Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Cyber Security in Data Mining Using Homomorphic Encryption

Ekta Chauhan* , **Sonia Vatta**
School of Computer Science and Engineering
Bahra University, India

Abstract— In this paper we discuss the issue of privacy preserving data mining and present the technique that provide the privacy on data mining application. We provide an overview of the different techniques and how they relate to one another. We used the asymmetric encryption to provide the privacy and used the RSA encryption to encrypt the data. We also present a client server architecture that connects to the multiple clients. Server need to receive data from client .Connect server to a data base and enter the data received from the client into database with client id. Our proposed protocol is to encrypt the data so we used the encryption technique to encrypt the data, we also used homomorphic encryption to secure the information. Without security our data may stand compromised.

Keywords—- Data Mining, Homomorphic encryption, Client Server Architecture.

I. INTRODUCTION

Computer security is information security as applied to computers and networks. The field covers all the processes and mechanisms by which computer-based equipment, information and services are protected from unintended or unauthorized access, change or destruction. Computer security also includes protection from unplanned events and natural disasters. The problem of privacy-preserving data mining has become more important in recent years because of the increasing ability to store personal data about users, and the increasing sophistication of data mining algorithms to leverage this information. Generally, data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information. Explosive progress in networking, storage, and processor technology has led to the creation of ultra large databases that record unprecedented amount of transactional information. The main problem is that with the availability of non-sensitive information or unclassified data, one is able to infer sensitive information that is not supposed to be disclosed. Despite its benefits in various areas such as marketing, business, medical analysis, bioinformatics and others, data mining can also pose a threat to privacy in database security if not done or used properly. Privacy preserving data mining, is a novel research direction in data mining and statistical databases, where data mining algorithms are analyzed for the side-effects they incur in data privacy. [7,8]. Homomorphic encryption scheme originally called a privacy homomorphic was introduced by Rivest, Adleman and Dertouzous shortly after the invention of RSA by Rivet, Shamir, and Adelman. Basic RSA is a multiplicatively homomorphic encryption scheme.

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on cipher text and obtain an encrypted result which decrypted matches the result of operations performed on the plaintext. For instance, one person could add two encrypted numbers and then another person could decrypt the result, without either of them being able to find the value of the individual numbers. Homomorphic encryption schemes are malleable by design. The homomorphic property of various cryptosystems can be used to create secure voting systems, collision-resistant hash functions, private information retrieval schemes and enable widespread use of cloud computing by ensuring the confidentiality of processed data. There are several efficient, partially homomorphic cryptosystems, and two fully homomorphic, but less efficient cryptosystems. Although a cryptosystem which is unintentionally homomorphic can be subject to attacks on this basis, if treated carefully homomorphism can also

RSA is a type of encryption known as public key encryption. That is anyone can access the specific public key and herefore encrypt. But the key is that the key is asymmetric: it will encrypt data but it will not decrypt data. Only someone with the private key can perform the decryption.

The underlying way in which RSA encryption works is to exploit what is effectively a one way function. That is, if you find the product of two large prime numbers, then the only way to get back to 'p' and 'q', the two base primes, is through a slow and arduous process of factoring. Now, to find 'p' and 'q' for the number '21' is pretty easy: we know through our maths that p and q are here 3 and 7. However, when the end number is the product of two 1024 bit prime numbers (very large numbers) then it is slightly time consuming to find the correct numbers .We also provide the client server architecture.

II. RESEARCH METHODOLOGY

Methodology of constructing the proposed system will consists of various modules .Each module uses different techniques and algorithms to perform its specific tasks. When a particular module completes its task, its output will become input for the next module. In the end the combined effort of each module will be displayed.

MODULE 1: Design database

- We create the data and then import that data into the wamp server.
- Connect dataset to mysql using the jdbc connector.
- Create the applications in the java net beans.

MODULE 2: Learning

- Apply the RSA encryption on the dataset.
- Then we apply the data mining .data mining abstract the information from the data base.
- Apply the decision tree and association rules in weka tool.
- Then the reporting module where we will create test reports and check performance with analysis tools that we are used.

MODULE 3: Testing

- Now we install the VM workstation and start the new window on the same machine.
- Send the data from the client to the server and encrypt that data using the RSA encryption.
- We can decrypt the data only on the server site or client side but can refresh the data only on server sit.

III. **IMPLEMENTATION**

All code for this project was implemented in Java and will describe security in data mining and create a server that can connect to the multiple clients Main class is RSAHomomorphic. This RSA encrypt the data so that no other person can read this data. In fig (a) we apply the decision tree on the weka tool and obtained the result .And in the next we will encrypt the field value. In fig (b) we create the database application and in fig (c) we select the field and encrypt the value. In fig (d) we again press the decrypt button and obtained the original value.

Results

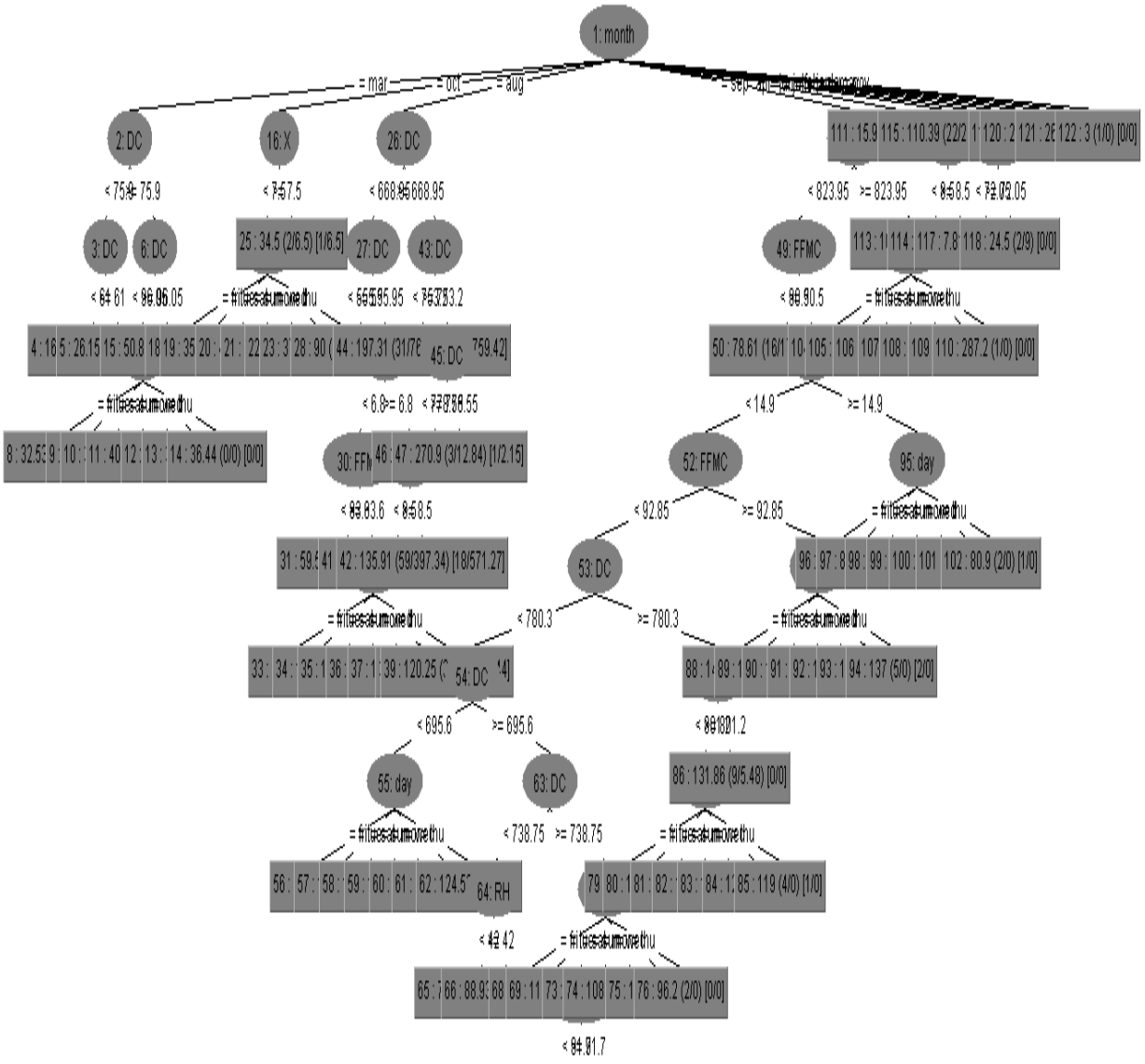


Fig (a)

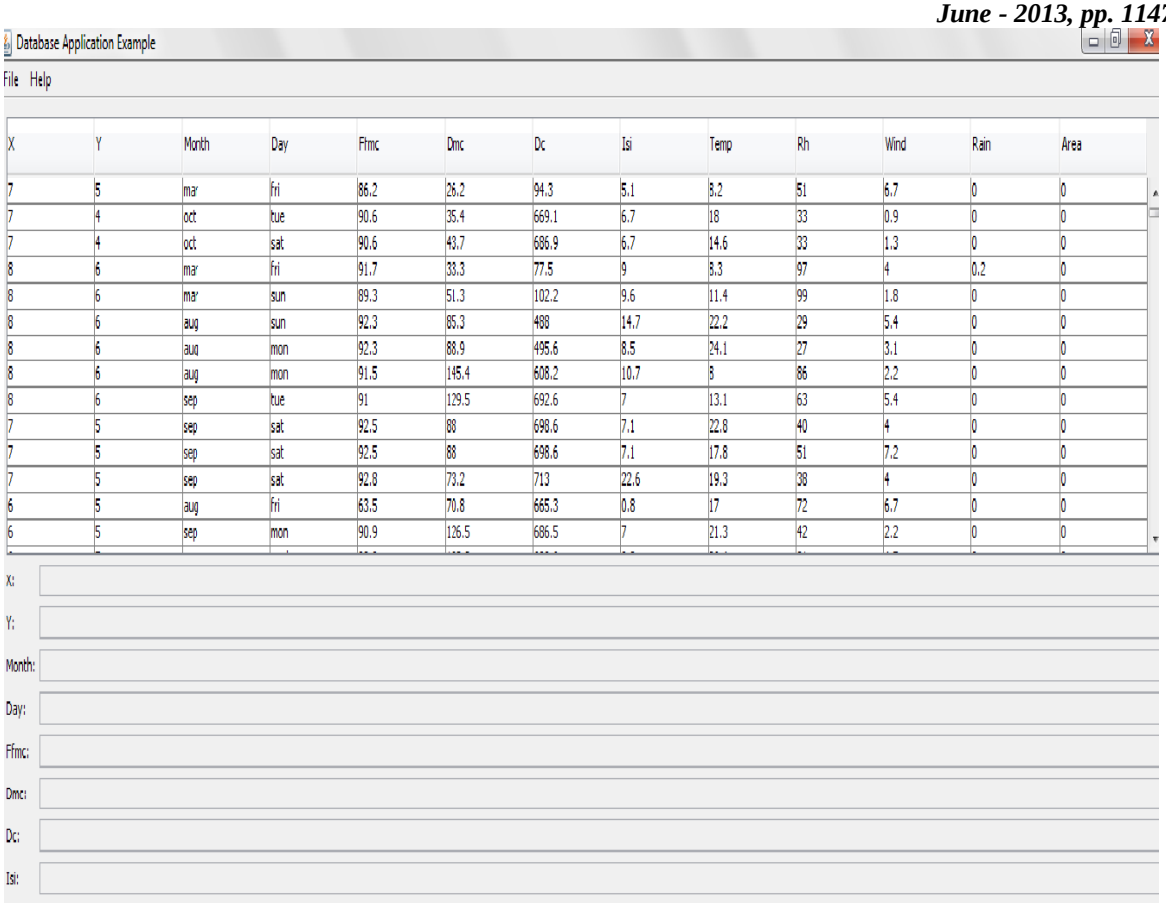


Fig (b) Database Application

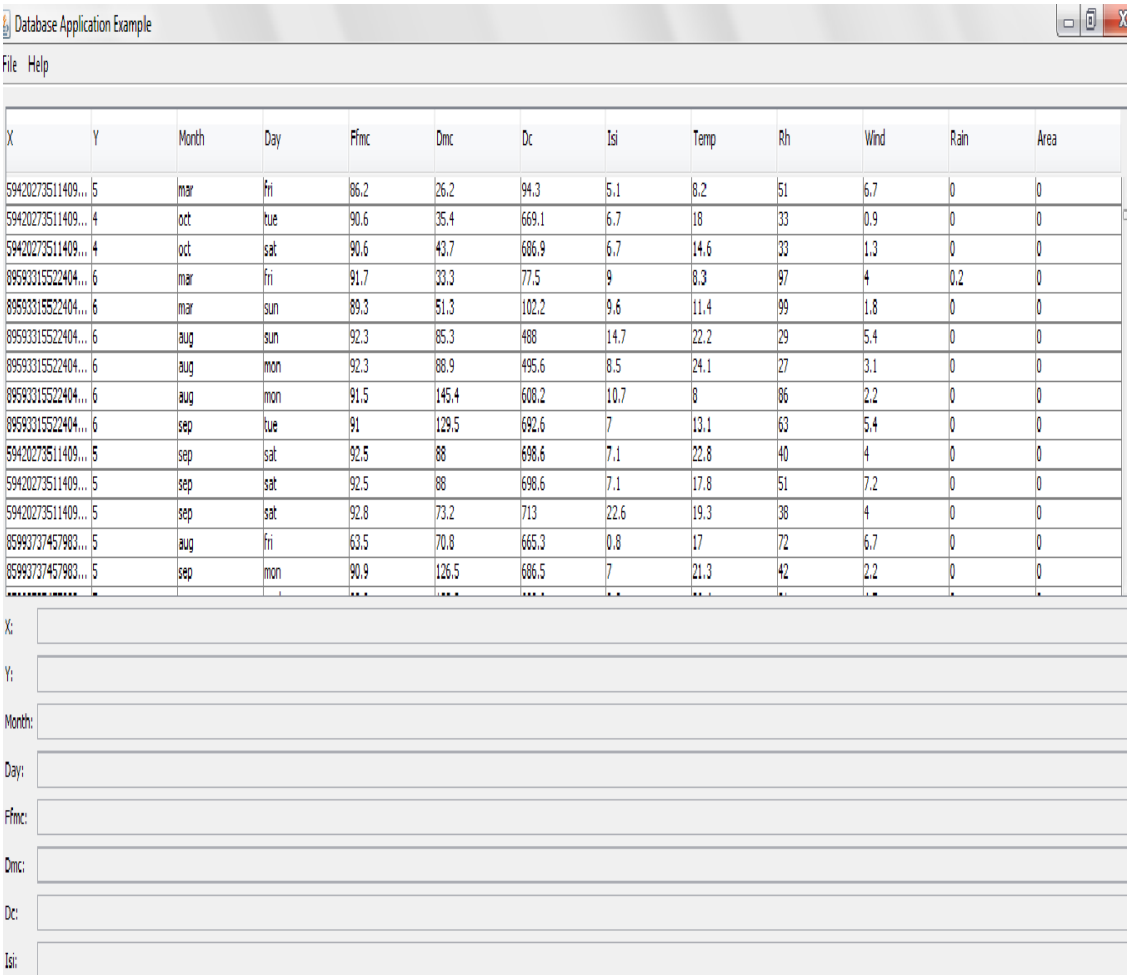
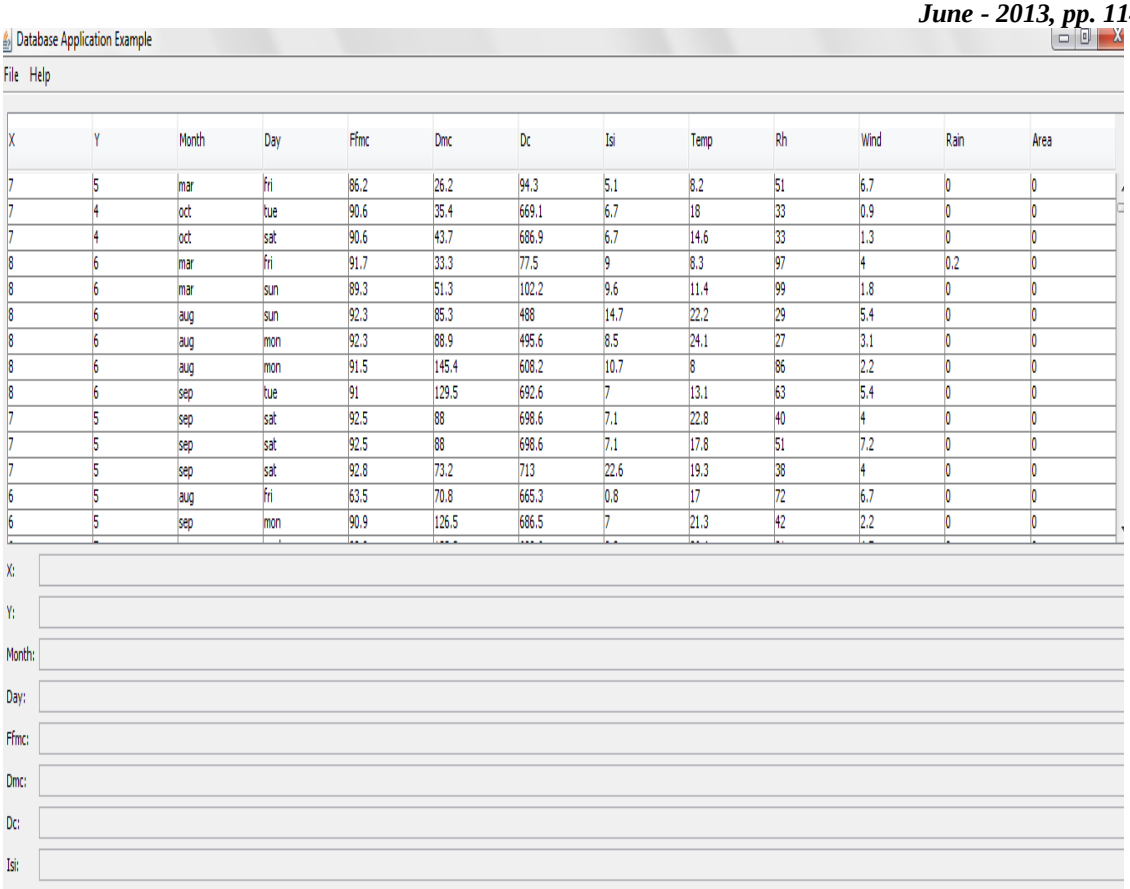


Fig (c) Encrypted form of database



X	Y	Month	Day	Ffmc	Dmc	Dc	Isi	Temp	Rh	Wind	Rain	Area
7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
7	4	oct	tue	90.6	35.4	669.1	6.7	18	33	0.9	0	0
7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0	0
8	6	mar	fri	91.7	33.3	77.5	9	8.3	97	4	0.2	0
8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0	0
8	6	aug	sun	92.3	85.3	488	14.7	22.2	29	5.4	0	0
8	6	aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0	0
8	6	aug	mon	91.5	145.4	608.2	10.7	8	86	2.2	0	0
8	6	sep	tue	91	129.5	692.6	7	13.1	63	5.4	0	0
7	5	sep	sat	92.5	88	698.6	7.1	22.8	40	4	0	0
7	5	sep	sat	92.5	88	698.6	7.1	17.8	51	7.2	0	0
7	5	sep	sat	92.8	73.2	713	22.6	19.3	38	4	0	0
6	5	aug	fri	63.5	70.8	665.3	0.8	17	72	6.7	0	0
6	5	sep	mon	90.9	126.5	686.5	7	21.3	42	2.2	0	0

Fig (d) Refresh form of database

V. CONCLUSIONS

Privacy preserving data mining is an ongoing research area and there are a lot of issues that needs to be addressed. First of all, the databases that are collected for mining are huge, and scalable techniques for privacy preserving data mining are needed to handle these data sources. In our approach, we have implemented privacy preservation in data mining by using the homomorphic encryption it add security so that any data mining technique does not lose his valuable data. Here, we have assumed that the decryption occurs entirely at the Server. For real time applications with crucial time-constraints like biomedical applications, the keys for decryption can be distributed to the user for faster decryption and retrieval of data. In further work we can also use elliptical cryptography and compare the different cryptography technique. In this current system we can refresh the data only on the server site and it has the fixed length. But in the future we can refresh the data on client site and can also increase the length size so that if we double click on encrypt button the value should be unencrypted not unreadable.

REFERENCES

[1] Rakesh Agrawal, Tomasz Imieliski, and Arun Swami. Mining association rules. between sets of items in large databases. In Proceedings of the 1993 ACM SIG-MOD international conference on Management of data, 1993.

[2] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and J Sander. Lof: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIG-MOD international conference on Management of data, 2000.

[3] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, September 1994.

[4] Wiener, Michael J. (May 1990). "Cryptanalysis of short RSA secret exponents". Information Theory, IEEE Transactions on:

[5] Varun Chandola and Vipin Kumar. Summarization. In Fifth IEEE International Conference on Data Mining, Houston, TX, November 2005.

[6] Eric Eilertson, Levent ErtÄoz, Vipin Kumar, and Kerry Long. Minds a new approach to the information security process. In 24th Army Science Conference.US Army, 2004

[7] Privacy Preserving Data Mining Using Cryptographic Role Based Access Control Approach. Lalanthika Vasudevan , S.E. Deepa Sukanya, N. Aarthi* 2008 Vol I IMECS 2008,.

[8] Anor F.A. Dafa-Alla, Eun Hee Kim, Keun Ho Ryu, *Yong Jun Heo "PRBAC: An Extended Role Based Access Control for Privacy preserving Data mining" In Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science(ICIS'05) of IEEE, 2005 .

