Privacy preserving Data mining methods

Miss. S. M. Dolas, Dr. V. M. Thakare, Prof. Y. M. Kurwade

Abstract-
Sensitive original information/data is extract of protected databases whether it is received or cut from the original database. The purpose of doing this is to prevent individual privacy against adverse data received. This paper focused on five different techniques such as tree framework, PPDM, Association mining rule, Rampart framework, anonymizing trajectories. . But some problems are persisting in each method. The paper proposes the method to overcome the existing problems and the improved method "Secure multi party computation and restricted data" is proposed in this paper.

Keywords-
Data mining, Metrics, Association rules, Data collector, Suppression.

I. INTRODUCTION

The overall goal of the data mining process is to extract the information and transfer it for the further use it is not the extract the data from itself. Because of the transforming the data it reduces the utility. To protect the leakage information the data process is used. PPDM is used to avoid the direct use of data or information which is available. This paper, discusses five different privacy schemes such as Tree framework[1], PPDM[2], Association mining rule[3], Rampart framework[4], Anonymizing trajectories[5].But these methods also have some problem so to overcome such problems improve version of mobility scheme that is "Secure multi party computation and restricted data" directions from web servers under the assailant's control [6]. Accordingly, distinguishing bots with electronic controlling is more unpredictable than bots with IRC-based controlling. In this study, we have experienced different systems for HTTP botnet discovery and techniques utilized in them.

II. BACKGROUND

Some studies on privacy models have been done to develop the privacy scheme in recent past years. Such schemes are:

Tree framework is proposed to define that gives a solution so that there is improved efficiency are done. so because of these it gives more security[1].PPDM methodology are construct for the privacy level so the use of the data largely used. so the performance of the data mining process of transferring of data are efficient[2].Association rule mining is conducted by the third party. The privacy preserving association rule mining increment the data and this is control by the target system. It is one time mining model because the overall data are not change through the process that is it does not increase or decrease through the process [3].The Rampart framework gives the security among the unnecessary data and gives the solution and also it can encourage for the variety of problem which are related to the privacy[4]. The anonymizing trajectories use the data before it publishing and then after this data are use for an analytical use. The three anonymizing algorithm are used are suppression of location, splitting of trajectory or both are used in these trajectories [5].
This paper introduces five privacy scheme i.e., a tree framework, PPDM, Association mining rule, Rampart framework, anonymizing trajectories.
The paper is organized as follows. **Section 1** Introduction. **Section 2** discusses Background. **Section 3** discusses previous work. **Section 4** discusses existing methodologies. **Section 5**discusses attributes and parameters and how these are affected on privacy models. **Section 6** proposed method and outcome result possible. **Section 7** conclude the outcome and possible results of paper. Finally **Section 8** includes the Conclusion.

III. PREVIOUS WORK DONE

In research literature, many privacy models have been studied to provide various privacy schemes and improve the performance in terms of transferring the data and improve their efficiency. JaideepVaidya et.al [1] has proposed A Random Decision Tree Framework for Privacy-Preserving Data Mining exploit the fact that RDTs can naturally fit into a parallel and fully distributed architecture. Z. Zhang et.al [2] has proposed a Cost-friendly differential privacy for smart meters exploiting the dual roles of the noise the pattern is defined as an expression to describe a subset of data or a model applicable to a subset. Y.Yang et.al [3] has proposed a differential privacy in data publication and analysis also it can measure the performance of association rule mining after applying differential privacy. G. Barbier et.al [4] has proposed a framework for categorizing and applying privacy preservation technique in big data mining that provenance data in the context of social networks, leverages special social

network features, including user profiles and interaction type and time, to obtain information provenance. D. Kopanaki et.al [5] has proposed cares about others privacy: Personalized anonymization of moving object trajectories and extend the work of personalized privacy parameters and with a technique that partitions trajectories to sub-trajectories.

IV. EXISTING METHODOLOGIES

Many mobility schemes have been implemented over the last several decades. There are different methodologies that are implemented for different mobility models i.e., tree framework, PPDM, Association mining rule, Rampart framework, anonymizing trajectories.

*4.1 Tree framework***:** The tree can be generated by the list of an attributes. The structure of the tree can be independent on the structured data. The two phases are generated training and classification. Training are used for building the tree structure and classification are used for the dividing the tree structure. The tree can be generated by random choosing the features it does not depend on the training data. The tree does not grow when it achieves the highest limit. The tree can be classified through their nodes in tree. The classification of tree is depends on their leaves. Also the data are scanned for update the random tree [1].

4.2 PPDM:

Mainly the PPDM are applied on the field of scientific interest. Their trade is between privacy and their utility. The PPDM method is depend on the data lifecycle. The lifecycle of he PPDM depends on the data collection, data publishing, data distribution, and the output of the data mining. The PPDM technique can modify the data or remove the data according their needs [2].

4.3 Association mining rule:

The association rule mining used the apriori algorithm. The algorithm first finds the combination of the item sets and then it derives the association rule. The privacy preserving association rule mining is nothing but the order dependency of the wavelet transform. There is a group is foam of the similar data and further it get combined so that the association rule can be foamed. If there is a group is foam the mining rule can be applied easily otherwise it get take a lots of amount of time. The association mining generated from the incremental mask data which is related to the mining rule [3].

4.4 Rampart framework:

The Rampart framework divides the approaches of the protection and it gives the solution associated with data which is discover from the knowledge. The Rampart gives both the technical solution and non-technical solution both. The technical solution consist of restriction, aggregation approaches and the non-technical solution consist of laws and regulation. Also it gives the approaches for solving the privacy issues. The scope of this method is that it can broadcast the privacy issues related to data mining [4].

4.5 Anonymizing trajectories:

The use of this trajectory is that it prevent the adversaries from their temporary knowledge. Suppose a person uses a card in a store for their requirement. Before it uses the company can match the person identity to their database. In many companies from person card the company can stores the person information. If unknown person uses the person card than there is the mass attack are applied to that unknown person. The card knows all about the person so this knowledge can anonymize the trajectory before it use. This is the main use of anonymizing methodology [5].

V. ANALYSIS AND DISCUSSION

The tree can shows that the positive and the perfect model can be generated with the less cost. The use of this tree is efferent and hence the Bayes optimal classifier (BOC) are implemented. The randomization tree technique improved efficiency and security from a different decision tree framework [1]. The Privacy level measures how the data is secure. The quantification of data mining algorithm and the machine learning in cyber security are evolved. The data are taken from the multiple sensors and because of these multiple sensors there is reduce the network traffic and hence it improve the battery life and sensor lifetime [2]. Association mining rule parameters both the high level threshold and low level threshold are effects on the time of the execution. The fewer rule are applied to the high threshold hence it takes the less execution of time[3]. The Rampart framework scheme shows that how to modify or reconstruct the update model for security of the sensitive information or data [4]. Anonymizing trajectories scheme showcased the effective of anonymizing algorithm in terms of utility preservation of data and its efficiency [5].

Mobility scheme	**Advantages**	**Disadvantages**
Tree Framework	Can be combine with the other decision techniques. Important insights can be generated based on expert describing the situation (its alternatives, probabilities and costs) and their preferences for outcomes.	They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree. They are often relatively inaccurate. Many other predicators performs better with similar data. This can be remedied by replacing a single decision tree with the random tree of the decision tree
PPDM	PPDM is a very advantageous in	One of the major problem of privacy

	development of a various data mining techniques. It allows sharing of large amount of privacy sensitive data for analysis purpose.	preserving data mining is the abundant availability of the personal data.
Association mining rule	This method improves the high interpretability and it can easily incorporation of previous knowledge.	This method has limited options for handling the missing values.
Rampart framework	It can protect the data in mined data. The rampart framework encourage the solution that associated with privacy problem.	The modified data is less useful so because of this the data mining must balance the data utility.
Anonymizing trajectories	This algorithm can easily transfer long and detailed projection to the smaller and simple one.	It does not provide the guarantee privacy against the other adversaries.

TABLE 1: Comparisons between different privacy schemes.

VI. PROPOSED METHODOLOGY

The privacy preserving is a process of providing the security for sensitive data. The sensitive data like an employee salary, annual income of company, transferring the money from one account to another account etc. providing the security to this data is very important. Privacy preserving scheme is important and difficult task to analyse and discuss about various methods based on different parameters i.e., diameter, average, distance, time, cost, accuracy, efficiency etc. for different privacy preserving models. There are still problems which trouble in this field such as storing the personal data about the user because of increasingly ability of the user. New privacy preserving method called "Secure multi party computation and restricted data" privacy model for preserving the data is propose here to overcome the problems of this scheme. As this scheme is depend upon the target location and the source location and current location is calculated using the central location while it is flowing through the source point.

As privacy preserving data mining has become increasingly popular because it allows sharing of privacy sensitive data for an analysis purpose. When database has query to send to query handler and these handling query can accepting the query data from the client and process the query with the data base and fetching the datasets from the data base. So by applying appropriate method that is Secure multi party computation and restricted data when the query is transmitted to query handler the this method will apply on the privacy preserving algorithm. It can learn from the environment or data with keeping the data privacy and transferring that data for secure purpose. Data hiding investigate about maintaining the privacy of data or information. These investigate concentrate on the exclusion of private information from the database before sharing the data with others. Sampling, transformation are the general techniques used to create transformed database which is apply on that secure data. In this way, when database is moves through the dataset, then with the help of this method it is easy to send the query to database without taking any non-useful data because this method perform as soon as any database moves through the dataset. With the help of secure multi party computation and restricted data concept the proposed method performed in small space and in less interval of time.

Basic steps of algorithm:

Step1: A database is an organized collection of the data and that database collect the queries, schemes and the data.

Step2: due to the database the query handler can accepting the query data from the client and process the query with the data base and fetching the datasets from the data base.

Step3: those data can be handle by the query handler are transmitted to the privacy preserving algorithm.

Step4: The privacy preserving algorithm can learn from the environment or data with keeping the data privacy and transferring that data for secure purpose.

Step5: Data hiding investigate about maintaining the privacy of data or information. These investigate concentrate on the exclusion of private information from the database before sharing the data with others. Sampling, transformation are the general techniques used to create transformed database.

Diagrammatic representation of proposed method is shown as follows:

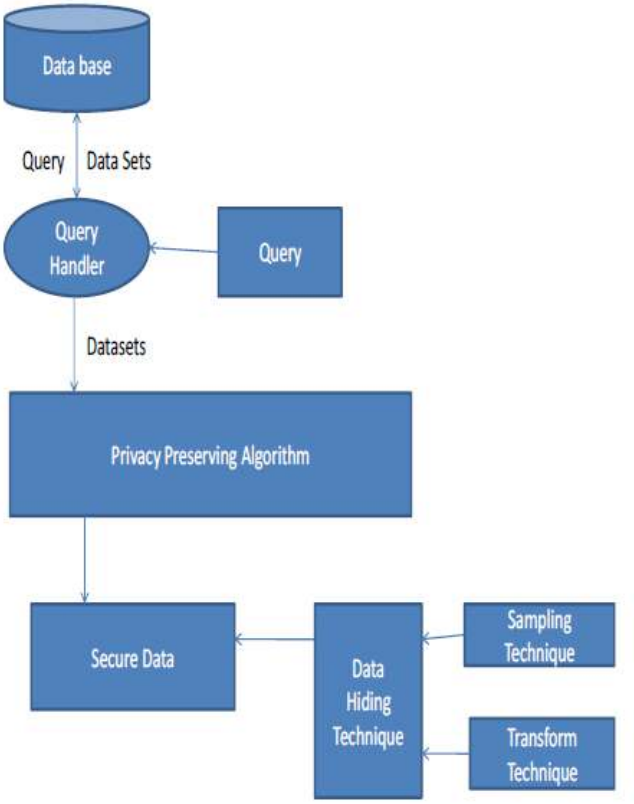

Fig: Block diagram of Reconstruction algorithm for PPDM.

VII. OUTCOME AND POSSIBLE RESULT

In this way the proposed method is perform for the privacy preserving scheme when database moves through the datasets. With the help of the speed and direction the proposed method calculate location of movable query in less time and store in database for sending the number of queries without having any delay.

CONCLUSION

This paper focused on the study of various privacy preserving scheme i.e., tree framework, PPDM, Association mining rule, Rampart framework, anonymizing trajectories**.** But there are some privacy problems associated with knowledge discovery from data so to improve this "Secure multi party computation and restricted data" privacy method for data mining is proposed here. When database moves through the dataset then the propose method provide the location for movement of the query in less time.

FUTURE SCOPE

From observations of the proposed method the future work will conduct a formal analysis of privacy with the help of heuristic analysis of privacy for an numeric data.

REFERENCES

1. JaideepVaidya, Senior Member,BasitShafiq "Tree frameworkfor the random decision" IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, Vol. 11, No. 5, September/October 2014.
2. Ricardo Mendes and Joaqo P. VllelaClsuc "Data mining for privacy preserving " IEEE ACCESS, Vol. 5, June 2017.
3. Madhu V. Ahluwalia, AryyaGangopadhyay, "Association rule mining for incremental the data"IEEE TRANSACTIONS ON SERVICES COMPUTING, Vol. 10, No. 4, July/August 2017.