www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 12 December 2014, Page No. 9568-9570

A Framework for Privacy Preserving Collaborative Data Mining

Gottipamula Padmavathi[1], T.V. Ramanamma[2]

[1]M.Tech, CSE, CSE Department,
Aurora's Technological & Research Institute, Hyderabad, Telangana, India
gpadmavathi005@gmail.com

[2]M.Tech, Sr.Asst.Professor, CSE Department,
Aurora's Technological & Research Institute, Hyderabad, Telangana, India

Abstract: ***Privacy Preserving Data Mining has become popular now-a-days to restrict the access of data from unauthorized parties i.e., it guarantees the protection of individual records of particular party. In order to make this possible Privacy Preserving Data Mining provides many techniques like Randomization & cryptography. In our application cryptographic technique has been used to provide security to parties data. Specifically, DES algorithm has been used to encrypt and decrypt the data which has been received from different parties and also, apriori algorithm has been used for analysis of collaborative data of multiple parties. Finally, we proposed an algorithm named privacy preserving Collaborative Data Mining (PPCDM) for successful realization of our framework.***

Keywords: Security, privacy-preserving data mining, horizontally partitioned data, vertically partitioned data.

I. Introduction

Data mining is used by enterprises in order to obtain business intelligence. However, there are some privacy issues that are to be addressed. Privacy preserving data mining is the research topic that assumed importance. Privacy or identity of the objects involved in data mining needs to be preserved. Many researchers contributed towards data mining and privacy preserving data mining. Recently researchers were focusing on collaborative data mining where multiple competing parties are involved. In the real world enterprises, multiple and related companies might need to have collaborative data mining for obtaining business intelligence to take well informed decisions. In such scenarios there is possibility of disclosing privacy which leads to security problems. There are some instances that reveal inference attacks launched by insiders and outsiders. In this context, privacy preserving collaborative data mining [1] assumes much importance.

The researchers in [2] focused on horizontally partitioned data. Other researches in [3] and[4] focused on vertically partitioned data. The competing parties are supposed to give genuine data as explored in [5] and [6]. Recently Kantarcioglu et al.[7] proposed incentive compatible approach that motivates competing parties to provide genuine data instead of giving less than ideal or incompatible data. Inspired by this research, in this paper, we propose a framework for privacy preserving collaborative data mining besides using incentives to enable competing parties to give genuine inputs.

The remainder of the paper is structured as follows. Section II provides review of literature on the prior works. Section III presents the proposed system to perform privacy preserving collaborative data mining. Section IV presents experimental results while Section V concludes the paper besides providing directions for future work.

.

II. Related Work

This section reviews the literature on related works. Privacy preserving data mining [8] is very important as the data mining should not disclose sensitive details. The researchers in [2] focused on horizontally partitioned data. Other researches in [3] and [4] focused on vertically partitioned data. Many privacy preserving data mining solutions assumed that the participants provide data truthfully. In the case of horizontally partitioned data, different sites collect the same set of information about different entities. The collaborative data mining where multiple participants or competing parties provide compatible data in a distributed environment it is important that all parties provide genuine data. In the case of vertically partitioned data, we assume that different sites collect information about the same set of entities, but they collect different feature sets.

The competing parties are supposed to give genuine data as explored in [5] and [6]. In this paper we follow an approach that provides incentives to the genuine data providers. Among the competing parties, it is observed for any incompatible data. When incompatible data is provided by any party, the incentives are reduced otherwise the incentives are increased. The idea is to ensure that competing parties are encouraged to provide genuine data.

III. System Architecture

The proposed system enables collaborative computing in privacy preserving fashion. The system supports taking inputs from multiple competing parties in distributed environment. The system checks the compatibility of inputs and takes care of incentive dynamics thereby encouraging the parties to provide genuine data. The architectural overview of the proposed system is as presented in Figure 1. Privacy preserving and analysis are the two important modules in the proposed work. In Privacy Preserving module, the registered users are assigned with unique keys by using random key generator. By using this we can ensure that the final analysis result can be only viewed by registered parties and the data given by parties are encrypted by using Data Encryption Standard (DES) algorithm thereby achieving privacy. The analysis part also works in parallel.

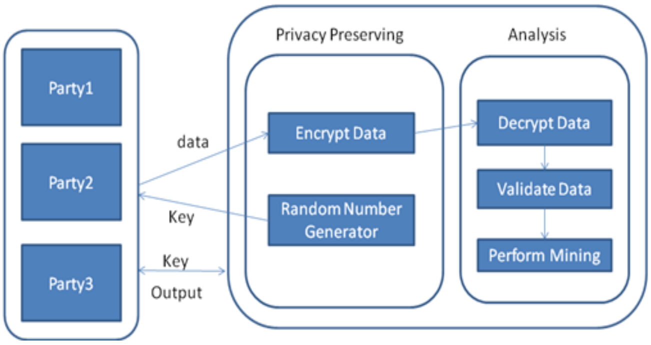

Figure 1: Architectural overview of the proposed system.

In the Analysis module, the encrypted data is decrypted in order to perform mining by using Data Encryption Standard (DES) algorithm. The decrypted data is then validated which takes care of incentive dynamics based on the validity of inputs. Once the inputs are validated then collaborative data mining takes place. We use apriori algorithm in order to perform collaborative data mining .We proposed an algorithm named Privacy Preserving Collaborative Data Mining (PPCDM) for successful realization of our framework. The privacy preserving data mining is achieved by using the algorithm presented in listing1.

```
Algorithm: Privacy Preserving Collaborative Data
                Mining (PPCDM)
Inputs     :  Data sets (d1, d2, d3….) from parties
Output     : Knowledge

From each party load data sets
 CASE 1: Horizontal Partition
Apply Privacy preserving
Do check  the Correctness
   IF Correctness is wrong THEN
      Reduce the incentives
   END IF
   IF Correctness is right THEN
       Collect the data based on common attribute
   IF  The data have all attributes Then
     Collaborate the data sets D=d1+d2+d3+…;
     Apply association rule mining to collaborative
      data D
END IF
ELSE
      Reduce the incentives
END
     send results to each party
   END IF

CASE 2:Vertical Partition
Apply Privacy preserving
Do check  the Correctness
   IF Correctness is wrong THEN
      Reduce the incentives
   END IF
   IF Correctness is right THEN
       Collect the data based on common attribute
IF  The data don't have same attributes Then
     Collaborate the data sets D=d1+d2+d3+…;
     Apply association rule mining to collaborative
     data
END IF
ELSE
      Reduce the incentives
END
     send results to each party
    END IF
```

Listing 1: Privacy preserving collaborative data mining algorithm.

As can be seen in Listing 1, it is evident that the algorithm supports both horizontal and vertical data for collaborative data mining besides preserving privacy. The proposed system also takes care of multi-party secure communications. Multiple parties are authenticated before allowing them to participate in privacy preserving collaborative data mining. Among the parties incentives are also provided to encourage the competing parties to provide genuine inputs. We built a prototype application that facilitates collaborative data mining.

The **proposed algorithm** is as following:

1) Registration of multiple parties, after registration each Party is assigned with unique keys by using Random Key Generator.
2) Each party's input data is encrypted by using DES algorithm.
3) Administrator logged in
4) Administrator decrypt's the received encrypted data by using DES algorithm and verifies the data and provides incentives.
5) Collaborate the data and perform Data Mining by using Apriori algorithm.
6) The results are viewed by only registered parties.

Multiple parties can collaborate and provide data for mining in distributed environment. Privacy preserving data mining is the important research area. In this paper we focused on building a data mining application that allows multiple parties to give compatible inputs in order to mine the data to produce business intelligence. Association rule mining is demonstrated with the application. However, it can be extended to support all mining operations. The collaborative data mining along with privacy preserving data analysis is possible with the help of the prototype application. The application has the notion of incentives. Incentives allow the competing parties to behave well. When they misbehave, the incentives will be decreased that will affect the credibility of that party. This is the motivation the competing parties get so as to provide genuine inputs to the system.

IV. Experimental Results

Experiments are made in terms of incentives and also the security aspects of the proposed systems. Incentives helped competing parties to give genuine inputs. Privacy preserving helped the system to be highly secure.

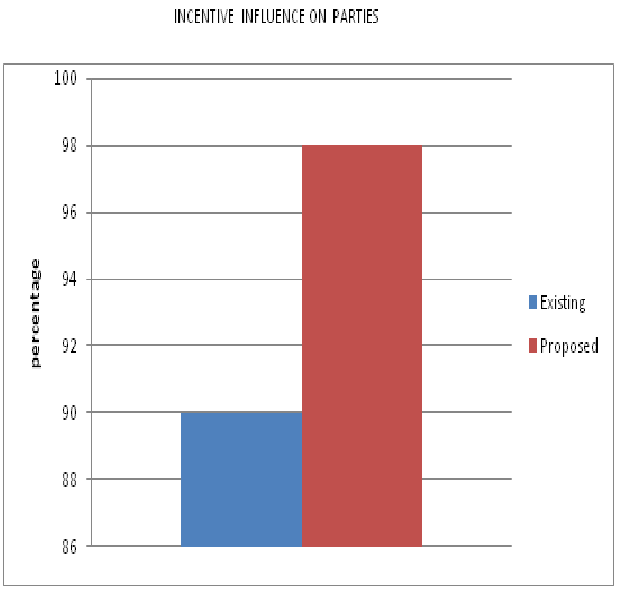

Figure 2: Comparison of the proposed system with existing system in terms of incentives.

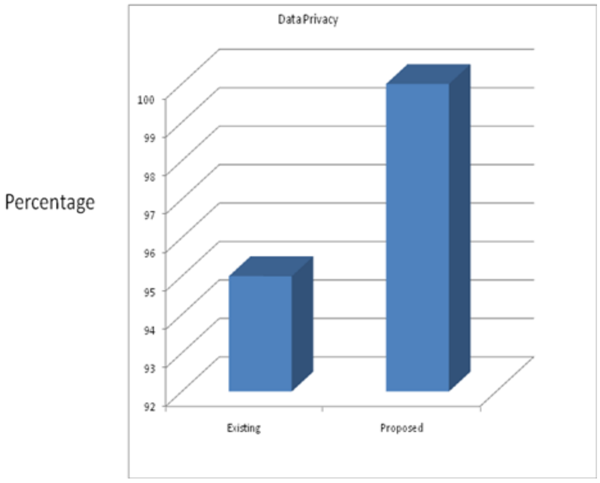

Figure 3: Comparison of the proposed system with existing system in terms of security.

As shown in Figure 2 and 3, it is evident that the proposed system has influenced the competing parties in order to provide genuine input data and also the data is highly secured.

V. Conclusion and Future Work

In this paper, we studied the problem of collaborative data analysis. When multiple competing parties are involved in providing data for privacy preserving data analysis, the parties might provide either genuine data or they provide incompatible data. In this paper we explored the possibilities to encourage all competing parties to provide genuine data so as to serve the purpose of collaborative data mining in order to acquire new business intelligence and device new business models. We presented secure and incentive compatible approach that lets competing parties to get incentives for genuine data. When incompatible data is suspected, the incentives are reduced. This is the motivation for the competing parties to behave well and provide genuine data. We built a prototype application that demonstrates the proof of concept. The empirical results revealed that the proposed solution is good for collaborative and privacy preserving data analysis in distributed environment. In future we extend this research into other aspects such as Big Data mining with Map Reduce programming using distributed programming frameworks like Hadoop.

VI. REFERENCES

[1] M.J. Atallah, M. Bykova, J. Li, and M.Karahan,"Private Collaborative Forecasting and Benchmarking," Proc. Second ACM Workshop Privacy in the Electronic Soc. (WPES), Oct. 2004.

[2] M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004.

[3] W. Du and Z. Zhan, "Building Decision Tree Classifier on Private Data," Proc. IEEE Int'l Conf. Data Mining Workshop Privacy, Security, and Data Mining, C.Clifton and V. Estivill-Castro, eds.,vol. 14, pp. 1-8, Dec. 2002.

[4] J. Vaidya and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02), pp. 639-644, July 2002.

[5] R. Agrawal and E. Terzi, "On Honesty in Sovereign Information Sharing," Proc. Int'l Conf. Advances in Database Technology, pp. 240-256, 2006.

[6] M. Kantarcioglu and R. Nix, "Incentive Compatible Distributed Data Mining," Proc. IEEE Int'l Conf. Soc. Computing/IEEE Int'l Conf. Privacy, Security, Risk and Trust, pp. 735-742, 2010.

[7] Murat Kantarcioglu and Wei Jiang, "Incentive Compatible Privacy-Preserving Data Analysis", ieee transactions on knowledge and data engineering, vol. 25, no. 6, june 2013.

[8] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," Proc.Int'l Conf. Advances in Cryptology (CRYPTO '00), pp. 36-54, Aug.2000.

Author Profile

1. Gottipamula Padmavathi ,pursuing M.Tech in Computer science and engineering branch from Aurora's Technological and Research Institute, uppal, Hyderabad, Telangana. Her Areas of interest are data mining.

2. T.V. Ramanamma working as a Sr.Asst.professor in the Department of Computer Science and Engineering in Aurora's Technological and Research Institute. She had received her M.Tech in Computer Science and Engineering. Her areas of interest are networking.