Homomorphic Encryption Schema for Privacy Preserving Mining of Association Rules

M.Sangeetha[1], P. Anishprabu[2], S. Shanmathi[3]
Department of Computer Science and Engineering
SriGuru Institute of Technology
Coimbatore, India.
Email: msangeetha2612@gmail.com[1], anishbtechit@gmail.com [2], shanmathi.shan@gmail.com[3]

***Abstract*-- Cloud computing and its model for IT services based on the internet and big data centers, the outsourcing of data and computing services .A company (data owner) lacking in expertise or computational resources can outsource its mining needs to a third party service provider (server). However, both the items and the association rules of the outsourced database are considered private property of the corporation (data owner). To protect corporate privacy, the data owner transforms its data and ships it to the server, sends mining queries to the server, and recovers the true patterns from the extracted patterns received from the server. The problem of outsourcing the association rule mining task within a business privacy-preserving framework. The comprehensive experiments on a very large and real transaction database demonstrate that our techniques are effective, scalable, and keep privacy.**

Index terms-- K-privacy, Privacy preserving, Outsourcing, Homomorphic Encryption, Association Rule Mining.

I . INTRODUCTION

Extraction of hidden predictive information from the large database is a new emerging technology in data mining. Data mining is the process of analyze data from dissimilar perspective and summarizing it into useful information that can be used to enlarge proceeds, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different extent or angles, classify it, and review the dealings identified. The actual data mining task is the automatic or semi-automatic analysis of large quantity of data to remove previously unknown interesting patterns such as groups of data records, unusual records and dependency.

This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. results by a decision support system. In general, association rule mining is mostly used for market analysis applications.

II . PRIVACY PRESERVING SECURITY

Data mining is special technical term related with the discovery of new and interesting pattern of data from large data sets. The extraction of hidden predictive information from large databases is a new emerging technology having the huge potential for the help of companies to focus on the important information in the data warehouse. The tools of data mining predicts the future trends and behaviors. This future trends and behavior allow the businesses to make proactive analysis and decision making for the growth of different aspects of the companies. This data mining automates the system to search the relevant information from the databases of data warehouse of the given enterprise which maintains the data warehouse. The data mining tools can answer the business questions which are traditionally very complicated task and take too much time to analyze and produce the result. Most of the companies already collect and refine huge quantities of data.

A. *Security*

Data mining is the process of posing a series of proper queries to extract information from large quantities of data in the database. Data mining techniques can be functional to handle problems in database security. On the other hand, data mining techniques can also be employed to cause security problems. Data mining techniques include those based on rough sets, inductive logic programming, machine learning, and neural networks, among others. Essentially one arrives at some hypothesis, which is the information extracted, from examples and patterns observed. These patterns are observed from posing a series of queries; each query may depend on the response obtained to the previous queries posed.

B. *Issues*

The main model here is that private data is collected from a number of sources by a collector for the purpose of consolidating the data and conducting mining. The collector is not trusted with protecting the privacy, so data are subjected to a random perturbation as it is collected. Techniques have been developed for perturbing the data so as to preserve privacy while ensuring the mined patterns or other analytical properties are sufficiently close to the patterns mined from original data. This body of work was pioneered by and has been followed up by several papers since. This approach is not suited for corporate privacy, in that some analytical properties are disclosed.

Another related issue is secure multiparty mining over distributed datasets. Data on which mining is to be performed is partitioned, horizontally or vertically, and distributed among several parties. The partitioned data cannot be shared and must remain private but the results of mining on the union of the data are shared among the participants, by means of multiparty secure protocols. It do not consider third parties. This approach partially implements corporate privacy, as local databases are kept private, but it is too weak for our outsourcing problem, as the resulting patterns are disclosed to multiple parties.

C. *Privacy Preserving In Data mining*

Some people consider that data mining itself is ethically neutral. While the term "data mining" has no ethical implications, it is often associated with the mining of information in relation to peoples' behavior. To be precise, data mining is a arithmetical method that is applied to a set of information (i.e., a data set). associate these data sets with people is an extreme narrowing of the types of data that are available. Examples could range from a set of crash test data for customer vehicles, to the performance of a group of stocks. These types of data sets make up a great proportion of the information available to be acted on by data mining methods, and rarely have ethical concerns associated with them. However, the ways in which data mining can be used can in some cases and context raise questions regarding privacy, legality, and ethics. In particular, data mining government or commercial data sets for national security or law enforcement purposes, such as in the Total Information Awareness Program or in ADVISE, has raised privacy concerns.

III . **HOMOMORPHIC ENCRYPTION**

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on ciphertext 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.Using such a scheme, any circuit can be homomorphically evaluated, effectively allowing the construction of programs which may be run on encryptions of their inputs to produce an encryption of their output. Since such a program never decrypts its input, it can be run by an untrusted party without revealing its inputs and internal state. The existence of an efficient and fully homomorphic cryptosystem would have great practical implications in the outsourcing of private computations, for instance, in the context of cloud computing.

The "homomorphic" part of a fully homomorphic encryption scheme can also be described in terms of category theory. If C is the category whose objects are integers (i.e., finite streams of data) and whose morphisms are computable functions, then (ideally) a fully homomorphic encryption scheme elevates an encryption function to a function from C to itself.

The utility of fully homomorphic encryption has been long recognized. The problem of constructing such a scheme was first proposed within a year of the development of RSA. A solution proved more elusive; for more than 30 years, it was unclear whether fully homomorphic encryption was even possible. During this period, the best result was the cryptosystem which supports evaluation of an unlimited number of addition operations but at most one multiplication.

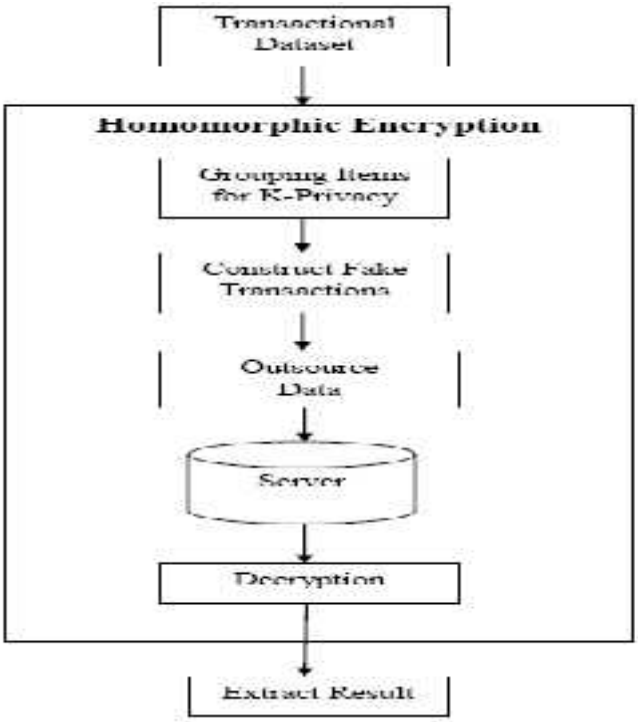
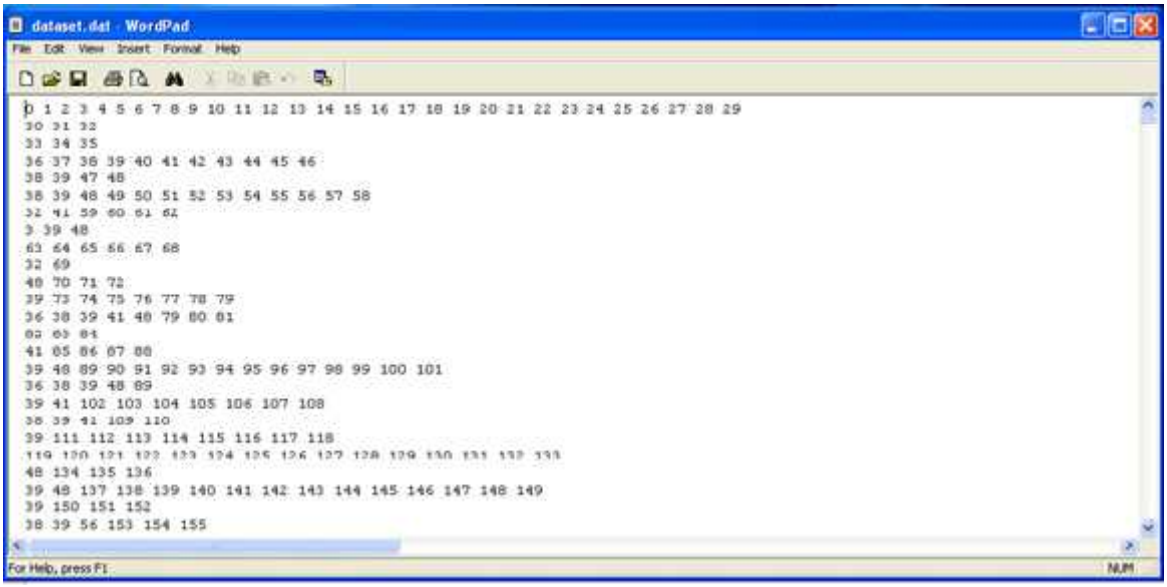


Fig 1.1 Architecture Diagram

IV . **RESULT**



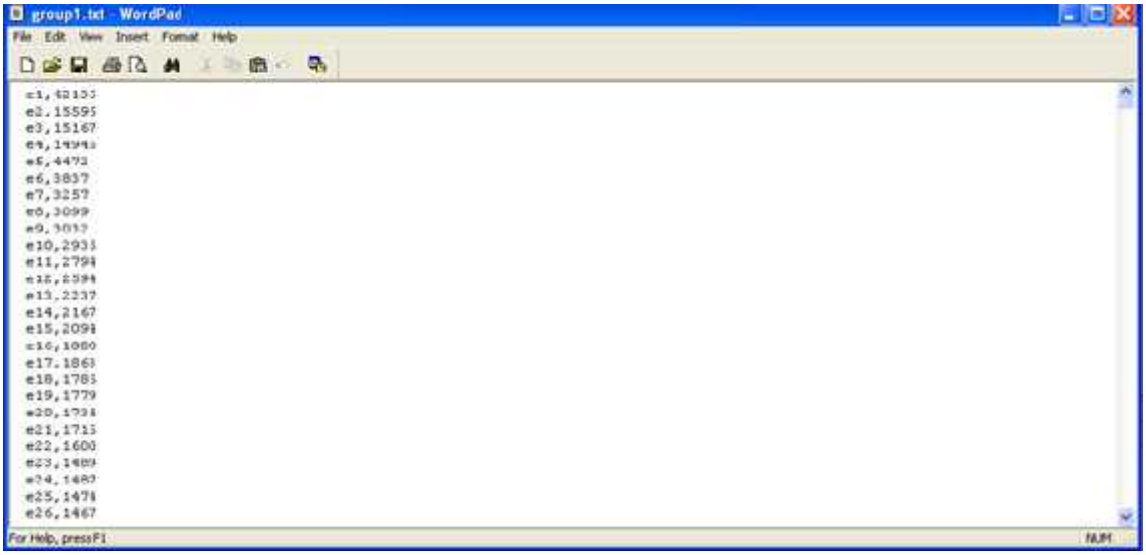
A. Dataset Collection and Encryption

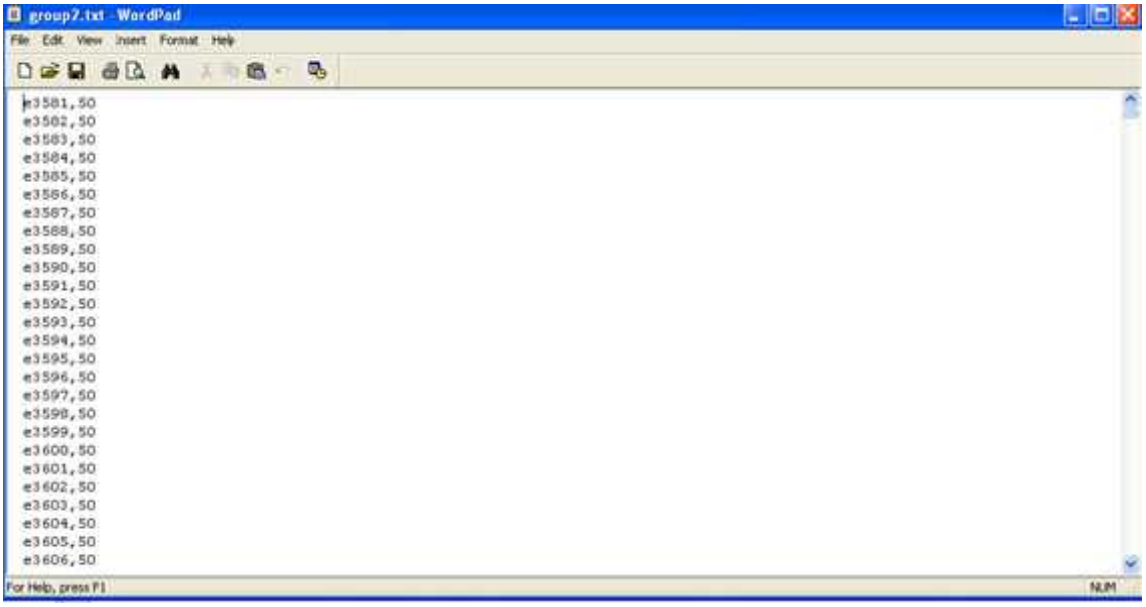
A dataset (or data set) is a collection of data. Dataset is collected from Belgium trade market dataset. It contains the retail market basket data from an unidentified Belgian retail store. The data are provided 'as is'. Basically, any use of the data is allowed as long as the proper acceptance is provided and a copy of the work is provided.

B. Grouping Items For K-Privacy

Given the items support table, several strategy can be adopted to cluster the items into groups of size k. To start from a simple grouping method. To assume the item support table is sorted in descending order of support and refer to cipher items in this order as e1,e2, etc.

Grouping = Maximum support value – Minimum support value / No.of. groups





C. *Constructing Fake Transactions*

To add fake transactions for each and every transaction with the original transactional data. It will be in the form of noise.

D. *Decryption*

When the client requests the effecting of a pattern pulling out query to the server, then the server will provide the exact true patterns.

V .**CONCLUSION**

Security issues of the data-mining as- a-service paradigm. One of the main security issues is that the server has access to valuable data of the owner and may learn sensitive information from it. For example, by looking at the transactions, the server (or an intruder who gains access to the server) can learn which items are always copurchased. However, both the transactions and the mined patterns are the property of the data owner and should remain safe from the server.

Homomorphic Encryption scheme based association rule mining when new data is added to the transactional database, encryption schema was performed by whole data. Improve the security of the system by modifying the existing encryption schema. And, to reduce the time complexity by using this scheme then develop the encryption scheme which enables more security of the cloud outsourcing data in the transaction database. Analysis the encryption scheme to achieve the provable privacy guarantee of outsourced transaction database.

ACKNOWLEDGEMENT

The authors would like to thank the staff and students of SriGuru Institute of technology, friends and family members for their support and guidance in bringing this research article. The authors would also like to thank them for their valuable support.

VI . REFERENCES

[1] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, "Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases," in IEEE Systems Journal,vol 7, No.3, September 2013.

[2] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "Security In Outsourcing Of Association Rule Mining," International Conference On Very Large Data Bases, 2007, pp. 111–122.

[3] L. Qiu, Y. Li, and X. Wu, "Protecting Business Intelligence And Customer Privacy While Outsourcing Data Mining Tasks," Knowledge Information System, vol. 17, no. 1, pp. 99–120, 2008.

[4] C. Clifton, M. Kantarcioglu, and J. Vaidya, "Defining Privacy For Data Mining," National Science Foundation. Workshop Next Generation Data Mining, 2002, pp. 126–133.

[5] I. Molloy, N. Li, and T. Li, "On The (In)Security And (Im)Practicality Of Data Mining, December 2009, pp. 872–877.

[6] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-Preserving Data Mining From Outsourced Databases," SPCC2010 Conjunction with CPDP, 2010, pp. 411–426.

[7] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining," ACM SIGMOD International Conference Management Data, 2000, pp. 439–450.

[8] S. J. Rizvi and J. R. Haritsa, "Maintaining Data Privacy In Association Rule Mining," International Conference Very Large Data Bases, 2002, pp. 682– 693.

[9] M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining Of Association Rules On Horizontally Partitioned Data," IEEE Transaction on Knowledge Data Engineering, vol. 16, no. 9, pp. 1026–1037, September 2004.

[10] B. Gilburd, A. Schuster, and R. Wolff, "k- ttp: A New Privacy Model For Large Scale Distributed Environments," International Conference Very Large Data Bases, 2005, pp. 563–568.

M Sangeetha was born in Theni on 26th December 1990. She received her B.Tech.(IT) degree from Periyar Maniammai University, Thanjavur, Tamil Nadu in 2012. She is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu. She has Published research papers and articles in various international journals. She has presented a papers in various national and international conferences. She is interested in Secure Computing, Data mining and Audio mining.

P.Anishprabu was born in Namakkal on 14th May 1990. He received his B.Tech.(IT) degree from KSR College of Engineering, Tiruchengode, Tamil Nadu in 2011. He is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu.He has presented a papers in various conferences. He is interested in Data mining and Operating system.

S.Shanmathi was born in Tiruchengode on 8th Augest 1991. She received her B.E (CSE) degree from Sengundhar College of Engineering Tiruchengode, Tamil Nadu in 2012. She is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu. She has presented papers in various national conferences. Her areas of interest are Network Security, Datastructures and web designing.