



Recoverable Watermarking for Distributed Databases

Sana Hatture, Prof. Dr. Suhas D. Raut

ME Student, Dept. of Computer Science & Engineering, NK Orchid College of Engineering & Technology, Solapur
University, Solapur, Maharashtra, India

Project Guide, Dept. of Computer Science & Engineering, NK Orchid College of Engineering & Technology, Solapur
University, Solapur, Maharashtra, India

ABSTRACT: Distributed database security has become an important issue. This paper provides method for providing security to shared databases in distributed environment. Such approach is applied for protecting numerical data in relational databases. Advancement in information technology is playing an increasing role in the use of information systems comprising relational databases. These databases are used effectively in collaborative environments for information extraction; consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Watermarking is advocated to enforce ownership rights over shared relational data and for providing a means for tackling data tampering. The system provides security to the shared databases in distributed environment by recoverable watermarking where different users share their data in various proportions. It also preserves and maintains data quality.

KEYWORDS: Distributed Database, Relational Database, Recoverable Watermarking and Numerical Data

I. INTRODUCTION

Security of relational database is great concern in today's world because of sharing of data over internet. Data providers create services and make them available to users for searching and accessing purposes. Given that these services may attract more attacks. So it is desire to protect the data and hence providers need some technology that identify the threats and pirated copies unauthorized access to their databases. The increasing use of relational database creates a need for watermarking database. In today's internet based application environment, ownership rights protection on relational database is decisive issue because unauthorized changes to data may have serious consequences and result in significant losses for the organization. Hence right protection through watermarking becomes an important research topic [1]. Following is existing system:

- Robust and Reversible Watermarking (RRW): Genetic algorithm (optimization algorithm) is employed in the robust and reversible watermarking technique (RRW) to achieve an optimal solution that is feasible for the problem at hand and does not violate the defined constraints. An optimal watermark value is created through the GA and inserted into the selected feature of the relational database in such a way that the data quality remains intact. Mutual Information, a well-known information theory (concept), statistically measures the amount of information that one feature contains about the other features in a database. In RRW, mutual information is used to select a suitable (candidate) feature from the database for watermarking. In RRW, the knowledge of mutual information for every candidate feature is also employed to compute the watermark information. Thus, it is ensured that the data quality will not be affected. The RRW is based on recoverable watermarking numerical data of centralized relational databases [8]. Therefore implemented system provides database security in distributed environment providing recoverable watermarking.

The system provides the ownership protection by performing recoverable watermarking on numerical data in relational databases for distributed environment. The goal is how to embed watermark to numerical data known only by the data owner in order to prove the ownership of the data without lossless of its quality.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

II. RELATED WORK

In the digital world of today, data is excessively being generated due to the increasing use of the Internet and Cloud Computing. Data is stored in different digital formats such as images, audio, video, natural language texts and relational data. Relational data in particular is shared extensively by the owners with research communities and in virtual data storage locations in the Cloud. The purpose is to work in a collaborative environment and make data openly available so that it is useful for knowledge extraction and decision making. Following are the techniques to ensure security in terms of ownership protection.

- Watermarking techniques have historically been used to ensure security in terms of ownership protection and tamper proofing for a wide variety of data formats. This includes images, audio, video, natural language processing software, relational databases [2] and more.
- Reversible watermarking techniques can ensure data recovery along with ownership protection.
- Fingerprinting (transactional watermarks), data hashing, serial codes are some other techniques used for ownership protection [3].

Following are some systems:

- The first irreversible watermarking technique for relational databases was proposed by Agrawal et al. [2]. Similarly, the first reversible watermarking scheme for relational databases was proposed by Zhang et al. [4]. In this technique, histogram expansion is used for reversible watermarking of relational database. Zhang et al. proposed a method of distribution of error between two evenly distributed variables and selected some initial nonzero digits of errors to form histograms. Histogram expansion technique is used to reversibly watermark the selected nonzero initial digits of errors. This technique keeps track of overhead information to authenticate data quality. However, this technique is not robust against heavy attacks.
- Difference Expansion Watermarking techniques (DEW), exploit methods of arithmetic operations on numeric features and perform transformations [5]. The watermark information is normally embedded in the LSB of features of relational databases to minimize distortions.
- Another reversible watermarking technique proposed is based on difference expansion and support vector regression (SVR) prediction to protect the database from being tampered. The intention behind the design of these techniques is to provide ownership proof. Such techniques are vulnerable to modification attacks as any change in the expanded value will fail to detect watermark information and the original data[6].
- Genetic Algorithm based on Difference Expansion watermarking (GADEW) technique is used in a proposed robust and reversible solution for relational databases [7]. GADEW improves upon the drawbacks mentioned above by minimizing distortions in the data, increasing watermark capacity and lowering false positive rate. To this end, a genetic algorithm is employed to increase watermark capacity and minimize introduced distortion. However, watermark capacity decreases with the increase in watermarked tuples. GADEW used the distortion measures to control distortions in the resultant data.

III. SYSTEM ARCHITECTURE

Below figure1, depicts system architecture of recoverable watermarking for distributed databases (RWDD). System uses relational database. Here users share their data in various proportions. It provides security to shared databases in distributed environment. This is done by recoverable watermarking shared databases. The RWDD system consists of four servers. These servers are connected to each other via LAN and connected to respective client directly. Here database is relational database. Database design starts from global schema and processed by designing fragmentation and allocating fragments to different servers. Data is distributed by fragmenting the data and storing at different servers. Here horizontal fragmentation is done. Each Replica is located at different servers. Each server add watermark to numerical data in relational database. Only authorized client can access watermarked data. Here read authorization is provided to client and log is maintained for read only access. When authorized client send request to server, server checks validity of client. For this server maintains client table and checks user identification number and password match. If match found then client is valid otherwise refused. The protocol communication between client and server is inter process communication call that is remote procedure call. Procedure call can be local or remote. Server gives reply with watermarked data when procedure call is local. For remote procedure call server forward client request to all servers and once it receives reply send result of requested procedure to client. Server is the owner of data. Fragmented data is stored at different servers. Then each server embeds watermark on numerical data of original relational database. This is done by selecting one numeric attribute. Here two different watermarked data are created of original data. First watermarked data is created by selecting one numeric attribute and second is created by selecting another numeric attribute. Both watermarked data is saved with different name. When client send data

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

request for first time, server responds with first watermarked data created. For second request, server responds with second watermarked data. Watermarked data is only known by server as it is the owner of data. Server also extracts the watermark by detecting the watermark. This is recoverable watermarking. It also preserves and maintains data quality by recoverable watermarking.

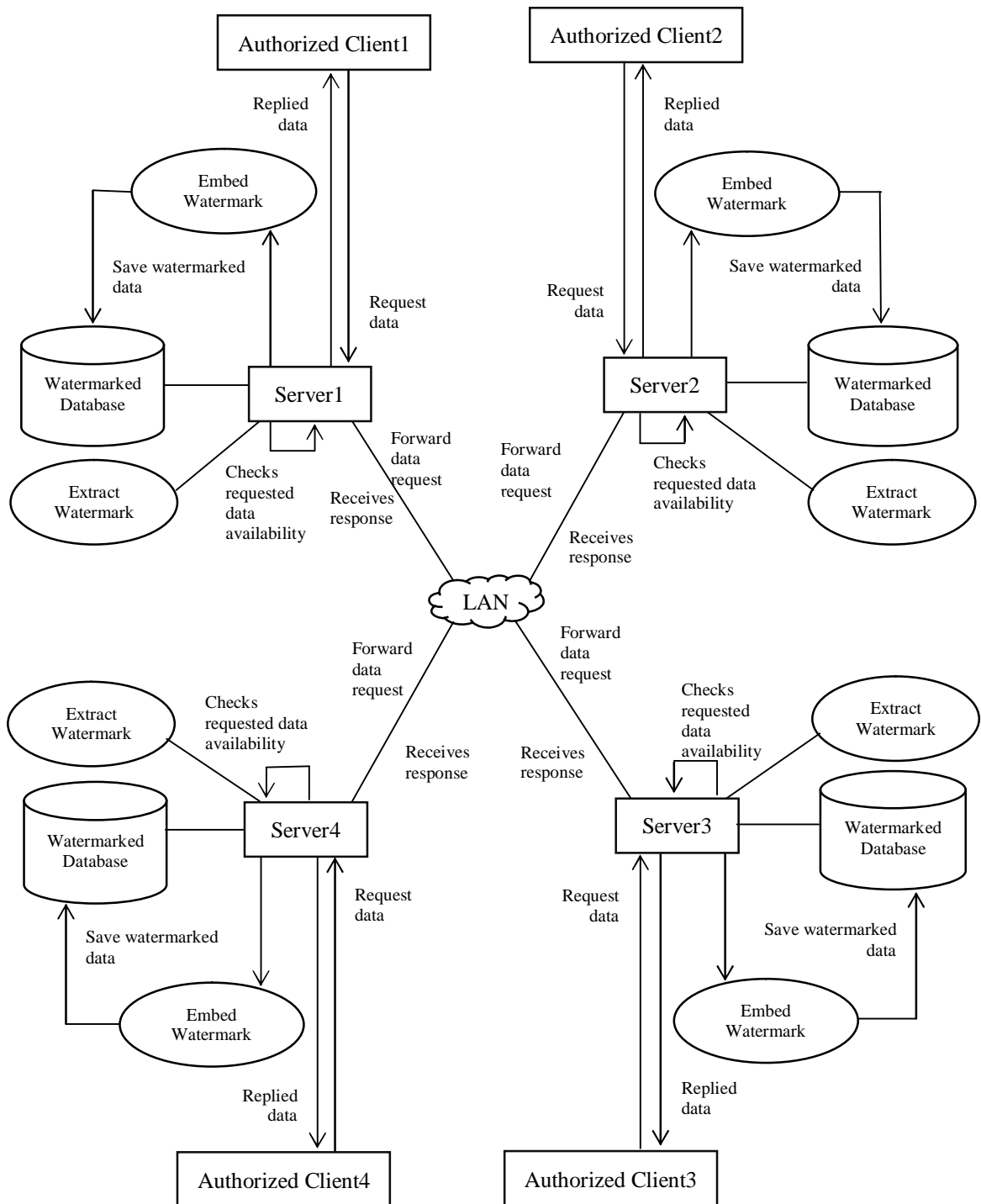


Figure 1: RWDD System Architecture



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

IV. IMPLEMENTATION METHOD

A. Description of the Implementation Method:

Following are the functional modules of the system:

1. Server Modules:

Server is the owner of data. First it creates the relational database which is to be watermarked to ensure ownership protection. Then data is distributed over servers. Here horizontal fragmentation is done. Then server performs two tasks: first embeds watermark in relational database by selecting one attribute of numeric data and server extracts watermark by detecting watermark embedded. This is recoverable watermarking. This also preserves and maintains quality of data. Server also performs authentication and authorization of client when it receives data request from client. For this server maintains different database for client. This database contains client details and log file. Client authentication is performed by user id and password. Server allows only read authorization to client for requested data.

i. Adding Watermark:

This module adds watermark to data. After creating and fragmenting database server adds watermark to data. First it selects one numeric data attribute which is to be watermarked. Numeric data column is taken from user input. Its parameters are calculated such as min, max, sum, mean and variance of original data. These are parameters before adding watermark. Watermarked string is generated through genetic algorithm and watermark length and beta is taken from user. Beta is watermarking strength, low values of beta preserves data quality. Watermark length, beta, database name, table name and watermarked table name, primary key column and data column are user inputs. After adding watermark again parameters such as min, max, sum, mean and variance of watermarked column is calculated to ensure watermark is embedded.

ii. Extract Watermark:

This is next module after adding watermark. This module extracts watermark. Watermark extraction is performed by detecting watermark. This is called recoverable watermarking. It also preserves and maintains quality of data. It depends on length of watermark.

2. Client Modules:

Client contains three functional modules registration of client, login of client and data request. New client registers on server. Already registered client logged in directly. After logging client send request to server for data. For this database name, table name and primary key column is taken as input from client. Then client send request to server and receives reply from server. This replied data is watermarked but client is unaware about this watermarked data.

V. IMPLEMENTATION TECHNIQUE

The implementation technique to embed watermark can be summarized in following steps:

- 1) Get the database name, table name, watermark table name, primary key column and numeric data column
- 2) Genetic Algorithm for generating watermark string
 - i. Set population size, crossover percentage, mutation rate, elite percentage, database name, table name, primary key column, data column
 - ii. Calculate min, max, sum, mean and variance of selected data column from original data
 - iii. Initialize population.
 - iv. Calculate fitness
 - v. Sort population
 - vi. Select elite chromosomes
 - vii. Create new chromosomes using crossover
 - viii. Mutate chromosome
 - ix. Goto step (iv) if end condition is not met
 - x. Generate watermark string from best chromosome
- 3) Add watermark string to selected data column

Following are steps of the proposed technique to extract watermark:

- 1) Get watermark length, database name, table name, watermarked table name, primary key column and data column.
- 2) Calculate difference of values in data column of original data and watermarked data
- 3) Add support vote for each watermark bit

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

4) Build watermark string by counting votes for each bit

Following are the steps of data request from client to server:

- 1) If new client then register else login
- 2) Get database name, table name and primary key column.
- 3) Send request to server
- 4) Receives reply from server in the form of watermarked data.

VI. RESULTS

System first adds watermark to original data. It can also extract watermark. This preserves data quality. Below figure 2, figure 3, figure 4, figure 5 depicts adding watermark to original relational database. Figure 6, figure 7, figure 8 and figure 9 depict watermark extraction. Figure 10 shows output of watermark extraction. Client registration is shown in figure 11. Figure 12 depicts client login. Client send data request to server is shown in figure 13 and figure 14.

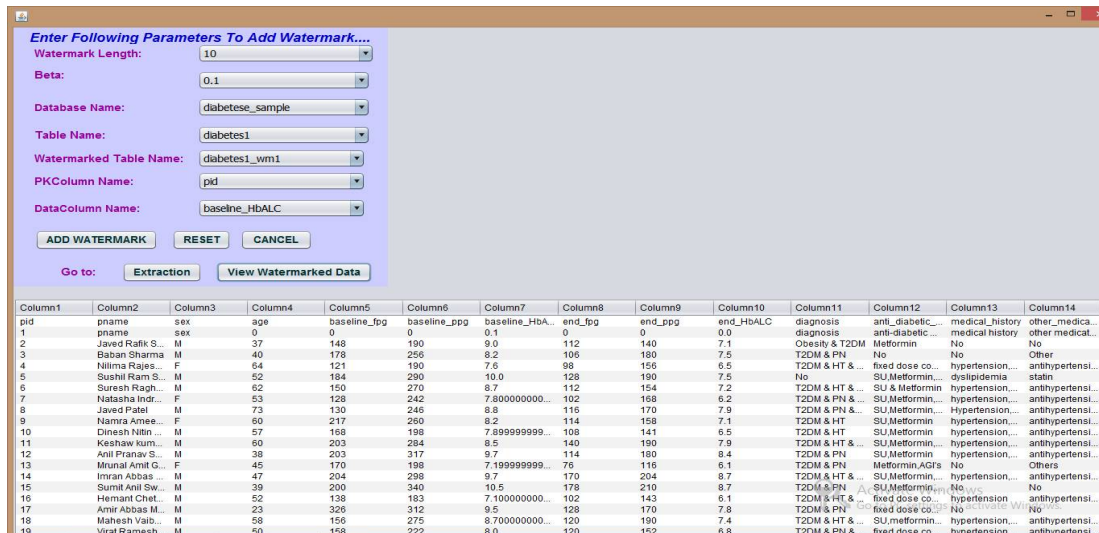


Figure 2: Screenshot of Adding Watermark to Relational Database

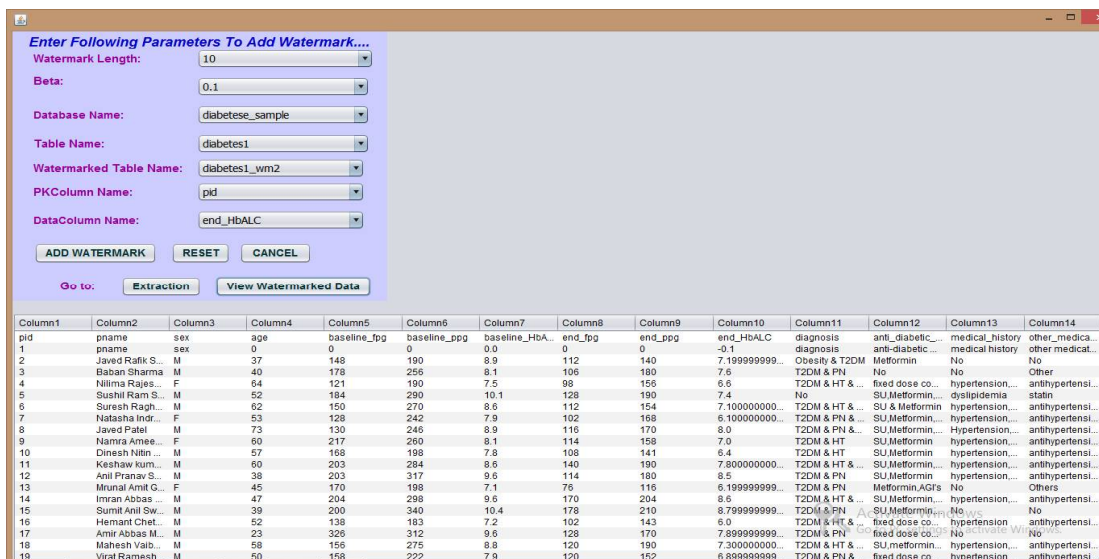


Figure 3: Screenshot of Adding Watermark to Relational Database

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

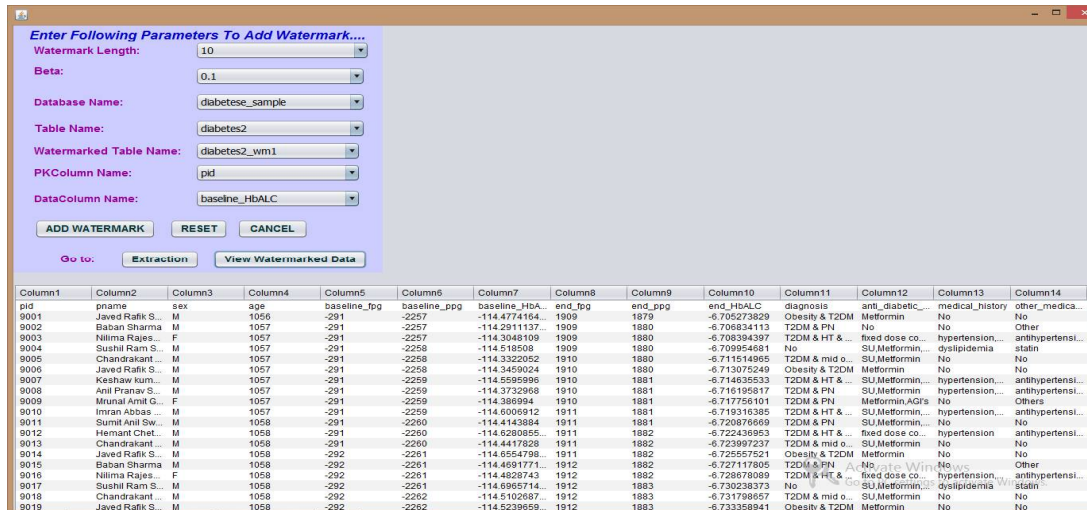


Figure 4: Screenshot of Adding Watermark to Relational Database

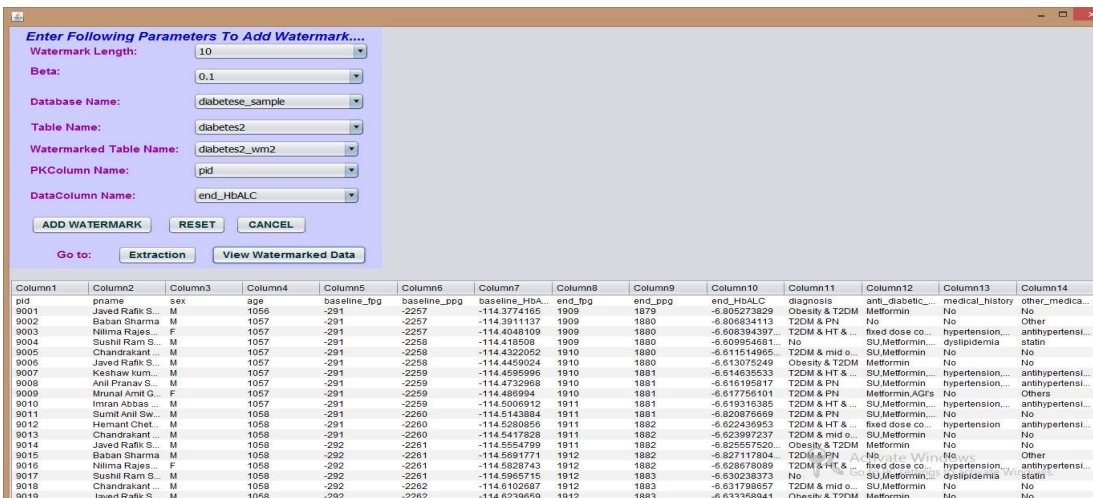


Figure 5: Screenshot of Adding Watermark to Relational Database



Figure 6: Screenshot of Watermark Extraction



Figure 7: Screenshot of Watermark Extraction

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016



Figure 8: Screenshot of Watermark Extraction



Figure 9: Screenshot of Watermark Extraction



Figure 10: Screenshot of Output of Watermark Extraction



Figure 11: Screenshot of Client Registration

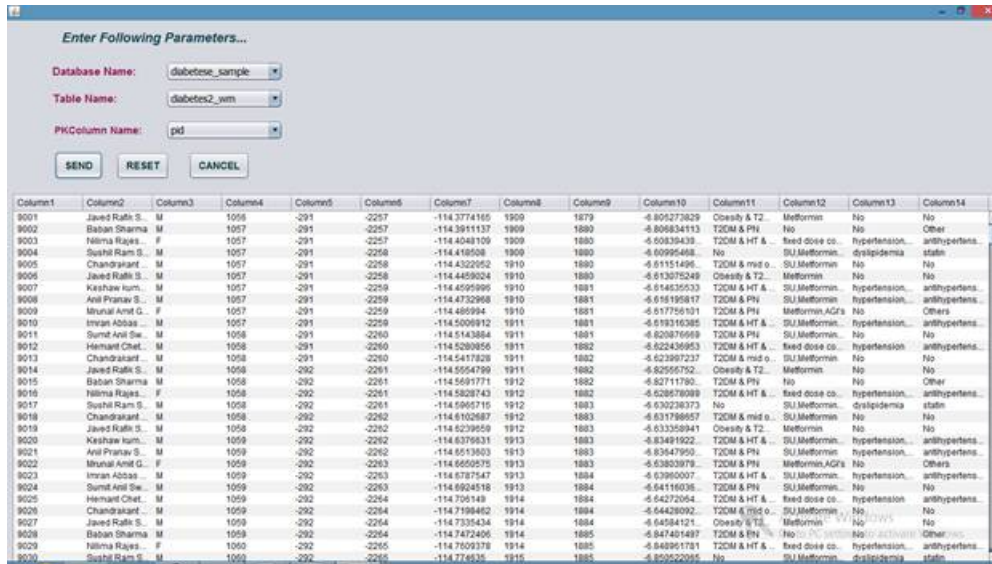


Figure 12: Screenshot of Client Login

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016



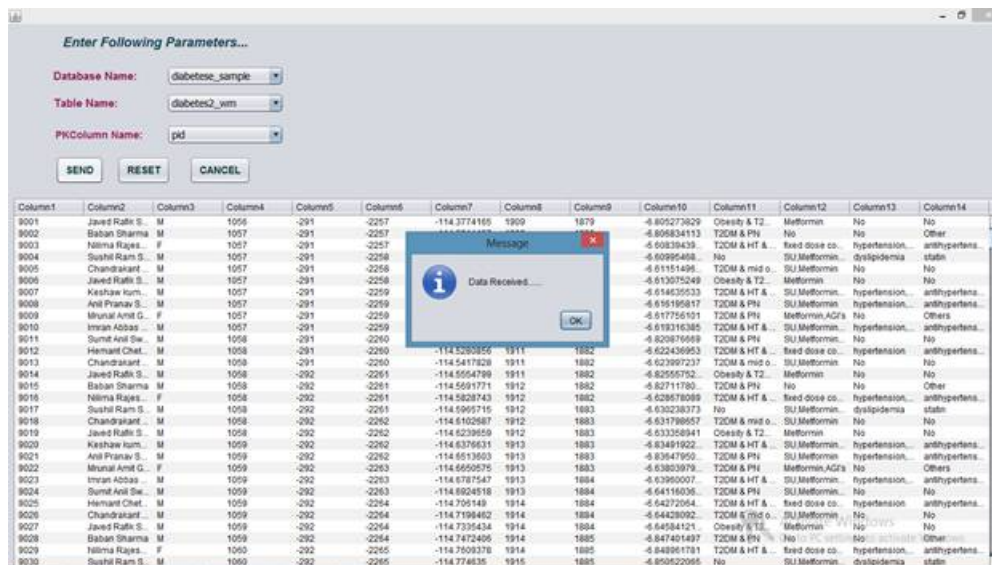
Enter Following Parameters...

Database Name: diabetes_sample
Table Name: diabetes2_ym
PKColumn Name: pid

SEND RESET CANCEL

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
9001	Javed Rafiq S.	M	1956	-291	-2257	-114.3774165	1909	1879	-8.805273829	Obesity & T2.	Metformin	No	No
9002	Babun Sharma	M	1957	-291	-2257	-114.3911137	1909	1880	-8.806834113	T2DM & PN	No	No	Other
9003	Nilima Rajas.	F	1957	-291	-2257	-114.4048109	1909	1880	-8.80839439	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9004	Sushil Ram S.	M	1957	-291	-2258	-114.418508	1908	1880	-8.80995468	No	SU.Metformin	dyslipidemia	statin
9005	Chandrasant	M	1957	-291	-2258	-114.4322052	1910	1880	-8.81151496	T2DM & mid o.	SU.Metformin	No	No
9006	Javed Rafiq S.	M	1957	-291	-2258	-114.4459024	1910	1880	-8.813075249	Obesity & T2.	Metformin	No	No
9007	Kashaw Kum.	M	1957	-291	-2259	-114.4595999	1910	1881	-8.814635533	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9008	Anil Pranav S.	M	1957	-291	-2259	-114.4732968	1910	1881	-8.816195817	T2DM & PN	SU.Metformin	hypertension	antihypertens.
9009	Mrunal Amr G.	F	1957	-291	-2259	-114.486994	1910	1881	-8.817756101	T2DM & PN	Metformin.AGRs	No	Others
9010	Imran Abbas	M	1958	-291	-2259	-114.5006912	1911	1881	-8.819316385	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9011	Sumit Anil Sw.	M	1958	-291	-2260	-114.5143884	1911	1881	-8.820876669	T2DM & PN	SU.Metformin	No	No
9012	Hemant Chat.	M	1958	-291	-2260	-114.5280856	1911	1882	-8.822436953	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9013	Chandrasant	M	1958	-291	-2260	-114.5417828	1911	1882	-8.823997237	T2DM & mid o.	SU.Metformin	No	No
9014	Javed Rafiq S.	M	1958	-292	-2261	-114.5554799	1911	1882	-8.825557521	Obesity & T2.	Metformin	No	No
9015	Babun Sharma	M	1958	-292	-2261	-114.5691771	1912	1882	-8.82711780	T2DM & PN	No	No	Other
9016	Nilima Rajas.	F	1958	-292	-2261	-114.5828743	1912	1882	-8.828678089	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9017	Sushil Ram S.	M	1958	-292	-2261	-114.5965715	1912	1883	-8.830238373	No	SU.Metformin	dyslipidemia	statin
9018	Chandrasant	M	1958	-292	-2262	-114.6102687	1912	1883	-8.831798657	T2DM & mid o.	SU.Metformin	No	No
9019	Javed Rafiq S.	M	1958	-292	-2262	-114.6239659	1912	1883	-8.833358941	Obesity & T2.	Metformin	No	No
9020	Kashaw Kum.	M	1959	-292	-2262	-114.6376631	1913	1883	-8.83491922	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9021	Anil Pranav S.	M	1959	-292	-2262	-114.6513603	1913	1883	-8.83647950	T2DM & PN	SU.Metformin	hypertension	antihypertens.
9022	Mrunal Amr G.	F	1959	-292	-2263	-114.6650575	1913	1883	-8.83803978	T2DM & PN	Metformin.AGRs	No	Others
9023	Imran Abbas	M	1959	-292	-2263	-114.6787547	1913	1884	-8.83959907	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9024	Sumit Anil Sw.	M	1959	-292	-2263	-114.6924519	1913	1884	-8.84115936	T2DM & PN	SU.Metformin	No	No
9025	Hemant Chat.	M	1959	-292	-2264	-114.7061491	1914	1884	-8.84271964	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9026	Chandrasant	M	1959	-292	-2264	-114.7198462	1914	1884	-8.84427992	T2DM & mid o.	SU.Metformin	No	No
9027	Javed Rafiq S.	M	1959	-292	-2264	-114.7335434	1914	1884	-8.84584021	Obesity & T2.	Metformin	No	No
9028	Babun Sharma	M	1959	-292	-2264	-114.7472406	1914	1885	-8.847400497	T2DM & PN	No	No	Other
9029	Nilima Rajas.	F	1960	-292	-2265	-114.7609378	1914	1885	-8.848960781	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9030	Sushil Ram S.	M	1960	-292	-2265	-114.774635	1915	1885	-8.850521065	No	SU.Metformin	dyslipidemia	statin

Figure 13: Screenshot of Client Data Request



Enter Following Parameters...

Database Name: diabetes_sample
Table Name: diabetes2_ym
PKColumn Name: pid

SEND RESET CANCEL

Message
Data Received...
OK

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
9001	Javed Rafiq S.	M	1956	-291	-2257	-114.3774165	1909	1879	-8.805273829	Obesity & T2.	Metformin	No	No
9002	Babun Sharma	M	1957	-291	-2257	-114.3911137	1909	1880	-8.806834113	T2DM & PN	No	No	Other
9003	Nilima Rajas.	F	1957	-291	-2257	-114.4048109	1909	1880	-8.80839439	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9004	Sushil Ram S.	M	1957	-291	-2258	-114.418508	1908	1880	-8.80995468	No	SU.Metformin	dyslipidemia	statin
9005	Chandrasant	M	1957	-291	-2258	-114.4322052	1910	1880	-8.81151496	T2DM & mid o.	SU.Metformin	No	No
9006	Javed Rafiq S.	M	1957	-291	-2258	-114.4459024	1910	1880	-8.813075249	Obesity & T2.	Metformin	No	No
9007	Kashaw Kum.	M	1957	-291	-2259	-114.4595999	1910	1881	-8.814635533	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9008	Anil Pranav S.	M	1957	-291	-2259	-114.4732968	1910	1881	-8.816195817	T2DM & PN	SU.Metformin	hypertension	antihypertens.
9009	Mrunal Amr G.	F	1957	-291	-2259	-114.486994	1910	1881	-8.817756101	T2DM & PN	Metformin.AGRs	No	Others
9010	Imran Abbas	M	1958	-291	-2259	-114.5006912	1911	1881	-8.819316385	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9011	Sumit Anil Sw.	M	1958	-291	-2260	-114.5143884	1911	1881	-8.820876669	T2DM & PN	SU.Metformin	No	No
9012	Hemant Chat.	M	1958	-291	-2260	-114.5280856	1911	1882	-8.822436953	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9013	Chandrasant	M	1958	-291	-2260	-114.5417828	1911	1882	-8.823997237	T2DM & mid o.	SU.Metformin	No	No
9014	Javed Rafiq S.	M	1958	-292	-2261	-114.5554799	1911	1882	-8.825557521	Obesity & T2.	Metformin	No	No
9015	Babun Sharma	M	1958	-292	-2261	-114.5691771	1912	1882	-8.82711780	T2DM & PN	No	No	Other
9016	Nilima Rajas.	F	1958	-292	-2261	-114.5828743	1912	1882	-8.828678089	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9017	Sushil Ram S.	M	1958	-292	-2261	-114.5965715	1912	1883	-8.830238373	No	SU.Metformin	dyslipidemia	statin
9018	Chandrasant	M	1958	-292	-2262	-114.6102687	1912	1883	-8.831798657	T2DM & mid o.	SU.Metformin	No	No
9019	Javed Rafiq S.	M	1958	-292	-2262	-114.6239659	1912	1883	-8.833358941	Obesity & T2.	Metformin	No	No
9020	Kashaw Kum.	M	1959	-292	-2262	-114.6376631	1913	1883	-8.83491922	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9021	Anil Pranav S.	M	1959	-292	-2262	-114.6513603	1913	1883	-8.83647950	T2DM & PN	SU.Metformin	hypertension	antihypertens.
9022	Mrunal Amr G.	F	1959	-292	-2263	-114.6650575	1913	1883	-8.83803978	T2DM & PN	Metformin.AGRs	No	Others
9023	Imran Abbas	M	1959	-292	-2263	-114.6787547	1913	1884	-8.83959907	T2DM & HT &	SU.Metformin	hypertension	antihypertens.
9024	Sumit Anil Sw.	M	1959	-292	-2263	-114.6924519	1913	1884	-8.84115936	T2DM & PN	SU.Metformin	No	No
9025	Hemant Chat.	M	1959	-292	-2264	-114.7061491	1914	1884	-8.84271964	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9026	Chandrasant	M	1959	-292	-2264	-114.7198462	1914	1884	-8.84427992	T2DM & mid o.	SU.Metformin	No	No
9027	Javed Rafiq S.	M	1959	-292	-2264	-114.7335434	1914	1884	-8.84584021	Obesity & T2.	Metformin	No	No
9028	Babun Sharma	M	1959	-292	-2264	-114.7472406	1914	1885	-8.847400497	T2DM & PN	No	No	Other
9029	Nilima Rajas.	F	1960	-292	-2265	-114.7609378	1914	1885	-8.848960781	T2DM & HT &	fixed dose co.	hypertension	antihypertens.
9030	Sushil Ram S.	M	1960	-292	-2265	-114.774635	1915	1885	-8.850521065	No	SU.Metformin	dyslipidemia	statin

Figure 14: Screenshot of Client Data Request

VI. CONCLUSION AND FUTURE WORK

The system provides security to the shared databases in distributed environment by recoverable watermarking numerical data where different users share their data in various proportions. It performs recoverable watermarking. First, it embeds watermark in numeric data of relational database. During extraction process system detects watermark hence it is recoverable watermarking. System also preserves and maintains data quality by recoverable watermarking. Also system maintains ownership protection by performing recoverable watermarking on numerical data in relational database for distributed environment. Future concern is to perform recoverable watermarking for distributed databases for non-numeric data.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

REFERENCES

1. Snehal S. Kshatriya and Dr. S. S. Sane, "A Study of Watermarking Relational Databases," in International Journal of Application or Innovation in Eengineering & Management , vol. 3,issue 10, October 2014 .
2. R. Agrawal and J. Kiernan, "Watermarking relational databases," in Proceedings of the 28th international conference on Very Large Data Bases. VLDB Endowment, 2002, pp. 155–166.
3. S. Subramanya and B. K. Yi, "Digital rights management," Potentials, IEEE, vol. 25, no. 2, pp. 31–34, 2006.
4. Y. Zhang, B. Yang, and X.-M. Niu, "Reversible watermarking for relational database authentication," Journal of Computers, vol. 17, no. 2, pp. 59–66, 2006.
5. G. Gupta and J. Pieprzyk, "Database relation watermarking resilient against secondary watermarking attacks," in Information Systems Security. Springer, 2009, pp. 222–236.
6. J.-N. Chang and H.-C. Wu, "Reversible fragile database watermarking technology using difference expansion based on svr prediction," in Computer, Consumer and Control (IS3C), 2012 International Symposium on. IEEE, 2012, pp. 690–693.
7. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," Journal of Systems and Software, 2013.
8. Saman Iftikhar, M. Kamran and Zahid Anwar, "A Robust and Reversible Watermarking Technique for Relational Data," in Knowledge and Data Engineering, val X, No : XX., IEEE, 2015.