# Advance Technique For Privacy Preservation Using Association Rules Based On Fuzzylization

Karishma khatri, Prof. Ruchika Pachori 2
*1Research Scholar, Mahakal Institute of Technology, Ujjain, M.P, India.
2Associate Professor, Mahakal Institute of Technology, Ujjain, M.P, India.
*Department of Information Technology
karishmakhatrirtm@gmail.com

## Abstract:

**In this paper we address the difficulty of the privacy preserving sequential prototype mining on perpendicularly distributed data. we analyzed the privacy-preserving in the term of accurate, efficient. existing privacy-preserving approach based on Boolean association rules ,the incompletely transform measure and moderately transforming method, which is appropriate for the larger ones. We proposed privacy preservation approach using distributed technique based on fuzzylization. Our technique can be used in multi-parties who wish for to together compute the response devoid of useful to each other their uniqueness and their private data.**

**Kewwords: Distributed Technique , Association Rules, Privacy Preservation Approach.**

## I. INTRODUCTION

In privacy preserving disseminated data mining, how the data is partition amongst dissimilar sites is extremely important. The three major partitioning techniques in distributed data base surroundings are horizontal, straight up and mixed mode. In case of horizontal partition, the similar schema is use to continue the data at each site while in straight up partition, dissimilar schemas are used at dissimilar sites, that is, diverse kind of data on the similar entities. The other partitioning technique is mixed partitioning where data is separation horizontally and then every fragment is advance partition into vertical and vice versa. Privacy preserving association rule mining algorithms can be alienated into three categories according to privacy protection technology. The three groups are heuristic-based approach, reconstruction-based approach and cryptography-based technique. In this paper cryptographic technique is accept to discover global association rules by preserving the privacy when no gathering can be treat as trusted party. The cryptography technique is extremely accepted for the subsequent two reason. It has a well conventional and well definite model intended for privacy which can in reality present good quality number of methodologies for verify and validate purpose. Cryptography branch has a extensive assortment of tool set to include privacy in data mining. a little of the appropriate works in privacy preserving data mining are accessible as follow. An indication of data mining technique and a comprehensive explanation of mining association rules are obtainable by the authors and this inspection is completed maintenance in view of data base researcher's point of view based upon the data mining techniques. The authors as well discuss a range of programme of data mining approach and its different features exists amongst them [1].Secure two party multiplication conception was primary introduce by  and presently comprehensive to multi party subtraction. In [2], the authors accessible ID3 classification for two party with horizontally  partitioned data by with secure protocols to accomplish absolute zero acquaintance leakage. The authors proposed in [4], four competent method namely protected sum, secure set union, secure size of set crossroads and scalar creation for privacy preserving data mining in distributed situation. In [5], the author discusses the problem of privacy preserving data mining of association rules while the data is partition horizontally. They proposed algorithm which use three essential ideas such as randomization, encryption of site consequence and secure calculation. The situation of sculpture in the district of privacy preserving data mining approach is accessible [6]. The authors also converse concerning classifications of privacy preserving technique and privacy preserving algorithms such as cryptography-based techniques ,heuristic-based techniques, , and renovation based technique. A framework for assess privacy preserving data mining algorithms and base on this framework single can measure the dissimilar description of privacy preserving algorithms according to dissimilar assessment criterion. An improved scheme is proposed by authors in [7], which a two stage for privacy is preserving disseminated data mining. In [8], the authors discussed the problem of privacy preserving data mining in distributed data bases. They suggested a new paradigm based on two separate entities, a minor and a calculator, both are not having any parts of the data base. They as well obtainable three algorithms based on this concept, one for horizontally partition data, particular for vertically partitioned data and one for several data mining method. The proposed a novel algorithm for mining association rules in dispersed homogeneous databases based on semi honest reproduction and insignificant collision probability. Many another the authors obtainable a classification, an comprehensive explanation and clustering of a variety of association rule mining algorithms. They also recommended additional research directions of privacy preserving association rule mining algorithms by analyzing the alive work. To every approach, the privacy and accuracy are analyzed, and the precision and probability are recognized by experiments.

## II. RELATED WORK

Data mining and information discovery are hot research file relating to the cooperative of artificial intelligence, catalogue and information. It is developed to determine formerly unidentified, potentially constructive knowledge, rules or model [1] from huge databases.

The assumption of data mining and familiarity detection is that the data is release to be used. except in real world it is not essential true. a quantity of database might contain private information that have to not be leak out. Thus method of data mining without leaking the private information is essential. Investigate on privacy preserving data mining is developed for this purpose.

Data mining and knowledge discovery enclose many characteristic problems, counting association rule mining, sequential pattern mining, and classification and clustering. respectively the privacy preserve data mining and knowledge finding should be developed intended at these problems. In this paper we attend to the problem of the privacy preserving in order pattern mining on upright distributed data. Our problem is explain as follow: suppose two parties, PartA and PartB, have confidential data set, D1 and D2, correspondingly, where D1 and D2 are perpendicular distributed databases, i.e., dissimilar sites get together information concerning the same set of entities and assemble dissimilar feature sets. These two party will implement a convinced kind of in order pattern mining algorithm on D1 D2 without continuation of a third party. It require the two party have to not leak their particular private information throughout calculation. Hence we propose a secure two-party calculation protocol based on homomorphic cryptography .

N V Muthu Lakshmi [1]proposed privacy preserve data mining base on random data , privacy preserving data pulling out has obtain wide concentration in the field of data mining and knowledge detection. quite a few techniques, counting data perturbation, encryption, and secure combined algorithm, have been projected in the literature, relating to data mining problems as association rule mining, categorization and clustering.

Arun K Pujari [2] proposed an algorithm to get scalar product of vectors, in which its security is based on the inability of either side to solve k equations in more than k unknowns.

S.Vijayarani [3]proposed planned two-party collaboration Bayesian Networks convince base on Paillier's homomorphic cryptography. except attacks on the protocols are simple to be success. recommend secure scalar invention multiplication protocol will arrangement with this problem.

## III. EXISTING PRIVACY PRESERVING TECHNIQUES

a. Data or rule hiding:
   The PPDM algorithms can be added confidential into Two types, data hiding and rule hiding [14], according to the rationale of hiding.
b. Data distribution
   Distributed data situation can be additional classified interested in horizontal and perpendicular data distributions.
   Cryptography-based approach similar to secure multiparty calculation where a computation is protected if at the end of the computation, no party know something except for its possess input and the consequence.

## IV. PROPOSED METHODOLOGY

The problem of preserving privacy in group rule mining when the database is distributed straight among what time no trusted festivity is measured. A model which adopt a hash based secure sum cryptography technique to discover the comprehensive association rules is proposed in this paper by preserve the privacy constraint. Double hashing function is adopt to improve the privacy additional. The proposed approach powerfully discover comprehensive frequent item sets yet when no site can be treat as trusted. By captivating sample databases, working of the proposed replica is explained. Efficiency of the proposed reproduction is analyzed in terms of privacy and connections and it illustrate that the proposed replica effortlessly and resourcefully find the global frequent item sets by agreeable all the privacy constraint. This replica can be functional for some number of sites and for several number of transactions in the databases of sites. A novel approach which exploit hash based secure sum cryptography is planned in this paper for horizontally partitioned databases without trusted party to find global association rules. The competence of the proposed technique in terms of privacy and communication is converse as follow:

In the procedure of compute partial carry value of every item set at every site, is subtract beginning its local support value and then a value is additional which is compute by subtract received precursor random number from its possess random number. So, ultimately the partial support value of an item set is obtained in masquerading form. To let alone the successor site from guess the predecessor site's private data information, a twofold hash function is distinct in this paper and is use to discover the mask value which will be additional to the masquerading form of inequitable support value to improve the privacy further. As masquerade value is compute by apply two dissimilar hash functions which perform numerous arithmetic computation such as modulus, addition, subtraction and exponentiation, it is not probable for several successor site to predict predecessor site's data information from the external partial support values. every beginning site prepares a list which consists of every nearby frequent item sets of its database of uncommon item sets (positive border item sets) whose support value is earlier to the minimum support to discover whether the item sets in this listing are globally frequent or not by extract every other site's confined supports. Few infrequent item sets are additional to a list of frequent item sets to avoid the situation that a successor site can forecast local frequent item sets of its predecessor site.

Phase 1: is conscious of the total database size and has concluding accumulate partial supports of every frequent item sets of every site. base on this information, Phase 1 can discover definite support of all globally frequent item sets but it is impracticable for Phase 1 to discover any site or sites local supports of any item set since it receive accumulate excess support values merely.

- Ultimately every site obtains the catalogue of global frequent item sets by means of support values received from Phase 1. base on this list no site can predict the input of previous site's database which make the item sets globally frequent as global frequent item sets might or may not be recurrent in all sites. The communication cost is deliberate in distributed situation based on the number of communications for data transfers amongst n sites. The quantity of data transfers at dissimilar stages in the projected replica is particular as follows:
- To broadcast least amount support threshold by Site1 to n-1 number of sites require (n-1) data transfers.
- n quantity of data transfers are necessary for sending every predecessor site's random number to its descendant site.
- A site which initiate to discover global frequent item sets beginning its local frequent item set list require n quantity of data transfers. This assignment is to be perform at every remaining sites. Hence the entirety number of data transfers required to discover every the global frequent item sets of n sites is n2. This is the maximum number of data transfers essential. though if some site or sites are not have local frequent item sets which are not process so far, requirements less number of data transfers compare to n2 data transfers. The minimum number of data transfers necessary forever depends on the databases of sites.
- The Phase 1 necessitate (n-1) data transfers to transmit global frequent item sets along with actual global support. The responsibilities particular in the last two points perform enormity data transfer for moderately a lot of item sets in its place of particular data transfer for each item set. we have proposed an algorithm to prevent

Discovery of sensitive categorization rules. The algorithm places unidentified values in put of identified values in the transactions that carry the sensitive rules. So, from the customized database, the sensitive rules are no longer produce.

## V. CONCLUSION

In this work, we have measured the database privacy effort cause by data mining technology and proposed algorithms for responsive data in association rules mining. The proposed algorithms are base on adapt the database transactions so that the assurance of the association rules can be reduced. more simulation should be carried out to show the possibility and efficiency of the proposed algorithms. alive privacy-preserving technique based on boolean association rules and the incompletely transform measure. The previous one base on moderately transforming method, which is appropriate for the larger ones.

## Reference

[1] N V Muthu Lakshmi1 and Dr. K Sandhya Rani 2,” Privacy Preserving Association Rule Mining Without Trusted Party For Horizontally Partitioned Databases” International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.2, March 2012.

[2] Arun K Pujari “Data Minig Techniques” Second edition 2010, universities(India) Private Limitd 2009.Reprint 2011,2012.

[3] S.Vijayarani, Dr.A.Tamilarasi, R.SeethaLakshmi: Privacy Preserving Data Mining Based on Association Rule- A Survey. International Conference on Communication and Computational Intelligence December,2010.pp.99-103

[4] Vikas Ashok, Ravi Mukkamala Virginia, USA- A Novel Approach To Privacy-Preserving Collaborative Distributed Data Mining WPES’11, October 17, 2011, Chicago, Illinois, USA 2011 ACM 978-1-4503-1002-4/11/10 .

[5] D.Karthikeswarant, V.M.Sudha, V.M.SureshA.Javed sultan a pattern based framework for privacy preservation through association rule mining IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.

[5] EHUD GUDES, TAMIR TASS, Secure Distributed Computation of Anonymized Views of Shared Databases ACM Transactions on Database Systems, Vol. 37, No. 2, Article 11, Publication date: May 2012.

[6] ArashGhorbanniaDelavar, NarjesRohani, Mehdi ZekriyapanahGashti,” ERPAC : A Novel Framework for Integrated Distributed Systems Using Data Mining Mechanisms” 2nd International Conference on Software Technology and Engineering(ICSTE) -2010.

[7] Liming Li, Qishan Zhang ,”A Privacy Preserving Clustering Technique Using Hybrid Data Transformation Method”, Proceedings of IEEE International Conference on Grey Systems and Intelligent Services, Nanjing, China, November 10-12, 2009.

[8] Aris Gkoulalas - Divanis and Vassilios S. Verykios , “Association Rule Hiding for Data Mining” , Advances in Database Systems ,Volume 41, Springer, 2010.

[9] T. Berberoglu and M. Kaya, “Hiding Fuzzy Association Rules in Quantitative Data”,The 3rd InternationalConference on Grid and Pervasive Computing Workshops, May 2008, pp. 387- 392.

[10] Manoj Gupta and R. C. Joshi, “Privacy Preserving Fuzzy Association Rules in in Quantitative Data”, International Journal of Computer Theory and Engineering, Vol. 1, No. 4, October, 2009, 382-388.

Elisa Bertino , Igor Nai Fovino Loredana Parasiliti Provenza (2005), A Framework for

Evaluating Privacy Preserving Data Mining Algorithms, Data Mining and Knowledge

Discovery, Vol. 11, 121–154.

[8] Chin-Chen Chang, Jieh-Shan Yeh, and Yu-Chiang Li (2006), Privacy-Preserving Mining of Association Rules on DistributedDatabases, IJCSNS International Journal of Computer Science and Network Security, Vol.6 No.11.

[10] Alex Gurevich, Ehud Gudes (2006), Privacy preserving data mining algorithms without the use of secure computation or perturbation, 10th international database Engineering and Applications Symposium IDEAS06 IEEE.

[11] Mahmoud Hussein, Ashraf El-Sisi,and Nabil Ismail (2008), Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Data Base, I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008,Part II,LNAI 5178, pp. 607–616, 2008.© Springer-Verlag Berlin Heidelberg.

[12] Jian Wang, Yongcheng Luo, Yan Zhao, Jiajin Le(2009), A Survey on Privacy Preserving Data Mining, First International Workshop on Database Technology and Applications, pp. 111-114.

[13] Anita A. Parmar1,Udai Pratap Rao2, Dhiren R. Patel3,” Blocking based approach for classification Rule hiding to Preserve the Privacy in Database” 978-0-7695-4443-4/11 - 2011 IEEE DOI 10.1109/ISCCS.2011.103.

[14] Dr.A.M.Natarajan, R.R.Rajalaxmi, “A Hybrid Data Transformation Approach for Privacy Preserving Clustering of Categorical Data” , Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, Springer ,2007.