iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** V **Month of publication:** May 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Analytical Representation on Secure Mining in Horizontally Distributed Database

Raunak Rathi[1], Prof. A.V.Deorankar[2]
[1,2]Department of Computer Science and Engineering, Government College of Engineering Amravati

Abstract— The security of the large database becomes a serious issue while sharing the data to the network against unauthorized access. However in order to provide the security many researchers cited the issue of Secured Multiparty Computation (SMC) i.e. Secure Third party that allows multiple parties to compute some function of their inputs without disclosing the actual input to one another. Secure sum computation method is popularly and widely accepted due to its simple and thorough solution. The outcomes of our proposed procedure provide a significant result so that it becomes impossible for semi honest party to know the private data of some other sites. Association rule mining is one of the Data Mining techniques used in distributed database. In distributed database the data may be partitioned into fragments and each fragment is assigned to one site. The issue of privacy arises when the data is distributed among multiple sites and no other party wishes to provide their private data to their sites but their main goal is to know the global result obtained by the mining process. However privacy preserving data mining came into the picture. As the database is distributed, different users can access it without interfering with one another. In distributed environment, database is partitioned into disjoint fragments and each site consists of only one fragment. Data can be partitioned in three different ways, that is, horizontal partitioning, vertical partitioning and mixed partitioning. Here we are using horizontal partitioning which is nothing but the data can be partitioned horizontally where each fragment consists of a subset of the records of relation R. Horizontal partitioning divides a table into several tables. The advantage of using horizontal partitioning is, the fact table is partitioned on the basis of time period. Here each time period represents a significant retention period within the business. For example, if the user queries for month to date data then it is appropriate to partition the data into monthly segments. We can reuse the partitioned tables by removing the data in them.
Keywords— secure,multi-party,horizontally,mining.

I. INTRODUCTION

Data mining methodology has emerged as a means of identifying patterns and trends from large quantities of data. Data mining go hand in hand: most tools operate by gathering all data into a central site, then running an algorithm against that data. The problem of computing association rules within such a scenario is addressed. Here there is problem of secure mining of association rules in distributed databases. In that there are various sites that holds the homogeneous databases where databases contain same schema holds the information of different entities. Where the inputs are partial databases and the output is the list of association rules that hold in united database with exceeding level of support and confidence. The aim is to _nd out association rules with predefined level of support and confidence and also protect the content of the information which is not only local but also more global. So that to overcome the problem of security another protocol is proposed for secure computation of union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplifies it significantly and contributes towards much reduced communication and computational costs). While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol. In Data mining, association rule is a popular and well researched method for discovering interesting relation between variables in large databases. Piatetsky-shapiro describes analyzing & presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal et al [3] introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets For example, the rule Found in the sales data of a supermarket would indicate that if a customer buys onions and as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis. Mining association rules plays an important role in knowledge discovery and data mining (KDD). Its purpose is mining the hidden knowledge in databases. The "Mining Generalized Association Rules" problem is mining association rules among items in

hierarchical tree that satisfy minSup and minConf. However, this study does not change the support in different hierarchical levels. The uniform minimum support in each level, items in the same level get the same minimum support. The purpose is mining association rules among item sets in the same level. Another association rule mining method with multiple minimum supports. This allows users identify the different minimum supports for different items, so can mine both frequent and rare rules. However, this method does not traverse the whole hierarchism so that it is very difficult to fund association rules among items in different levels. The research also considered in mining association rules among items in different levels. However, the limitations of this method are still based on Apriori method and used a uniform minimum support for each level [1].

II. IMPLEMENTATION OF PROPOSED WORK

A. *Implementation*

Proposed work involves the use of hash key concept which increases the efficiency of the work. Mainly, hashing involves applying a hashing algorithm to a data item, known as the hashing key, to create a hash value. Hashing is used so that searching a database can be done more efficiently, data can be stored more securely and data transmissions can be check for tampering. The whole concept depends on the concept of hashing so let's see.

End-user has to register by including all details. While registering, user has to select group on which he wants to store his data. Hashing technique is applied at group selection process so that directly memory location of data get accessed.

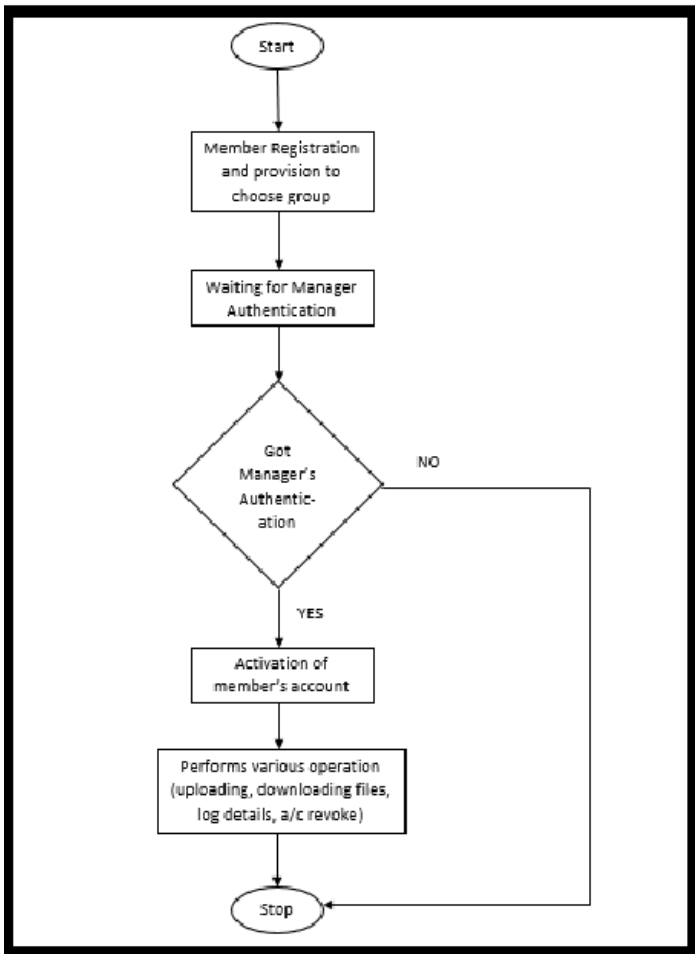


Fig. 1 Flowchart of Proposed Work

Volume 3 Issue V, May 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

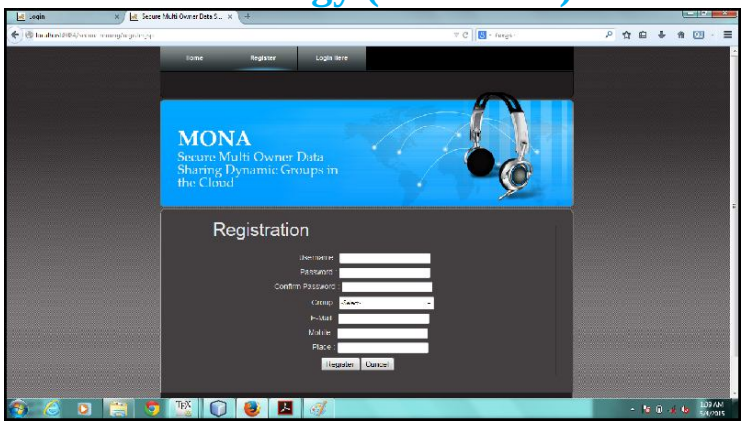


Figure 1: Member Registration

As soon as user creates account, its the job of Group Manager to activate the member registration so that he can do various operation such as
File Upload
File Download
Log Details
Account Revoke

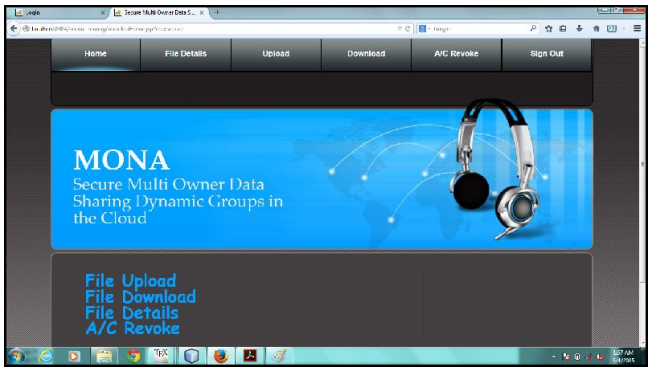


Figure 2: Various operation user can perform after activation

Group Manager activates the end-user by clicking on View Group button given in the following figure.

Figure 3: Member Activation

After activation, member can logged-in and perform operations like uploading, downloading files, deleting registration. For uploading file automatically _le key is generated to provide encryption. While delete the registration it will pop-up one dialog box, to delete an account or not.

Volume 3 Issue V, May 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

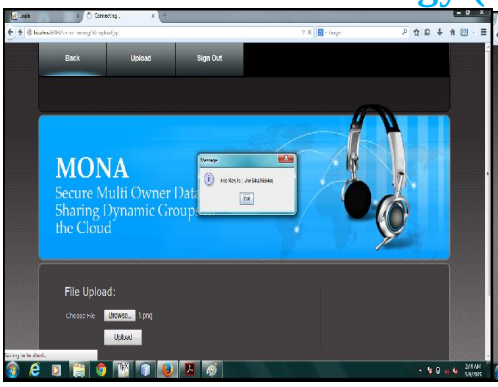


Figure 4: File Uploading

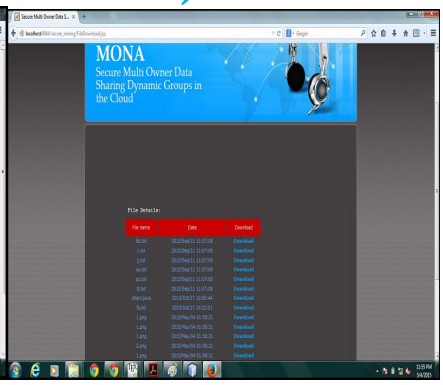


Figure 5: File Downloading

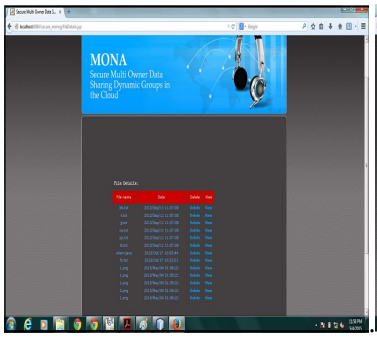


Figure 6: Viewing various files

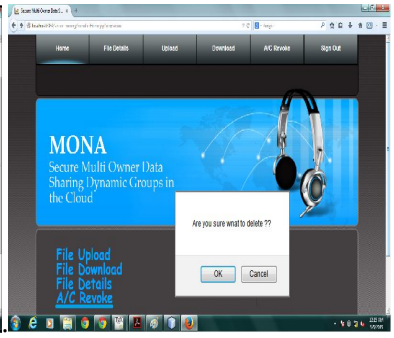


Figure 7: Deleting an account

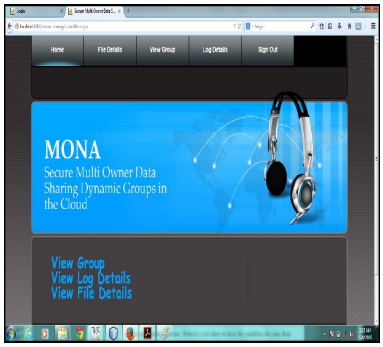


Figure 8: Group Manager Operation

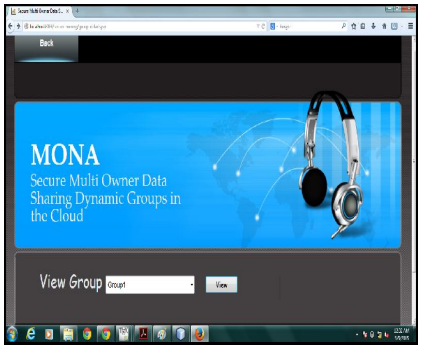


Figure 9: Viewing Profile

Figure 10: Viewing Profile by Selecting Group

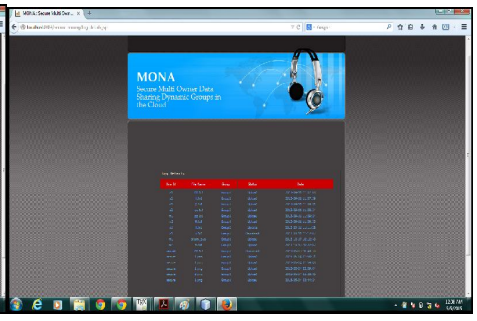


Figure 11: Viewing Log Details

Volume 3 Issue V, May 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A. Performance Analysis

We ran experiment set, where each set tested the dependence of the total computation time and total message size on a different parameter: Above figures describes the computational and communication cost versus the no. of transactions. It has been cleared that no. of transaction has little effect on the run-time of the applied method. Total computational time for the larger value of no. of transaction retains at the same pattern. Total computational time with hash key is 10.5 µsec whereas without hash key is 10.5 µsec. The second set of experiments shows how the computation and communication costs increase with no. of end-user. From the above figures, we came to know that as the no. of end-users increases the computational time required for the method with hash key is less as compared to the

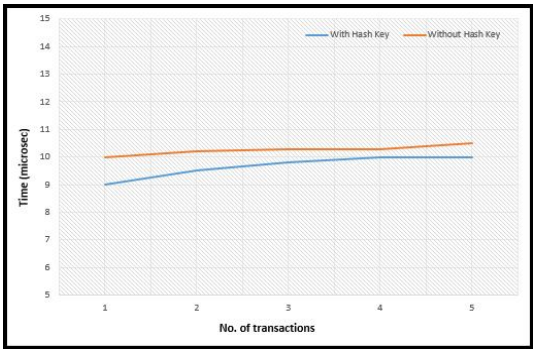


Fig. 12 Total Computation Time with respect to no. of transaction

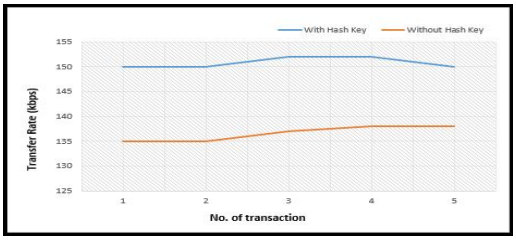


Figure 13: Total Message Time with respect to no. of transaction

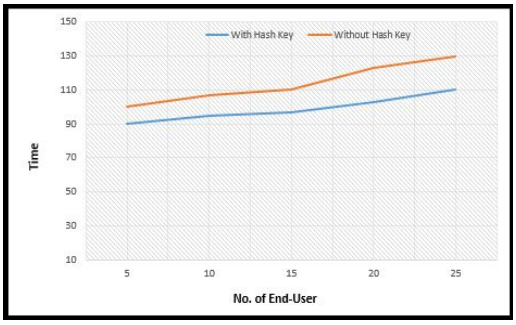


Figure 14: Total Computation Time with respect to no. of end-user

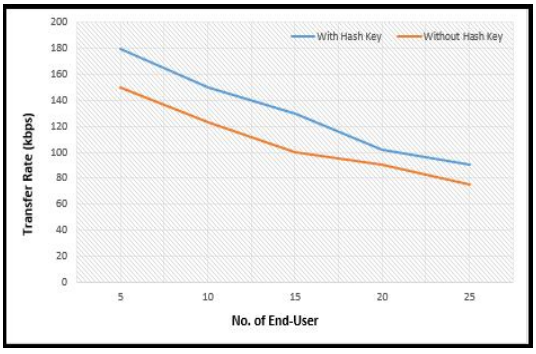


Figure 15: Total Message Time with respect to no. of end-user

computational time required for the method without hash key. Also, the total message time is less in case of without hash key time because of overloading as compared to with hash key method.

III. CONCLUSION

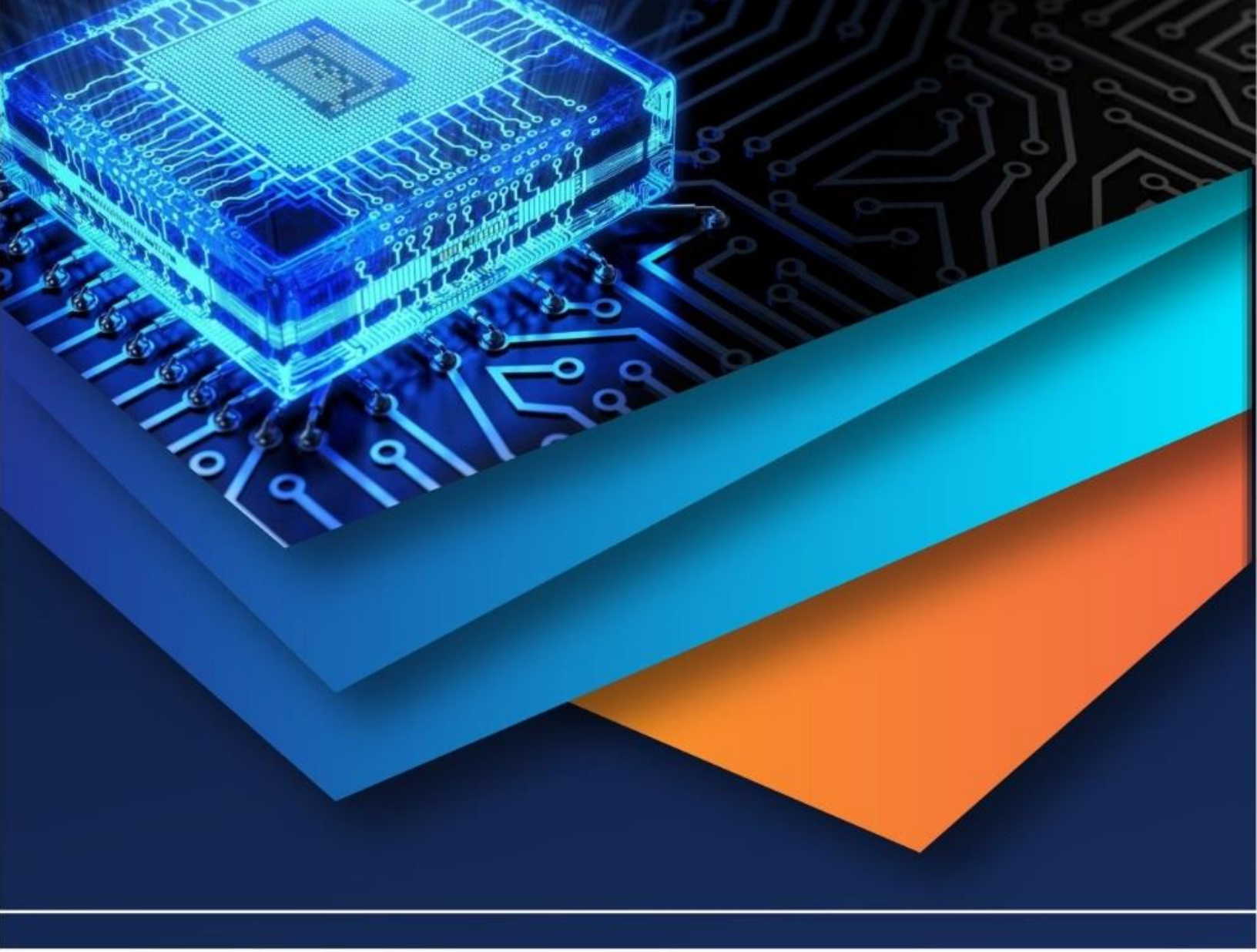
We propose a protocol for secure mining of association rules in horizontally distributed databases. The current leading protocol is that of Kantarcioglu and Clifton. The proposed protocol like theirs, is based on the Fast Distributed Mining (FDM) algorithm, which is an unsecured distributed version of the Apriori algorithm. The main ingredients in our protocol are two novel secure multi-party algorithms-one that computes the union of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. The protocol offers enhanced privacy with respect to the protocol in . In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost. Hence, we proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency. One of the main ingredients in proposed protocol is a novel secure third-party protocol for computing the union (or intersection) of private subsets that each of the interacting players hold. Another ingredient is a protocol that tests the inclusion of an element held by one player in a subset held by another. Those protocols exploit the fact that the underlying problem is of interest only when the number of players is greater than two. The main problem is to devise an e_cient protocol for inequality verifications that uses the existence of a semi-honest third party. Such a protocol might enable to further improve upon the communication and computational costs of the second and third stages of the protocol of [1]. Other research problems is that the implementation of the techniques presented here to the problem of distributed association rule mining in the vertical setting [3], [41], the problem of mining generalized association rules [2]. Also, besides third-party security, the main objective of this secure third-party protocol is to increase the accuracy so that user.

IV. ACKNOWLEDGMENT

We would like to show our sincere gratitude to them who directly or indirectly support us in completion of the manuscript.

REFERENCES

[1] B. Liu, W. Hsu, Y. Ma, "Mining association rules with multiple minimum supports", in: Proc. 1999 Int. Conf. on Knowledge Discovery and Data Mining San
[2] Qiu, L., Li, Y., Wu, "Preserving privacy in association rule mining with bloom
[3]R. Agrawal, R. Srikant, "Fast algorithms for mining generalized association rules", in: Proceedings of the 20th International Conference on Very Large Database
[4] O. Goldreich, "Foundations of Cryptography", Volume 2, Basic Applications,
[5]Han, J., Pei, J., Yin, Y. and Mao,"Mining frequent pattern without candidate generation: a frequent pattern tree approach", Data Mining and Knowledge Discovery, Vol. 8, No. 1, 2004, pp.53-87.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY