PROTECTED REMOVAL OF CONNECTION RULES IN STRAIGHT DISSEMINATED DATABASE

[1]Raunak Rathi, [2]A.V. Deorankar, [3]Roshni Sherkar
[1]M.TECH STUDENT GCOEA,[2]HOD I.T. GCOEA, [3]M.TECH STUDENT GCOEA
Email:[1]Raunakrathi.rathi@gmail.com , [2]avdeorankar@gmail.com, [3]rinku.sherkar@gmail.com

Abstract— In today's world the knowledge extraction plays a crucial role. The knowledge extraction know totally depends on scattered or distributed database. In this paper we are studying about the security of distributed database and implementing the hash key concept to improvise the computational speed of the algorithm. The automatic hash key concept will increase the efficiency and improvise the enhancement in the field of secure data mining.

Keywords — mining, association, secure, classification.

I. INTRODUCTION

The knowledge extraction is very important measure in the area of database. In distributed database system the system should also take care of the privacy of an data. To overcome this measure the paper[1]. Given a protocol which introduce third party system. In this case the protocol comes to solution that the parties should chose a trusted third party an by introducing such third party, we can achieve the goal to maintain the privacy of data. One more problem with the data in distributed database is the data in network. To overcome this measure the protocol use encryption and decryption method. The data is first encrypted and when the data travel in network and then the data is decrypted when it comes to the client.

Data mining and KDD(Knowledge discovery in database) are two different kind of research area which examine the auto extraction of earlier unidentified pattern from huge amount of data. To find the solution of secure mining has become more essential in upcoming years due to the rising capability to save personal data about users and the rising complexity of data mining algorithm to influence this information. A number of technique as such classification, kanonymity, association rule mining, clustering had been recommended in upcoming years in order to performed secure data mining. Besides, the difficulty has been discussed in several community such as the database community, the statistical disclosure control community and the cryptography community. Data mining technique has been evolved successfully to extract knowledge in such to maintain a variety of domains weather, national security, forecasting, medical diagnosis, and marketing. Although it is confront to mine such kind of data without violating the data owner's privacy. For example, how to mine an employee private data is an ongoing problem in multinational company's application. As such data mining become more enveloping, secure concern are rising.

II. HASH KEY

A hash function is the function so as to be used to plot digital data of random size to digital data of permanent size, with small difference in input data producing very large difference in output data. The ideals returned by a hash function are called hash

ideals, simply hash, or hashes codes. One sensible make use of is a data structure called a hash table, extensively used in computer software for fast data hunt for. Hash functions speed up table or database search for by detecting duplicated report in a big file. An pattern is finding similar stretch in DNA sequence. They are also helpful in cryptography. A cryptographic hash function allow one to with no trouble verify that some input data match a stored hash value, but makes it hard to rebuild the data from the hash alone. This standard is used by the PGP algorithm for data justification and by a lot of password examination system.

Hash functions are linked to (and often confused with) ciphers, error-correcting codes , randomization functions, fingerprints, check digits and checksums. Although these concepts partly cover to some extent, each has its own uses and necessities and is considered and optimized differently. The Hash Keeper database maintain by the American National Drug Intelligence Center, for instance, is more aptly described as a catalog of file fingerprints than of hash values.

This concept will increase the computation cost of the protocol and will enhance the mechanism of the protocol.

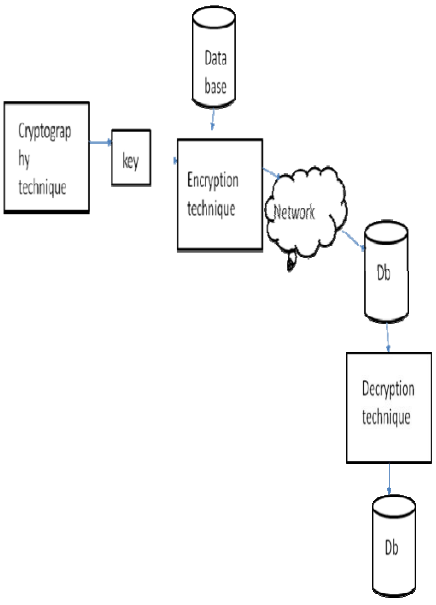
III. EXISTING WORK

Data mining is a beneficial technique used to extract data/knowledge from large collection of data, but the collection of data is in distributed form many times. In such case privacy plays an important role to maintain the privacy of data or some part of knowledge about the data. The problem here we will discuss from paper[1]. In horizontal distributed data plays an important factor to look for is the distributed database. Here several players that access homogenous databases, i.e., the database that share the same schema but hold different information. The paper[1] support at least S and confidence C, for some given minimum support size of S and confidence C, that hold in united database, while reducing the information release about the secure (or private) database accessed by such players.

The paper deals with the problem of secure multi-party calculation. If a trusted third party would be present, then the players could devote to such party and such party would evaluate and send them such resulting output. If such third party would not be present, it is need to develop a protocol that player can use on own in order to get their required output Y. If no player learn from such view, these protocol is consider perfectly secure more than that the third party would learn the ideal settings where the calculation is carried out by the trusted third party. The protocol that we used here calculates a parameterized family of functions, which we can say as a threshold function, in which the two excessive cases match up to the problem of calculating the union and insertion of private subsets. Those can be said as general purpose protocol that can be used in other part as well. One more problem regarding secure multi-party calculation is the set of addition problem; namely, the problem in which Bobs holds a private subsets of several ground set, and Alice hold an element in the ground set, and they desire to decide whether Alice's element is within Bob's subsets, exclusive of revealing to either of them Knowledge about the other party's input beyond the above describe addition. Here the existing work is an alternative protocol for the secure calculation of the union private subsets. The protocol get better when we use hash key function which will we elaborate in proposed work part

The methodology is given in the architecture in the next system. The architecture is good and efficient but the architecture we proposed will decrease the computational cost than the present work.

IV. SYSTEM ARCHITECTURE



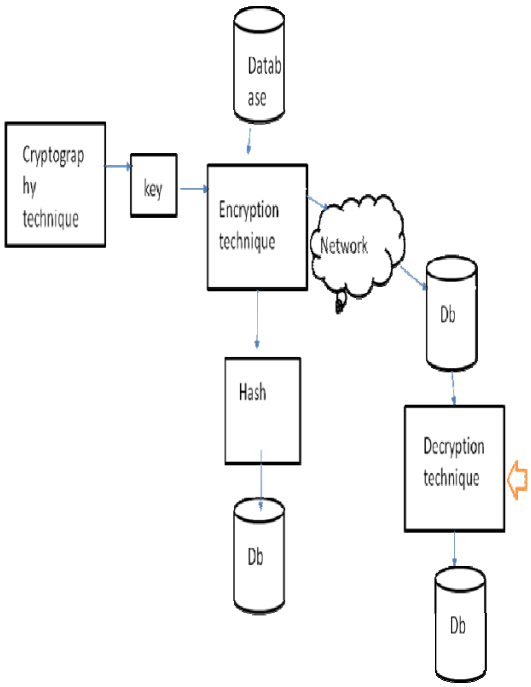
V. PROPOSED WORK

An rising number of databases have become web easily reached from end to end HTML form-based search interfaces. The data units return from the fundamental database are typically encoded into the outcome pages dynamically for human browse. For the programmed data unit to be machine procedure able, which is necessary for many application such as deep web data collected works and internet link shopping, they require to be extract out and assigned meaningful labels. In this paper, we present an automatic explanation approach that primary aligns the data units on a consequence page into dissimilar groups such that the data in the similar group have the similar semantic. Then, for every group we explain it from dissimilar aspect and combined the different annotations to forecast a final explanation label for it. An explanation wrapper for the look for site is automatically construct and can be used to explain new end result pages from the similar web database.

VI. ADVANTAGES

As a rising subject, data mining is playing an increasingly important role in the decision support activity of every walk of life. Get Efficient Item set result based on the customer request.

VII. PROPOSED SYSTEM ARCHITECTURE



VIII. CONCLUSION

The paper gives the brief idea about the enhancement of the existing model through the use of hash key. The automatic generation of the hash key is possible my making the group and enhancing the data mining speed.

REFERENCES

[1] Tamir tassa, “Secure Mining of Association Rules inHorizontally Distributed Databases”, IEEE transactions on knowledge and data engineering, 2013.

[2] LiWu Chang and Ira S. Moskowitz, *Parsimonious downgrading and decision trees applied to the inference problem*, In Proceedings of the 1998 New Security Paradigms Workshop (1998), 82–89.

[3] Mike J. Atallah, Elisa Bertino, Ahmed K. Elmagarmid,Mohamed Ibrahim, and Vassilios S. Verykios, *DisclosureLimitation of Sensitive Rules*, In Proceedings of the IEEE Knolwedge and Data Engineering Workshop (1999), 45–52.

[4] LiWu Chang and Ira S. Moskowitz, *An integrated framework for database inference and privacy protection*,Data and Applications Security (2000), 161–172, Kluwer, IFIP WG 11.3, The Netherlands.

[5] Nabil Adam and John C. Wortmann, *Security-Control Methods for Statistical Databases: A Comparison Study*,ACM Computing Surveys **21** (1989), no. 4, 515–556.

[6] Rakesh Agrawal and Ramakrishnan Srikant, *Privacy-preserving data mining*, In Proceedings of the ACM SIGMOD Conference on Management of Data (2000), 439–450.

[7] David W. Cheung, Jiawei Han, Vincent T. Ng, Ada W. Fu, and Yongjian Fu, *A fast distributed algorithm for*
mining association rules, In Proceedings of the 1996 International Conference on Parallel and Distributed
Information Systems (1996).

[8] Chris Clifton, Murat Kantarcioglou, Xiadong Lin, and Michael Y. Zhu, *Tools for privacy preserving distributed*
data mining, SIGKDDExplorations **4** (2002), no. 2.

[9] Chris Clifton and Donald Marks, *Security and privacy implications of data mining*, In Proceedings of the ACM

SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (1996), 15–19.
[10] Elena Dasseni, Vassilios S. Verykios, Ahmed K. Elmagarmid, and Elisa Bertino, *Hiding Association Rules byusing Confidence and Support*, In Proceedings of the 4th Information Hiding Workshop (2001), 369–383.
[11] Wenliang Du and Mikhail J. Attalah, *Secure multiproblem computation problems and their applications: A review and open problems*, Tech.
[12] Wenliang Du and Zhijun Zhan, *Building decision tree classifier on private data*, In Proceedings of the IEEE ICDM Workshop on Privacy, Security and Data Mining (2002).
[13] Alexandre Ev.mievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke, *Privacy preserving mining of association rules*, In Proceedings of the 8th ACM SIGKDDD International Conference on Knowledge Discovery and Data Mining (2002).
[14] Ioannis Ioannidis, Ananth Grama, and Mikhail Atallah, *A secure protocol for computing dot products in clustered and distributed environments*, In Proceedings of the International Conference on Parallel Processing (2002).
[15] Murat Kantarcioglou and Chris Clifton, *Privacypreserving distributed mining of association rules onhorizontally partitioned data*, In Proceedings of the ACM SIGMOD Workshop on Research Isuues in Data Mining and Knowledge Discovery (2002), 24–31.