ADVANCE SOLUTION FOR PRIVACY PRESERVING IN DISTRIBUTED DATA MINING

[1]Ms. A.G. Raut, [2]Dr. Mrs. S. S. Sherekar, [3]Dr.V. M. Thakare

[1]Student ME, [2]Professor, [3]Professor

[1] PG Department of Computer Science and Engineering,
[1]SGBAU, Amravati, India

Abstract : Association rule mining and frequent item set mining are two popular and widely used data analysis techniques for a range of applications. In this paper, focus is on privacy-preserving mining on vertically partitioned databases. In such a scenario, data owners wish to learn the association rules or frequent item sets from a collective data set and disclose as little information about their (sensitive) raw data as possible to other data owners and third parties. Stimulated by developments such as cloud computing, there has been considerable recent interest in the model of data mining-as-a-service. A company (data owner) absent in expertise or computational resources can outsource its mining to a third party service provider (server).

IndexTerms - Association rule Mining ,Privacy-Preserving outsourcing.

I. INTRODUCTION

In the Random projection approach the path finder is the probability of using multiplicative random projection matrices for privacy preserving distributed data mining. It specifically considers the problem of computing statistical aggregates like the inner product matrix, correlation coefficient matrix, and Euclidean distance matrix from distributed privacy sensitive data possibly owned by multiple parties. Random projection class of category is directly equivalent to many other data-mining problems such as clustering, principal component analysis, and classification [1].Some methods make initial contributions on two divergent grounds. This paper, discusses five different privacy technique, to ensure data privacy, design an efficient homomorphic encryption scheme and a secure comparison scheme [2].Also proposed a cloud-aided frequent itemset mining solution, which is used to build an association rule mining solution [3].In comparison to the only known solution achieving a similar privacy level as proposed solutions, the performance of proposed solutions is three to five orders of magnitude higher [4].To preserve privacy while mining large amounts of data distributed among different parties, cryptography based privacy preserving distributed data mining (PPDDM) has emerged as an important alternative [5].

This paper implements the proposed protocols and analyze the computation and communication cost, and security. It also compare the performance of the proposed protocols with the existing ID3-based protocols. Main contributionis to realize that RDTs can provide good security with very high efficiency.

II. BACKGROUND

As per the studies on Privacy Preserving many models and approaches have been develop for the Providing Privacy in data mining in recent past years. Such approaches are:

A Random Decision tree framework for privacy –Preserving data mining approach is based on Random decision trees in RDT same code can be used for multiple tasks [1]. Incentive compatible privacy-preserving distributed classification method propose game theoretic mechanism to encourage truthful data sharing for distributed data mining [2]. Privacy preserving outsourced association rule mining on vertically portioned databases technique propose a cloud aided privacy preserving frequent itemset mining solution for vertically portioned database, which is then used to build a privacy preserving association rule mining solution [3].Random projection basesd multiplicative data perturbation for perturbation for privacy preserving distributed data mining method specifically considers the problem of computing statistical aggregates like the inner product matrix and Euclidean distance matrix from distributed privacy sensitive data possibly owned by multiple parties Privacy-Preserving mining of association rules from outsourced transaction databases method study the problem of outsourcing the association rule mining task within a corporate privacy preserving framework [4].Privacy-Preserving outsourced Association rule mining on vertically partitioned databases method focus on privacy preserving mining on vertically partitioned databases [5].This paper introduces five technique for privacy preserving Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining, Privacy-preserving mining of association rules from outsourced transaction databases, Incentive compatible privacy preserving distributed classification, Privacy Preserving outsourced association rule mining on vertically partitioned databases, A random decision tree framework for privacy preserving data mining.

The paper is organized as follows:

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on mining techniques. **Section VI** proposed method and outcome of result. Finally **Section VII** Conclude this analytical paper.

III. PREVIOUS WORK DONE

In research literature, many privacy Preserving techniques are used for provide effictiveness, scalability.

Kun Liuet al.(2015) [1] has proposed an Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining worked on with the exception of, it do not utilize a third-party server to server to compute the mining result. Some solutions use asymmetric encryption to compute the supports of itemsets, while other solutions use a secure scalar product protocol.

Fosca Giannotti et al. (2013) [2] has proposed an Privacy Preserving Mining of Association Rules From Outsourced Transaction Databases in that developed an encryption scheme, called Rob Frugal, that is based on 1-1 substitution ciphers for items and adding the fake transaction to make each cipher item share the same frequency as > K-.

Robert Nix et al. (2012) [3]has proposedan Incentive Compatible Privacy Preserving Distributed Classification in this method which reward players based on their contribution to the model is incentive compatible. If a deviation from the truth affords a player a payout advantage, then this means that the deviation has necessarily increased the overall accuracy of the final classifier.

Lichun Liet al.(2014) [4] has worked on Privacy Preserving outsourced Association Rule Mining on Vertically Partitioned databases in this method propose a symmetric homomorphic encryption scheme (using only modular additions and multiplications), which is significantly more efficient than asymmetric schemes.

Jaideep Vaidya, et al.(2014) [5] has proposed A Random Decision Tree Framework for privacy –Preserving Data Mining in that method the proposed scheme implement the RDTs algorithm builds multiple (or m) is o-depth RDTs. One important aspect of RDTs is that the structure of a random tree is constructed completely independent of the training data. The RDT algorithm can be broken into two stages, training and classification.

IV. EXISTING METHODOLOGIES

Many Privacy preserving approach have been implemented over the last several decades. There are different technique that are implemented for providing privacy in data mining i.e Random Projection base multiplicative data perturbation for privacy preserving data mining, Privacy-preserving mining of association rules from outsourced transaction databases, Incentive compatible privacy preserving distributed classification, Privacy Preserving outsourced association rule mining on vertically partitioned databases, A random decision tree framework for privacy preserving data mining,

A) Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining:

In the Random projection approach it path finder the probability of using multiplicative random projection matrices for privacy preserving distributed data mining .It specifically considers the problem of computing statistical aggregates like the inner product matrix, correlation coefficient matrix, and Euclidean distance matrix from distributed privacy sensitive data possibly owned by multiple parties. This method makes initial contributions on two divergent grounds [1].

$$U(t)=Rx(t);$$

Where, x(t)=(x1(t),x2(t); . . . xm(t) denotes a m- dimensional vector collecting the m independent source signals xi(t)= 1; 2; . . .;m. Here, t indicates the time dependence.

B) Privacy-preserving mining of association rules from outsourced transaction databases:

Spurred by developments such as cloud computing, there has been considerable recent interest in the paradigm of data mining-as-a-service. 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 [2].

C) Incentive compatible privacy preserving distributed classification:

In this method, propose game-theoretic mechanisms to encourage truthful data sharing for distributed data mining. One proposed mechanism uses the classic Vickrey-Clarke-Groves (VCG) mechanism, and the other relies on the Shapley value. Instead, incentivize truth telling based solely on the data mining result. This is especially useful for situations here privacy concerns prevent verification of the data. Under reasonable assumptions, prove that these mechanisms are incentive compatible for distributed data mining. [3].

D) Privacy Preserving outsourced association rule mining on vertically partitioned databases:

In this paper focus on privacy-preserving mining on vertically partitioned databases. In such a scenario, data owners wish to learn the association rules or frequent itemsets from a collective data set and disclose as little information about their (sensitive) raw data as possible to other data owners and third parties. To ensure data privacy, design an efficient homomorphic encryption scheme and a secure comparison scheme. Then propose a cloud-aided frequent itemset mining solution, which is used to build an association rule mining solution.[4].

$$EPK\ (m1 + m2) = EPK(m1) \times EPK\ (m2)$$

$$EPK\ (m1 \times m2) = EPK(m1) \times EPK\ (m1) \times ...EPK(m1)(m2\ \text{multiplications}) = EPK(m1)m2$$

Where, EPK () be the function of encrypting with the public key, and "•" is modular multiplication in Paillier.

E) A random decision tree framework for privacy preserving data mining:

In this method, develop methods to securely construct RDTs for both horizontally and vertically partitioned data sets. also implement the proposed protocols and analyze the computation and communication cost, and security. Also compare the performance of the proposed protocols with the existing ID3-based protocols .In this method main contribution is to realize that RDTs can provide good security with very high efficiency.[5].

ANALYSIS AND DISCUSSION

Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining focus on how much privacy perturbation technique can preserve when the adversary has different kinds of prior knowledge of the data and when the basic assumptions of this technique are not satisfied.[1].Privacy-preserving mining of association rules from outsourced transaction databases method study would be interesting to enhance the framework and the analysis by appealing to cryptographic notions such as perfect secrecy[2].Incentive compatible privacy preserving distributed classification in this method discuss how to incentivize data sharing in privacy-preserving distributed data mining applications.[3].Privacy Preserving outsourced association rule mining on vertically partitioned databases uses comparison scheme based on the symmetric homomorphic encryption scheme[4]. A random decision tree framework for privacy preserving data mining method derive the computation and communication cost, assuming k sites, n attributes, j instances, p class values, and m random trees, and then go through the security analysis.[5].

TABLE 1:Comparison between different privacy preserving techniques.

Proposed techniques	**Advantages**	**Disadvantages**
A Random Decision Tree Framework for Privacy-Preserving Data Mining	This proposed method provide many benefit over the traditional classification. also this method provide more accuracy.	This proposed method can't work for general solution that can work for arbitrarily portioned data.
Privacy Preserving Outsourced Association Rule Mining On vertically Partitioned Databases.	This method improve the performance of the fault prediction model. The fault prediction models are then validated using a cost evaluation framework.	The drawback of this method is cross validation cannot be used in practice..
Privacy Preserving Mining of Association Rules From Outsourced Transaction Databases.	This proposed method provide better performance Scalability and protect Privacy.	The drawback of this method need to improve the encryption schemes and Rob Frugal algorithm to minimize the number of spurious pattern.
Incentive Compatible Privacy Preserving Distributed Classification	In this approach is that players based on their contribution to the model is incentive compatible reward.	This method have more cost.

Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining	In this approach, explores the probability of using multiplicative random projection matrics for privacy preserving distributed data mining.	This method need to abundant arability of personal data.

. PROPOSED METHODOLOGY

Providing privacy in data mining is important and difficult task to analyse and discuss about various methods based on different parameters i.e. accuracy, quality, cost, time, flexibility, effectiveness, etc for different privacy preserving models. There are still problems which trouble in this field. New privacy preserving method called "Privacy Preserving for data mining " is proposed for more effective and more accurate privacy preserving so as to overcome the problems of previous models..The proposed method uses encryption decryption algorithm that is easier as compare to other methods.

Diagrammatic representation of proposed method is shown as follows:

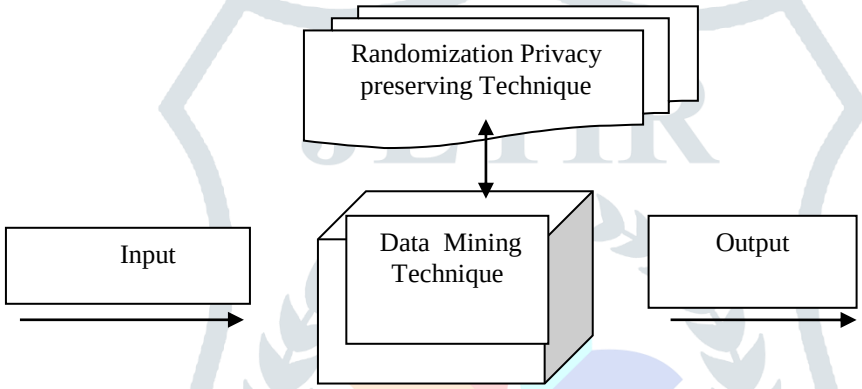

Fig**:** Block Diagram of PPDM technique

In above figure there is input given and on that input the data mining technique is performed. Also randomization method of privacy preserving is applied. The proposed method uses encryption decryption algorithm that is easier as compare to other methods. Then the possible result is desired.

VII. OUTCOMES AND POSSIBLE RESULT

In this way the proposed method is to perform effectively. A number of algorithmic techniques have been designed for the Privacy preserving data mining Methods of randomization, K- anoymization and distributed privacy preserving data mining.With the help of the randomization algorithm of the proposed method, numerical and categorical data performance of model is enhanced.

VIII. CONCLUSION

In this way the proposed method is to perform effectively. A number of algorithmic techniques have been designed for the Privacy preserving data mining Methods of randomization, K- anoymization and distributed privacy preserving data mining. With the help of the randomization algorithm of the proposed method, numerical and categorical data performance of model is enhanced. The proposed scheme implements and demonstrated that general and efficient distributed privacy preserving knowledge discovery is truly feasible.

IX. FUTURE SCOPE

From observations of the proposed method the future work will includeother geometric transformation techniques like scaling, translation, and rotation and also Demonstrating the utility of the proposed homomorphic encryption scheme .

REFERENCES

[1] Kun Liu, Hillol Kargupta, Senior Member, IEEE, and Jessica Ryan,“ Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining”, IEEE TRANSACTIONS ON KNOWLEDGE ANDDATA ENGINEERING,VOL.18,NO.1, JANUARY 2015.

[2] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang“Privacy-preserving mining of association rules from outsourced transaction databases*”*,IEEE SYSTEMS JOURNAL,VOL.7,NO.3 2013.

[3] Robert Nix and Murat Kantarcioglu, “Incentive compatible privacy preservingdistributed classification”,IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 09, NO. 4, AUGUST 2012.
[4] Lichun Li, “Privacy Preserving outsourced association rule mining on vertically partitioned databases”, IET Software, Vol. 8, Iss. 3, 2014.

[5] Jaideep Vaidya, Senior Member, IEEE, Basit Shafiq, Member, IEEE, Wei Fan, Member, IEEE, Danish Mehmood, and David Lorenzi,“A random decision tree framework for privacy preserving data mining”, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, Vol.11, NO.5, OCTOBER 2014.