International Journal of scientific research and management (IJSRM)
\||Volume||3||Issue||6||Pages|| 3008-3010||2015||
Website: www.ijsrm.in ISSN (e): 2321-3418

Making Data Analysis into Incentive Compatibility Mode Using DNCC Approach

P Vamsi Naidu[1], K Lavanya[2]

[1]PG Scholar, LBRCE, Mylavaram,
Krishna District, Andhra Pradesh, India
Pvn1232@gmail.com

[2]Senior Grade Assistant Professor, Department Of IT,
LBRCE, Mylavaram, Krishna District, Andhra Pradesh, India
lavanya.kk2005@gmail.com

Abstract: *Generally the organizations wanted to conduct data analysis on their private data in order to getting better predictions for their benefits .Even though we have some Privacy Preserving Data Analysis (PPDA) techniques that are not sure about the participating parties are true about their input or not. This leads one organization may confidentially breach the other organizations data and misuses it. This misuse leads severe harm to the organization which lost its information. This point raises a question of how to design PPDA techniques that motivates participating parties to provide truthful inputs. In this paper we clearly see the current approaches that are being in use and proposing a new methodology called DNCC and its way working and we will see how this approach makes the data analysis into incentive compatible*

Keywords: Privacy, Security, Secure Multi-party Computation (SMC), Non-Cooperative computation (NCC), Deterministic Non-Cooperative computation (DNCC)**..**

1. Introduction

In the area of privacy-preserving data mining, a differentially private mechanism intuitively encourages People to share their data because they are at little risk of revealing their own information. Privacy and security, particularly maintaining confidentiality of data, have become a challenging issue with advances in information and communication technology. The ability to communicate and share data has many benefits, and the idea of an omniscient data source carries great value to research and building accurate data analysis models. For example, for credit card companies to build more comprehensive and accurate fraud detection system, credit card transaction data from various companies may be needed to generate better data analysis models. [1] Secure multi-party computation (SMC) [2], [3], [4] has recently emerged as an answer to this problem. Informally, if a protocol meets the SMC definitions, the participating parties learn only the final result and whatever can be inferred from the final result and their own inputs. A simple example is Yao's millionaire problem [4]: two millionaires, Alice and Bob, want to learn who is richer without disclosing their actual wealth to each other. Recognizing this, the research community has developed many SMC protocols, for applications as diverse as forecasting [5], decision tree analysis [6] and auctions [7] among others.

Example: Let $\mathbf{x_i}$ be the $\mathbf{i^{th}}$ company's sales amount. In order to estimate the sample mean, companies need to calculate $\boldsymbol{\mu = 1/n \sum_{i=1}^{n} x_i}$ and similarly variance $\mathbf{s^2 = 1/n-1 \sum_{i=1}^{n} (x_i - \mu)^2}$ Any company may **exclusively** learn the **correct** result by lying about its input. Company i may report $\mathbf{x_i`}$ instead of the correct $\mathbf{x_i}$. Given the wrong mean $\boldsymbol{\mu`}$ and variance $\mathbf{s`^2}$ (computed based on $\mathbf{x_i`}$ and truthful values from the other parties), the company **i** can calculate the **correct** sample mean $\boldsymbol{\mu}$ by setting $\boldsymbol{\mu = \mu` + x_i - x_i`/n}$.

The correct sample variance s^2 can be calculated as $\mathbf{S^2 = s`^2 + (x_i^2 - x_i`^2/n-1) + (n(\mu`^2 - \mu^2))/n-1.}$

Let us consider three organizations (e.g., n =3) are formed as a group and wanted to predict future results on their "Total Sales" using SMC model. The following figure will clearly explains Working of SMC

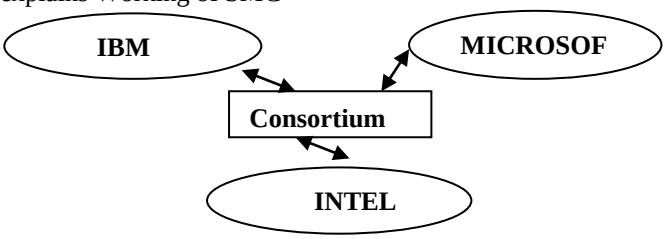

Figure 1: Explanation of SMC Model

No of organizations **n=3**,
Total Sales (IBM) $\mathbf{x_1 = 3000}$,
Total Sales (Microsoft) $\mathbf{x_2 = 4000}$,
Total sales (Intel) $\mathbf{x_3 = 5000}$,
Then compute mean using above mentioned formula we get **mean μ=4000** and **variance=1000000** but 'Organization 1(IBM)'lied about its input and provided $\mathbf{x`_1 = 6000}$ instead **3000** and gathers others inputs and misuses it
If the above situation was happened then no company would have the incentive to be truthful

2. Literature Survey

In this paper, we analyze what types of distributed functionalities could be implemented in an incentive compatible fashion. In other words, we explore which functionalities can be implemented in a way that participating parties have the incentive to provide their true private inputs upon engaging in the corresponding SMC protocols. We show how tools from theoretical computer science in general and

non-cooperative computation [8] in particular could be used to analyze incentive issues in distributed data analysis framework. This is significant because input modification cannot be prevented before the execution of any SMC-based protocol. (Input modification could be prevented during the execution of some SMC-based protocols, but these protocols are generally very expensive and impractical.) [9].

The theorems developed in the paper can be adopted to analyze whether or not input modification could occur for computing a distributed functionality. If the answer is positive, then there is no need to design complicated and generally inefficient SMC-based protocols.

Following are the terms used in the paper.

NCC: Non-Cooperative Computation

DNCC: Deterministic Non-Cooperative Computation

PPDA: Privacy Preserving (Distributed) Data Analysis

SMC: Secure Multi-party Computation

TTP: Trusted Third Party

In this paper, we assume that the number of malicious or dishonest participating parties can be at most n − 1, where n is the number of parties. This assumption is very general since most existing works in the area of privacy preserving data analysis assume either all participating parties are honest (or semi-honest) or the majority of participating parties are honest. Thus, we extend the non cooperative computation definitions to incorporate cases where there are multiple dishonest parties. In addition, we show that from incentive compatibility point of view, most data analysis tasks need to be analyzed only for two party cases. Furthermore, to show the applicability of our developed theorems, we use these theorems to analyze under what conditions, common data analysis tasks, such as mean and covariance matrix estimation can be executed in an incentive compatible manner

3. Related Work

Even though privacy-preserving data analysis techniques guarantee that nothing other than the final result is disclosed, whether or not participating parties provide truthful input data cannot be verified. Although certain PPDA techniques guarantee that nothing other than the final analysis result is revealed, it is impossible to verify whether or not participating parties are truthful about their private input data. In other words, unless proper incentives are set, even current PPDA techniques cannot prevent participating parties from modifying their private inputs.

3.1 Privacy-Preserving Data Analysis

All the previous privacy preserving data analysis protocols assume that participating parties are truthful about their private input data. Recently, game theoretical techniques have been used to force parties to submit their true inputs [2]. The techniques developed in [2] assume that each party has an internal device that can verify whether they are telling the truth or not. In our work, we do not assume the existence of such a device. Instead, we try to make sure that providing the true input is the best choice for a participating party.

The following figure denotes the architecture of the privacy preserving system model. The architecture of a privacy preserving system gives the detailed explanation about the process of the security system in which it allows only the authorized person not others. Suppose, if any fraud user is trying to access the data security system will not allow the user and also the access will be denied for the particular user. Then the appropriate data are retrieved from the database according to the request given by the user

Figure 2: Secure User Interaction with System

3.2Non-Cooperative Computation

Recently, research issues at the intersection of computer science and game theory have been studied extensively. Among those research issues, algorithmic mechanism design and non-cooperative computation are closely related to our work. The field of algorithmic mechanism design tries to explore how private preferences of many parties could be combined to find a global and socially optimal solution [10]. Usually in algorithmic mechanism design, there exists a function that needs to be maximized based on the private inputs of the parties, and the goal is to devise mechanisms and payment schemes that force individuals to tell their true private values. In our case, since it is hard to measure the monetary value of the data analysis results, devising a payment scheme that is required by many mechanism design models is not viable (e.g., Vickers-Groves-Clarke mechanisms [9]). Instead, we adopt the non-cooperative computation model [11] that is designed for parties who want to jointly compute the correct function results on their private inputs. Since data analysis algorithms can be seen as a special case, modifying non-cooperative computation model for our purposes is a natural choice [12].

4. Proposed Work (DNCC)

In design incentive compatible privacy-preserving data analysis techniques that motivate participating parties to provide truthful input data. The Deterministic non-cooperative computation (DNCC) model is the each party participates in a protocol to learn the output of some, given a function f over the joint data inputs of the parties. In the first step, all participating parties send their private inputs data to a trusted third party (TTP) for secure sharing. And then the second step is TTP computes the function f and sends back the result to every participating party. The Deterministic Non Cooperative Incentive Compatible model makes the following assumptions:

Correctness: **every participating party is to learn the *correct* result;**

Exclusiveness: **every participating party prefers to learn the *correct* result *exclusively*.**

Under the ***correctness*** and ***exclusiveness*** assumptions, the Non Cooperative Incentive Compatible model is formally defined as follows:

Given a set of n parties, for a party i,

1. Each party i send v_i' (not necessarily the correct private input) to a TTP.

2. The TTP computes $f(v') = f(v'_1,\ldots,v'_n)$ and sends the results back to the participating parties.

3. Each party i compute f (v) based on f (v′) received from TTP and v_i.

Considering the above protocol does not limit its generality. The incentive compatible Privacy Model function f over the joint inputs of the parties specified, that is derived from the secure code computation process.

In this paper, we first develop key theorems, then base on these theorem, we analyze what types of privacy-preserving data analysis tasks could be conducted in a way that telling the truth is the best choice for any participating party.

The following fig will clearly explains the working of DNCC model where user can send "two inputs" (ex: two security questions) to the TTP then the TTP computes "Security code" and sends back to the organizations. By using security code the data analysis would be done in healthy manner.

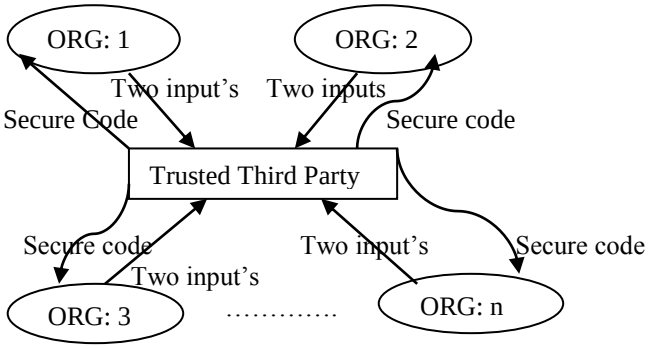

Figure 3: Practical Approach of DNCC

5. Working Modules

5.1 Privacy-Preserving Data Analysis:

In this module we assume that providing true input is the best choice

5.2 Non-Cooperative Computation:

All participating parties send their inputs to TTP ,then TTP computes the function 'f' and sends back the result

5.3 Analyzing Data Analysis Tasks in the NCC Model:

Combining the two concepts DNCC and SMC, we can analyze privacy preserving data analysis tasks that are incentive compatible.

5.4 Privacy Preserving Association Rule Mining:

The association rule mining and analyze whether the association rule mining can be done in an incentive compatible manner over horizontally and vertically partitioned databases.

6. Conclusion

The PPDA tasks analyzed in the paper can be reduced to evaluation of a single function. Now, the question is how to analyze whether a PPDA task is in DNCC if it is reduced to a set of functions. In other words, is the composition of a set of DNCC functions still in DNCC? We will formally answer this question in the future. Another important direction that we would like to pursue is to create more efficient SMC techniques tailored towards implementing the data analysis tasks that are in DNCC.

References

[1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In VLDB '94, pages 487–499,Santiago, Chile, September 12-15 1994. VLDB.

[2] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game a completeness theorem for protocols with honest majority. In 19th ACM Symposium on the Theory of Computing, pages 218–229, 1987.

[3] Andrew C. Yao Protocols for secure computation. In Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science, pages 160–164. IEEE, 1982.

[4] Andrew C. Yao How to generate and exchange secrets. In Proceedings of the 27th IEEE Symposium on Foundations of Computer Science, pages 162–167. IEEE, 1986.

[5] Mikhail J. Atallah, Marina Bykova, Jiangtao Li, & Mercan Karahan. Private collaborative forecasting and benchmarking. In Proc. 2d. ACM Workshop on Privacy in the Electronic Society (WPES), Washington, DC, October 28 2004.

[6] Yehuda Lindell and Benny Pinkas. Privacy preserving data mining. Journal of Cryptology, 15(3):177–206, 2002.

[7] Moni Naor, Benny Pinkas and R. Sumner. Privacy preserving auctions and mechanism design. In Proceedings of the 1st ACM Conference on Electronic Commerce. ACM Press, 1999.

[8] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin, and Yannis Theodoridis. State-of-the-art in privacy preserving data mining.SIGMOD Rec., 33(1):50–57, 2004.

[9] Noam Nisan and Amir Ronen. Algorithmic mechanism design (extended abstract). In STOC' 99, pages 129–140, New York, NY, USA, 1999. ACM Press.

[10] Yoav Shoham and Moshe Tennenholtz. Non-cooperative computation:boolean functions with correctness and exclusivity. heor. Comput. Sci., 343(1-2):97–113, 2005.

[11] Murat Kantarcıo˘glu and Chris Clifton. Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE TKDE, 16(9):1026–1037, September 2004.

[12] Yehuda Lindell and Benny Pinkas. Privacy preserving data mining. In Advances in Cryptology – CRYPTO 2000, pages 36–54. Springer-Verlag, August 20-24 2000.

[13] Robert McGrew, Ryan Porter, and Yoav Shoham. Towards a general theory of non-cooperative computation (extended abstract). In TARK IX, 2003.

[14] Jaideep Vaidya and Chris Clifton. Privacy preserving association rule mining in vertically partitioned data

Author Profile

Mr. P Vamsi Naidu currently pursuing M.Tech degree in Software Engineering from Laki Reddy Bali Reddy College of Engineering (LBRCE)

Mrs. K Lavanya working as a Senior Grade Assistant Professor in Department of IT, LBRCE. She has sanctioned a project titled "Design and Development of Robust Biometric tool for ATM Security" by DST under Young Women Scientist Scheme.