ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Novel approach to Incentive Compatible Secure Data Analysis

Sowdamini. K[1]
M.Tech Student, Dept.of CSE
Amrita Sai Institute of Science & Technology
Krishna Dt. A.P.

Chiranjeevi. P[2]
Asst.Prof, Dept.of Computer Applications
Amrita Sai Institute of Science & Technology
Krishna Dt. A.P.

Abstract: ***Data mining techniques enable the organizations gather beneficial knowledge in multiple data provider environment using Privacy Preserving Data Analysis (PPDA) without loss of privacy. Current PPDA techniques guarantee the confidentiality of the input data by revealing the final analysis results only. However, cannot guarantee the truthfulness of input data. With proper incentives the PPDA techniques may make the competing parties to provide truthful data for analysis.***

Without an efficient incentive strategy the present PPDA techniques may not reach the goal of motivating competing parties to provide truthful data. The theorems developed in this paper can be adopted to analyze certain important PPDA tasks that could be conducted in a way that providing the truth is the best choice for any participating party.

Keywords: Privacy; Privacy Preserving Data Analysis; Data mining; Non-cooperative Computation.

I. INTRODUCTION

Advancements in technology lead to the applications like online shopping to the next level offering various benefits. Online shopping is an application which provides a user interaction interface that provides more security for sensitive or personal data. Privacy preservation is a most important area in data mining.

Data mining is a process that explores the relationship in data sets to make valid predictions for data analysis tools. The main objective of data mining is to describe the data in a suitable form for analysis. It means that summarizes the statistical attributes of data, visually review by using charts and graphs and search for meaningful links in the data.

The main objective of this paper is to propose a strategy for data sharing that motivates the competing parties to provide truthful data for analysis. The real challenge in the multiple data provider environment is, to provide confidentiality of data. The ability to communicate and share data in secure environment has potential benefits and the use of a universal data source carries potential use for building accurate data analysis models. Certain PPDA techniques guarantee that nothing other than the final analysis results is revealed, but they cannot guarantee the truthfulness of the input data provided by the participating parties.

Recently emerged model, Secure Multi-Party Computation (SMC) proved to be successful in providing the secure data sharing for analysis without having to share or disclose the actual data. SMC model assumes that participating praties provide truthful data for analysis. SMC protocols cannot guarantee about the palpable of input data, unless proper incentives are set.

This paper analyses the types of distributed functionalities suitable for implementing in an incentive compatible fashion. The theorems adopted in this paper analyze whether or not input modification could occur for computing a distributed functionality.

II. LITERATURE SURVEY

In literature many privacy preserving data analysis protocols have been proposed using cryptographic techniques. In these techniques, data are assumed to be either vertically or horizontally partitioned. PPDA protocols are developed for both horizontally and vertically partitioned data to build decision trees, generating K-means clustering, KNN classifiers and mining association rules. All the exiting techniques assume that each participating party provides true data during the distributed data mining protocol execution.

Data mining includes various privacy preserving techniques such as data distribution, data mining algorithms,rule hiding and data modification, privacy preservation i.e. encryption methods. The PPDA mainly consider four categories of models such as Trust Third Party model, Semi-honest Model, Malicious Model and other models-Incentive Compatibility.

Murat et al [2], proposed procedures to mine distributed association rules on horizontally partitioned data and showed that distributed association rule mining can be done efficiently under reasonable security assumptions. Vaidya et al Presented a two party algorithm on vertically partitioned data, for efficiently discovering frequent itemsets with minimum support levels. Which are not revealing individual transaction values [1]. They proposed a privacy preserving association rule mining algorithm given a privacy preserving scalar product protocol and an efficient protocol for computing scalar product while preserving privacy of the individual data.

Mc Grew and Shoham included a game theoretic model in standard secure multiparty computation. They proposed a frame work of non-cooperative computation (NCC) by considering four components of each agent's utility function a) the wish to know the correct value of the function b) the wish to prevent others from knowing it c) the wish to prevent others from knowing one's own private input d) the wish to know other agent's private inputs. The NCC model can be seen as an example of applying game theoretical ideas in distributed computation setting [5].

III. SYSTEM ARCHITECTURE

Organizations who have private data may collaboratively conduct PPDA tasks to get useful data models or analysis results. For example, various credit card issuing banks may try to build models for credit card fraud detection through PPDA tasks. A PPDA model is said to be efficient when it has the ability to compute the desired constructive outcome of data sharing without having to actually share or disclose one's private data.

The system architecture is such that an unauthorized person is not allowed. In case, if any fraud user is trying to access the data security system will reject access for the particular user and the data is retrieved from the database according to the request given by the user.

Compared to conventional privacy preserving models, the incentive compatible privacy preserving models provide more security for individual private data. It does not require fraud signatures and yet is able to detect frauds by considering the data analysis model result. Another important advantage of incentive compatibility is a drastic reduction in the number of false positive transactions. The model tries to find any anomalies in the transaction based on the data analysis model

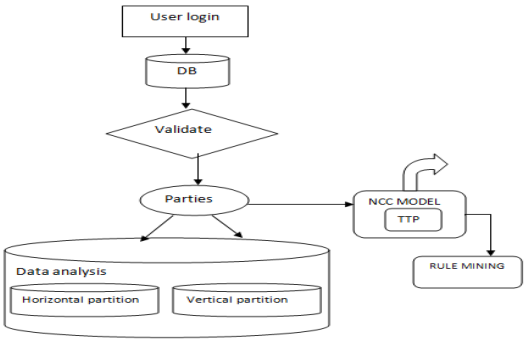

Fig. 1.System Architecture

In PPDA tasks multiple parties can participate. This is in contrast to SMC protocol, which is used in the existing system. Each participating party learns the output of some function f over the joint input of the parties in a PPDA model. All participating parties first send their private inputs securely to a trusted party (TTP) as shown in fig. 1, and then TTP computes f and sends back the result to every participating party. Here TTP will check either participants give truthful inputs or not. This paper analyses the incentive issues and display the list of the organizations in which the individual takes part. The techniques used are discussed below.

Non-cooperative Computation Mode to:

Each party participating in a NCC model protocol is to learn the output of some given function f over the joint inputs of the parties. The first step in NCC model is, all participating parties send their private data securely to a trusted third party (TTP) then TTP computes f and send back the result to every participating party.

The following assumptions are made in NCC model:

Correctness: The first preference for every participating party is to learn the correct result.

Exclusiveness: Every participating party if possible, choose to learn the correct result exclusively.

Under these assumptions the NCC model is defined as follows: Given a set of n parties, for party i, a) each party i sends V□ to a TTP b) The TTP computes function f on all parties and send results to the participating parties. $f(v') = f(v_1', v_2', ...v_n')$ c) Each party i compute f(v) based on f(v') received from TTP. Considering the given protocol does not limit its generality.

In SMC model, the TTP can be replaced such that the required functionality is still computable without violating privacy regarding each participating party. The participating parties are expected to provide their genuine inputs to correctly evaluate a function that satisfies the NCC model.

Any functionality that satisfies the NCC model inherently incentive compatible under the assumption that participating parties prefer to learn the function results correctly. NCC model considers two types of models one is Deterministic Non cooperative computation (D-NCC) and another Probabilistic Non cooperative Computation (P-NCC).

The D-NCC model, the function f is in the DNCC because the proof needs to consider all possible g_i and t_i pairs. The strategy t_i defines a way to change the input and the strategy g_i defines a method to reconstruct the actual result based on the genuine input, modified input and the result computed based on the modified input and other parties input data

Secure Code Computation Process:

It is providing incentive compatible secret code question for the NCC model. The theorem consists of the following steps

Step1: From bank database, select two fields from customer details as input for secure code computation process

Step2: Keep one field constant (for ex: say username) and other fields consist of information like occupation, account number, mailed, etc...

Step3: First field data is partitioned vertically and then attach a second field in the middle of partition data using secure sum process technique.

Association Mining Rule:

Given a set of n parties and let D be the domain of possible inputs, $v \in D^n$. f is defined as $f : \mathrm{D}^n \mapsto R$ with range R then f is deterministically Non-cooperative computable(DNCC) ,if the following holds For any party i, every strategy (t_i, g_i) and every $v_i \in D_i$, it is the case that

» Either $\exists v_{-i} \in D_{-i}, g_i(f(\mathrm{t}_i(v_i), v_{-i}) \neq f(v_i, v_{-i})$

» Or $\forall v_{-i} \in D_{-i}, f(\mathrm{t}_i(v_i), v_{-i}) = f(v_i, v_{-i}).$

The rule simply states what function could be computed in NCC setting deterministically and no party could correctly compute the correct result once the party lies about his or her inputs in a way that changes the original function result.

IV. CONCLUSION

The paper proposed an incentive compatible privacy preserving data analysis technique which motivates the participating parties to provide truthful inputs.NCC model provides secure data sharing during the information transaction and further reduces the number of false positive transactions. It tries to find any fraud in the transaction based on the data analysis model.

References

1. 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.
2. Murat Kantarcioglu and Wei Jiang, "Incentive Compatible Privacy Preserving Data Analysis", IEEE Transactions on Knowledge and Data Engineering, Vol.25, No.6, June 2013.
3. M Kantracioglu and O.Kardes, "Privacy Preserving Data Mining in the Malicious Model," Int'l Journal of Information and Computer Security, Vol.2, pp.353-375, Jan.2009.
4. Y.Shoham and M.Tenneholtz, "Non-Cooperative Compution:Boolean Functions with Correctness and Exclusivity," Theoretical Computer Science,Vol.343,Nos.1/2,pp.97-113,2005
5. R.Mc Grew, R Porter and Y.Shoham, "Towards a General Theory of Non Cooperative Computation," Proc.conf.Theoretical Aspects of Rationality and Knowledge (TARK IX), 2003.
6. S.Han and W K Ng , "Preemptive Measures against Malicious Party in Privacy-Preserving Data Mining," Proc.SIAM Int'l Conf. Data Mining(SDM),PP.375-386,2008.
7. G.Jagannadhan and R.N.Wright, "Privacy Preserving Distributed K-Means Clustering over Arbitrarily Partitioned Data," Proc.ACMSIGKDD Int'l Conf.Knowledge Discovery and Data Mining, pp.593-599, Aug.2005.
8. Dhanalakshmi.M ,Siva Sankari.E, "Efficient Incentive Compatible Model for Secure Data Sharing," Int'l Journal of computer science and mobile applications,Vol.2,Issue.1,Jan 2014,pp.170-179.
9. D.S.Deepika ,R.Nandini, "Secure Sharing of Private Data using Privacy Preserving Data Analysis,"Int'l Journal of Adavanced Research in computer Science and information technology,Vol.2,Issue.1,2014.
10. H.Kargupta,K.Das and K.Liu, "A Game Theoretic Approach toward Multi-Party Privacy Preserving Distributed Data Mining," Proc.11th European Conf.Principles and Practice of Knowledge Discovery in Databases,pp.523-531,Sept.2007.

AUTHOR(S) PROFILE

Smt.K. Sowdamini, Pursuing M.Tech (CSE) from AMRITA SAI INSTITUTE OFSCIENCE ANDTECHNOLOGY, Paritala, Krishna Dist, A.P. and her areas of interest are Cloud Computing Data Mining and Network Security.

Mr. P.Chirajeevi, an efficient teacher, received M.Tech (C.S.E) from JNTU Kakinada is working as an Assistant Professor in Department of Computer Applications of AMRITA SAI INSTITUTE OF SCIENCE AND TECHNOLOGY Having 6+ years of teaching experience areas of interest are Cloud Computing Data Mining and Network.

