algorithms for encoding, decoding and for data recovery as follows:

Algorithm for Watermark Encoding

For inserting watermark into data stream use watermark encoding algorithm.

```
Input: D, ω, β
Output:  $D_{w,}$
for ω =1 to l do
//loop will iterate for all watermark bits  ω from 1 to
length l of  the watermark
for  r =1 to R do
//loop will iterate for all tuples of the data
if  b, ω = = 0 then
//the case when the watermark bit is 0 .
Changes are calculated by  $ηr =D_r * ς$
data is watermarked by using Equation  $Dw_r =D_r - β$
if  br, ω = =1 then
// the case when the watermark bit is 1.
changes are   calculated by ηr  = Dr  * ς
data is watermarked by using Equation  ηr  =Dr + β
insert ηr into
end  if
end for
end for
return   DW, .
```

Algorithm for Watermark Decoding

For decoding least significant bit of the stream using watermark use watermark decoding algorithm.

```
Input: $D_W$ or  D’w,  l
Output: $W_D$
for  r = 1to R do
//loop will iterate for all tuples of the data
for b = l to 1 do
//loop will iterate for all watermark bits b from 1 to
length  l  of the watemark.
ηdr  (r)  * ς
ηΔr = ηdr - ηr
If  ηΔr   0 then
Detected watermark bit (dtW) is 1
Else if  ηΔr > 0 and  ηΔr    1 then
Detected  watermark bit   (dtW) is 0
end  if
end  for
end  for
WD = mode (dtW(1,2,…..,l))
return $W_D$
```

Algorithm for Data Recovery:

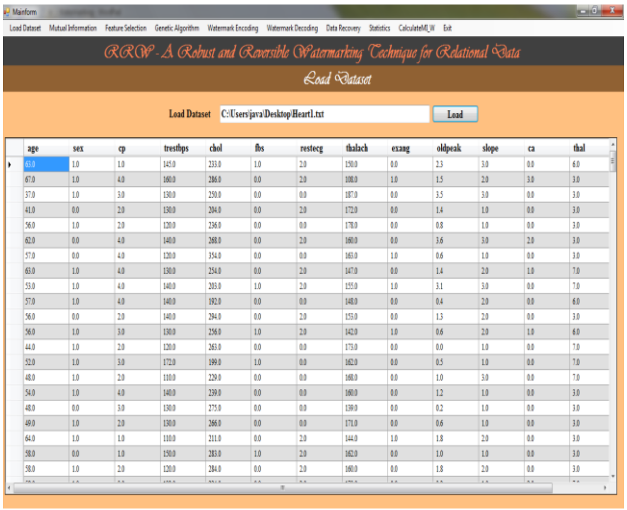
For recovery of original data from watermark data use watermark recovery algorithm.

```
Input : DW or  D’w, b
Output: $D_r$
for r = 1 to R do
//loop will iterate for all tuples of the data
for b = l to 1 do
//loop will iterate for all watermark bits b from 1 to    length
of the watermark.
If dtW(r,b) = = 1then
//0 or 1 watermark bit is detected from every tuple
data is recovered by  $Dr= dw_r+ β$
else
data is recovered by   $Dr=dw_r-β$
else
end if
end for
end  for
return  D
```

Table 1: Algorithm Description

Symbol	Description	Symbol	Description
D	Original data	b	Watermark bit
Dw	Watermarked data	$μ_d$	Mean of original data using PEEW technique
R	No of tuples	$μ_D$	Mean of original data of RRW
ω	Watermarked bits	a	Value of feature
ηr	Percentage changing in encoding	r	A tuple in database table
l	Length of database	ηΔr	Difference in change in values during encoding and decoding
Dr	Recovered data	dwr	Data recovered withwatermarkbits

4. IMPLEMENTATION DETAILS



Load Dataset C:\Users\java\Desktop\Heart1.txt

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0
67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0
37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0
41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0
56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0
62.0	0.0	4.0	140.0	268.0	0.0	2.0	160.0	0.0	3.6	3.0	2.0	3.0
57.0	0.0	4.0	120.0	354.0	0.0	0.0	163.0	1.0	0.6	1.0	0.0	3.0
63.0	1.0	4.0	130.0	254.0	0.0	2.0	147.0	0.0	1.4	2.0	1.0	7.0
53.0	1.0	4.0	140.0	203.0	1.0	2.0	155.0	1.0	3.1	3.0	0.0	7.0
57.0	1.0	4.0	140.0	192.0	0.0	0.0	148.0	0.0	0.4	2.0	0.0	6.0
56.0	0.0	2.0	140.0	294.0	0.0	2.0	153.0	0.0	1.3	2.0	0.0	3.0
56.0	1.0	3.0	130.0	256.0	1.0	2.0	142.0	1.0	0.6	2.0	1.0	6.0
44.0	1.0	2.0	120.0	263.0	0.0	0.0	173.0	0.0	0.0	1.0	0.0	7.0
52.0	1.0	3.0	172.0	199.0	1.0	0.0	162.0	0.0	0.5	1.0	0.0	7.0
48.0	1.0	2.0	110.0	229.0	0.0	0.0	168.0	0.0	1.0	3.0	0.0	7.0
54.0	1.0	4.0	140.0	239.0	0.0	0.0	160.0	0.0	1.2	1.0	0.0	3.0
48.0	0.0	3.0	130.0	275.0	0.0	0.0	139.0	0.0	0.2	1.0	0.0	3.0
49.0	1.0	2.0	130.0	266.0	0.0	0.0	171.0	0.0	0.6	1.0	0.0	3.0
64.0	1.0	1.0	110.0	211.0	0.0	2.0	144.0	1.0	1.8	2.0	0.0	3.0
58.0	0.0	1.0	150.0	283.0	1.0	2.0	162.0	0.0	1.0	1.0	0.0	3.0
58.0	1.0	2.0	120.0	284.0	0.0	2.0	160.0	0.0	1.8	2.0	0.0	3.0

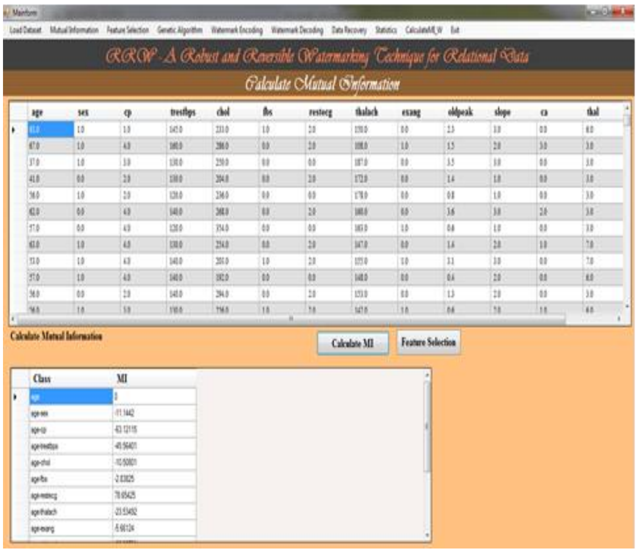
Figure 2: Load Dataset

In Main form following menus are available :load dataset, mutual information, feature selection, Genetic algorithm, Watermark Encoding, Watermark Decoding, Data Recovery ,Statistics, Calculate MI_W, Exit.

It consists of following steps:

Step 1: Firstly ,we load dataset of Medical history of heart patient ,then following features are loaded as follows :age ,sex ,cp ,trestbps ,chol, fbs ,restec, oldpeak,slope,ca,thal.

Step 2: Feature selection:.In this step, features are selected using knowledge discovery , features are age ,sex, chol ,fbs .these features are then applied to the model.

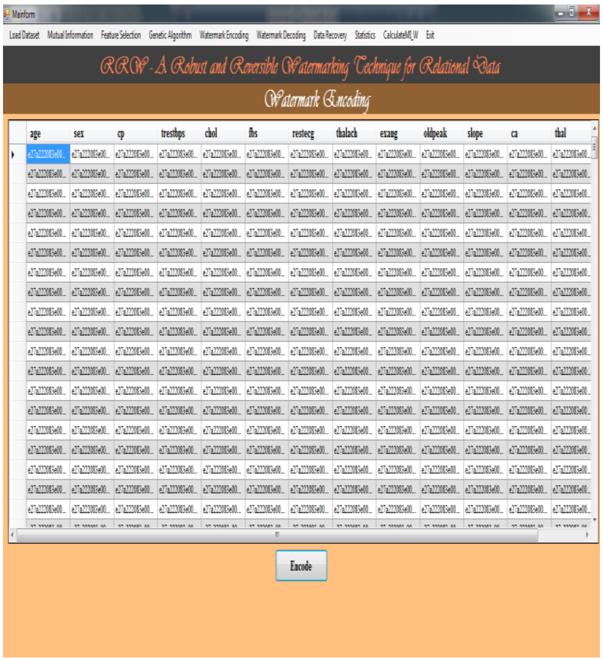


age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0
67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0
37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0
41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0
56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0
62.0	0.0	4.0	140.0	268.0	0.0	2.0	160.0	0.0	3.6	3.0	2.0	3.0
57.0	0.0	4.0	120.0	354.0	0.0	0.0	163.0	1.0	0.6	1.0	0.0	3.0
63.0	1.0	4.0	130.0	254.0	0.0	2.0	147.0	0.0	1.4	2.0	1.0	7.0
53.0	1.0	4.0	140.0	203.0	1.0	2.0	155.0	1.0	3.1	3.0	0.0	7.0
57.0	1.0	4.0	140.0	192.0	0.0	0.0	148.0	0.0	0.4	2.0	0.0	6.0
56.0	0.0	2.0	140.0	294.0	0.0	2.0	153.0	0.0	1.3	2.0	0.0	3.0

Class	MI
age	0
age-sex	-11.1442
age-cp	63.12115
age-trestbps	-45.56401
age-chol	-10.50801
age-fbs	-2.03025
age-restecg	78.65425
age-thalach	-23.53492
age-exang	-5.66124

Figure 2: MI Calculation

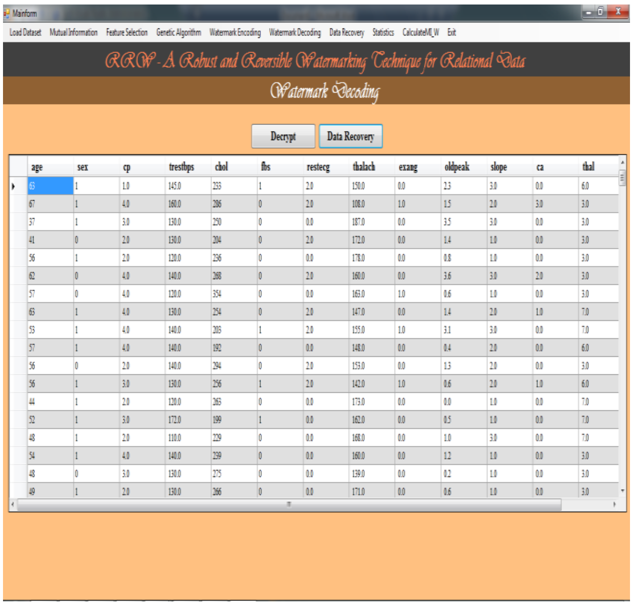
Step 3:Here, first an optimize value is selected using an optimization scheme, mutual information of every features with all other feature is given by mutual information(MI)MI(A,B)=ΣaΣbPAB(a,b)logPAB(a,b) /PA(a)PB(b)where MI(A,B) measures the degree of correlation of a feature measuring marginal probability distribution, PA(a)PB(b) and the Joint Probability Distribution and accordingly features are selected.



age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...
e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...	e27a222083e00...

Figure 3: Watermark Encoding.

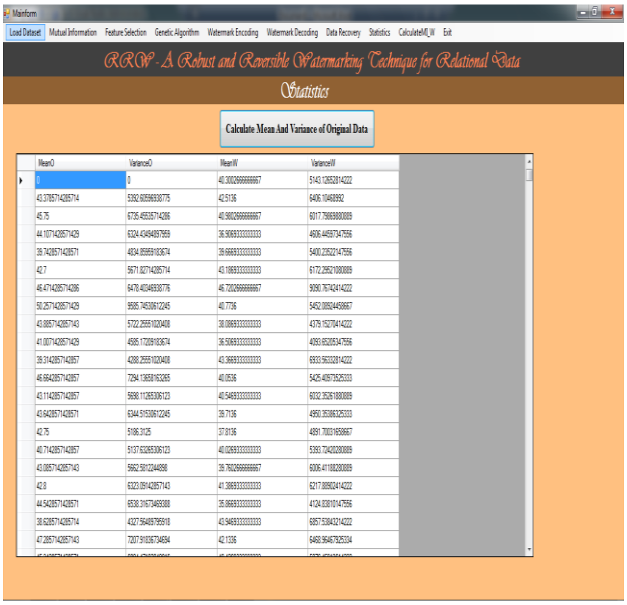
Step 4: Watermark Encoding; after selection of features using MI, next we encode this features. In this phase, watermark information is inserted in tuple .it does not affects the data quality for creating optimal watermark we use genetic algorithm. In optimal watermark we use optimal chromosomal string and β values and β value is computed using genetic algorithm it used for watermark encoding.in this feature value is not affected.



age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63	1	1.0	145.0	233	1	2.0	150.0	0.0	2.3	3.0	0.0	6.0
67	1	4.0	160.0	286	0	2.0	108.0	1.0	1.5	2.0	3.0	3.0
37	1	3.0	130.0	250	0	0.0	187.0	0.0	3.5	3.0	0.0	3.0
41	0	2.0	130.0	204	0	2.0	172.0	0.0	1.4	1.0	0.0	3.0
56	1	2.0	120.0	236	0	0.0	178.0	0.0	0.8	1.0	0.0	3.0
62	0	4.0	140.0	268	0	2.0	160.0	0.0	3.6	3.0	2.0	3.0
57	0	4.0	120.0	354	0	0.0	163.0	1.0	0.6	1.0	0.0	3.0
63	1	4.0	130.0	254	0	2.0	147.0	0.0	1.4	2.0	1.0	7.0
53	1	4.0	140.0	203	1	2.0	155.0	1.0	3.1	3.0	0.0	7.0
57	1	4.0	140.0	192	0	0.0	148.0	0.0	0.4	2.0	0.0	6.0
56	0	2.0	140.0	294	0	2.0	153.0	0.0	1.3	2.0	0.0	3.0
56	1	3.0	130.0	256	1	2.0	142.0	1.0	0.6	2.0	1.0	6.0
44	1	2.0	120.0	263	0	0.0	173.0	0.0	0.0	1.0	0.0	7.0
52	1	3.0	172.0	199	1	0.0	162.0	0.0	0.5	1.0	0.0	7.0
48	1	2.0	110.0	229	0	0.0	168.0	0.0	1.0	3.0	0.0	7.0
54	1	4.0	140.0	239	0	0.0	160.0	0.0	1.2	1.0	0.0	3.0
48	0	3.0	130.0	275	0	0.0	139.0	0.0	0.2	1.0	0.0	3.0
49	1	2.0	130.0	266	0	0.0	171.0	0.0	0.6	1.0	0.0	3.0

Figure 4: Watermark Decoding.

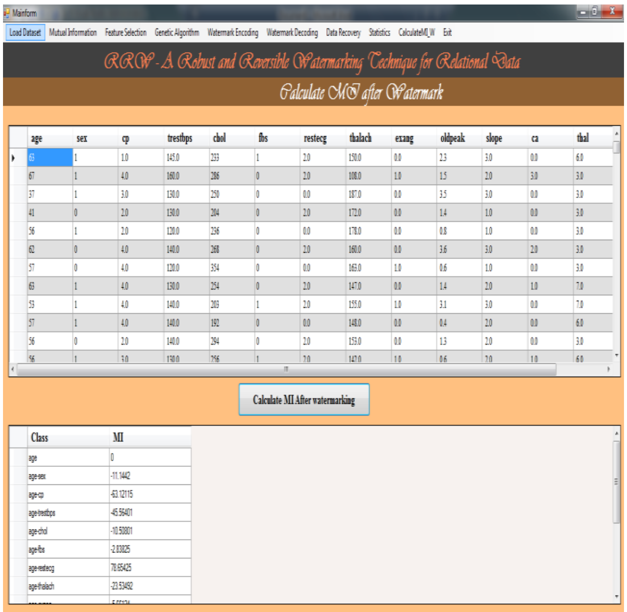
Step 5:Here the bits are decoded from lsb i.eleast significant bit to msb i.e is most significant bit.it works in reverse manner as compared to watermark encoding.



MeanO	VarianceO	MeanW	VarianceW
0	0	40.3002666666667	5143.12652814222
43.3785714285714	5392.60596938775	42.5136	6406.1046892
45.75	6735.45535714286	40.9802666666667	6017.79869880889
44.1071428571429	6324.43494897959	36.9069333333333	4606.44597347556
39.7428571428571	4834.85959183674	39.6669333333333	5400.23522147556
42.7	5671.82714285714	43.1869333333333	6172.29521080889
46.4714285714286	6478.40346938776	46.7202666666667	9090.76742414222
50.2571428571429	9585.74530612245	40.7736	5452.08924458667
43.8857142857143	5722.25551020408	38.0869333333333	4379.15270414222
41.0071428571429	4585.17209183674	36.5069333333333	4093.65205347556
39.3142857142857	4288.25551020408	43.3669333333333	6933.56332814222
46.6642857142857	7294.13658163265	40.0536	5425.40973525333
43.1142857142857	5698.11265306123	40.5469333333333	6032.35261880889
43.6428571428571	6344.51530612245	39.7136	4950.35386325333
42.75	5186.3125	37.8136	4891.70031658667
40.7142857142857	5137.63265306123	40.0269333333333	5393.72420280889
43.0857142857143	5662.5812244898	39.7602666666667	6006.41188280889
42.8	6323.09142857143	41.3869333333333	6217.88902414222
44.5428571428571	6538.31673469388	35.8669333333333	4124.83810147556
38.6285714285714	4327.56489795918	43.9469333333333	6857.53843214222
47.2857142857143	7207.91836734694	42.1336	6468.96467925334

Figure 5: Mean and Variance calculation

Step 6: In this mean and variance of original data is calculated.



age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63	1	1.0	145.0	233	1	2.0	150.0	0.0	2.3	3.0	0.0	6.0
67	1	4.0	160.0	286	0	2.0	108.0	1.0	1.5	2.0	3.0	3.0
37	1	3.0	130.0	250	0	0.0	187.0	0.0	3.5	3.0	0.0	3.0
41	0	2.0	130.0	204	0	2.0	172.0	0.0	1.4	1.0	0.0	3.0
56	1	2.0	120.0	236	0	0.0	178.0	0.0	0.8	1.0	0.0	3.0
62	0	4.0	140.0	268	0	2.0	160.0	0.0	3.6	3.0	2.0	3.0
57	0	4.0	120.0	354	0	0.0	163.0	1.0	0.6	1.0	0.0	3.0
63	1	4.0	130.0	254	0	2.0	147.0	0.0	1.4	2.0	1.0	7.0
53	1	4.0	140.0	203	1	2.0	155.0	1.0	3.1	3.0	0.0	7.0
57	1	4.0	140.0	192	0	0.0	148.0	0.0	0.4	2.0	0.0	6.0
56	0	2.0	140.0	294	0	2.0	153.0	0.0	1.3	2.0	0.0	3.0
56	1	3.0	130.0	256	1	2.0	142.0	1.0	0.6	2.0	1.0	6.0

Class	MI
age	0
age-sex	-11.1442
age-cp	-63.12115
age-trestbps	-45.56401
age-chol	-10.50801
age-fbs	-2.83825
age-restecg	78.65425
age-thalach	-23.53492

Figure 6: MI after watermarked data

Step 7: In Statistics, the mean and variance of data is calculated and their changes are observed. Here results of the reported mean and variance and watermarked mean and variance is compared and accordingly data is classified using data mining purposes the data is classified on the basis of cardiac symptoms where 0 shows absence of disease and 1,2,3,4, shows various types of cardiac disease .

5. CONCLUSION

Reversible watermarking techniques are used to cater to such scenarios because they are able to recover original data from watermarked data and ensure data quality to some extent. This paper, a novel robust and reversible technique for watermarking numerical data of relational databases is presented. The main contribution of this work is that it allows recovery of a large portion of the data even after being subjected. to malicious attacks. Evaluation of RRW is done where the watermark is detected with maximum decoding accuracy in different scenarios through attack analysis. Experiments shows that if an intruder deletes, alter upto 50 %of tables, recovery of original information can be done along with embedded watermark. One of our future concerns is to watermark shared databases in distributed environments where different members share their data in various proportions. In future it can be applied to non numeric data also.

6. REFERENCES

[1] Y.-C. Liu, Y.-T. Ma, H.-S. Zhang, D.-Y. Li, and G.-S. Chen, "A method for trust management in cloud computing: Data coloring by cloud watermarking," Int. J. Autom. Comput., vol. 8, no. 3, pp. 280–285, 2011.

[2] F. A. Petitcolas, "Watermarking schemes evaluation," IEEE SignalProcess. Mag., vol. 17, no. 5, pp. 58–64, Sep. 2000.

[3] J. T. Brassil, S. Low, and N. F. Maxemchuk, "Copyright protection for the electronic distribution of text documents," Proc. IEEE, vol. 87, no. 7, pp. 1181–1196, Jul. 1999.

[4] R. Agrawal and J. Kiernan, "Watermarking relational databases,"in Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 155–166.

[5] R. Sion, M. Atallah, and S. Prabhakar, "Rights protection for cate gorical data," IEEE Trans. Knowl. Data Eng., vol. 17, no. 7, pp. 912-926, Jul. 2005.

[6] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans. Image Pro- cess., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[7] I. Cox, M. Miller, J. Bloom, and M. Miller, Digital Watermarking. Burlington, MA, USA: Morgan Kaufmann, 2001.

[8] P. W. Wong, "A public key watermark for image verification and authentication," in Proc. IEEE Int. Conf. Image Process., 1998, vol. 1, pp. 455–459.