International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.02,Issue.19,
December-2013,
Pages:2218-2222

www.semargroups.org,
www.ijsetr.com

Secure Multi-Party Computation Incentive in Data Analysis

V. SRINIVAS[1], ARSHIA BANOO[2]

[1]Assoc Prof, Dept of CSE, Aurora's Scientific, Technological and Research Academy, AP-INDIA.
[2]PG Scholar, Dept of CSE, Aurora's Scientific, Technological and Research Academy, AP-INDIA,
Email: akarshkhan@gmail.com.

Abstract: In many cases, competing parties who have private data may collaboratively conduct privacy-preserving distributed data analysis (PPDA) tasks to learn beneficial data models or analysis results are often the most have different incentives competing parties. Although certain PPDA techniques guarantee that nothing other than the final analysis result are revealed, and it is impossible to verify whether participating parties are truthful about their private input data. To the proper, current PPDA techniques cannot prevent incentives are participating in their private inputs parties and modifying. This raises the question of how to design incentive compatible privacy-preserving data analysis techniques that motivate participating parties to provide truthful inputs. First we develop in this paper key theorem, and then based on these theorems; we can analyze certain important privacy preserving data analysis tasks that could be conducted in a way that telling the truth is the best choice for any participating party.

Keywords: The Privacy Secure Multi-Party Computation, Non-Cooperative Computation.

I. INTRODUCTION

Privacy and security, particularly maintaining data, have confidentiality become a challenging issue with advances in information and communication technology to communicate and share data, and the idea have many benefits an omniscient data source carries great value to research and building accurate data analysis models, for credit card companies to build more comprehensive and accurate fraud, credit card transaction data detection system from various companies may be needed to generate better data analysis models. Of Energy supports department research on building much more efficient diesel engines [8]. It has an ambitious task requires the collaboration of geographically and universities are distributed industries, national laboratories. At those institutions (including potentially competing industry partners) need to share their private data for building data analysis models to understand the underlying physical phenomena. An omniscient data source eases misuse, such as the identity theft of growing problem. To prevent misuse of data, there is a recent surge in laws mandating protection of confidential data, such as the privacy European Community standards [10], U.S. health-care laws. In this protection however it comes with a real cost through both added security expenditure and penalties and costs are associated with disclosure. Are we need this ability to compute the desired "beneficial outcome" of data sharing for analyzing without having to actually share or disclose data should be provided by maintain the security separation of control while still obtaining the benefits of a global data source. The Secure multi-party computation (SMC) [12], has recently to this problem emerged as answer. Informally, if a protocol meets the participating parties the SMC definitions, learn only the final result and whatever can be inferred from their own inputs the final result.

The simple example is Yao's millionaire problem: want to learn two millionaires, Alice and Bob who is richer without disclosing their each other actual wealth. To recognizing this, the research community has developed many applications SMC protocols, as diverse as forecasting [6], decision tree analysis and auctions among others. Nevertheless, the SMC model does not guarantee that data provided by parties are truthful participating. In many real life situations, data needed for building data analysis models are distributed among multiple parties with potentially conflicting interests. In this instance, a credit card company that has a superior data analysis model for fighting credit card fraud may increase its profits as compared to its engine design company may want to exclusively learn the data analysis models that may enable it to build much more efficient diesel engines are clearly, as described above, building data analysis models is generally performed among parties that have conflicting interests. In SMC, we generally assume that participating parties provide truthful inputs. In this assumption is normally justified by the fact that learning the correct data analysis models or results is in the best interest of all participating parties. From SMC-based protocols require

Copyright @ 2013 SEMAR GROUPS TECHNICAL SOCIETY. All rights reserved.

participating parties to perform expensive computations, does not if any party want to learn data models and, the party analysis results should not participate in the protocol assumption does not guarantee the truthfulness of the private input data when participating parties want to learn the final result exclusively.

The example we consider here is, a drug company may lie about its private data so that it can exclusively learn the data analysis model. For the SMC protocols guarantee that nothing other than the final data analysis results are revealed, it is impossible to verify whether or not participating their private input data parties are truthful about an unless proper, current SMC techniques incentives are set cannot prevent input modification by participating parties. In this problem to better illustrate, and we can consider a case from management where competing companies (e.g., Texas Instruments, IBM and Intel) establish a consortium company's send the consortium their sales data, and key manufacturing costs and times of the consortium analyzes the data and statistically summarizes them in a report of which industry trends is made available back to consortium members is in the interest of companies to learn true industry trends while revealing their private data as little as possible though SMC protocols can prevent the revelation of, they private data do not guarantee that companies send their true sales data and other required information can be assume that n companies would like to learn the sample mean and variance of the sales data for a particular type of product.

Example 1: Let x_i be the i^{th} company's sales amount in order to estimate the sample mean, companies need to calculate $\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ and similarly $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$ for sample variance. If any company may exclusively learn the correct result by lying about its input i may report x'_i instead of the correct x_i. Given the wrong mean μ' and variance s'^2 (computed based on x'_i and truthful values from the other parties), the company i can calculate the correct sample mean μ by setting:

$$\mu = \mu' + \frac{x_i - x'_i}{n} \tag{1}$$

If the correct sample variance s^2 can be calculated as:

$$s^2 = s'^2 + \frac{x_i^2 - x_i'^2}{n-1} + \frac{n\left(\mu'^2 - \mu^2\right)}{n-1} \tag{2}$$

As illustrated above, any company may have the incentive to lie about its input in order to learn the exclusively result, and at the same time, the correct result (e.g., μ) can be computed from, modified input and its original input the incorrect final result (e.g., x_i, x'_i and μ'). In this situation can be occurred, always no company would have the incentive to be truthful may be intrinsic nature of a function determines whether the situation (demonstrated by the above example) could occur.

A. Our Contributions

In this paper, we can analyze what types of distributed functionalities could be implemented in an incentive compatible fashion are 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 are show how tools from theoretical computer science in general and non-cooperative computation in particular could be used to analyze incentive issues in distributed data analysis framework is significant because input modification cannot be prevented before the execution of SMC-based any protocol. (Input modification could be prevented during the execution of SMC-based some protocols, but these protocols are generally expensive.) In this theorems developed in the paper can be adopted to analyze whether or not input modification could occur for computing a distributed functionality is positive, then there is no need to design complicated and generally inefficient SMC based protocols.

TABLE I: NOTATIONS AND TERMINOLOGIES

NCC	Non-Cooperative Computation
DNCC	Deterministic NCC
PPDA	Privacy Preserving 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. In 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. If, we can extend the non-cooperative computation definitions to incorporate cases where there are multiple dishonest parties are we show that from incentive of view, compatibility point most data analysis tasks need to be analyzed only for two party cases to show the applicability, we use of our developed theorems these theorems to analyze under the conditions, common data analysis tasks, such as mean and covariance matrix estimation; can be executed in an incentive compatible manner. The paper is organized as follows: Section II provides Related Work. In Section III, we propose Game Theoretic Background, Section IV concludes the paper with possible future research directions.

II. RELATED WORK

Cryptography and game theory have common, great deal in terms of the goals they try to achieve this problems

tackled by cryptography generally seek to assure that participants in certain activates are forbidden to deviate (profitably) from the prescribed protocol by rendering such actions or computationally infeasible detectable, impossible. It is similarly, mechanism design to seeks, but it does so by rendering the deviations unprofitable and it is understandable that a fair amount of work has been done to use the techniques of one to solve the problems of the other work is related to not directly ours, since a fair amount of the game theoretic security work deals with an individual specific functions, and the steps of the computations of those functions. If we define the class of NCC, or non-cooperatively computable functions, and define specifically the NCC boolean functions which are addition, the paper defined two additional classes, which stand for p-NCC and s-NCC, probabilistic-NCC and subsidized-NCC, respectively. If the P-NCC is the functions which are computable with some probability non-cooperatively and s-NCC is the functions which are computable when external monetary motivation is allowed. For this expanded to consider different motivations, and coalitions. Here our work does involve making functions computable in a competitive setting; it involves more specifies mechanisms and complicated functions, to ensure computability.

In addition to this, much work seeks to include a game theoretic model in standard secure multi-party computation are considering players which are honest, semi-honest, or malicious, these works simply consider players to be rational, in the game theoretic sense of this work concentrates on the problem of dividing is secret sharing that secret number among players such that any quorum (sufficiently large subset) of them can reconstruct the secret number is first studied by later re-examined by each other protocols for this problem were outlined the paper by hybridizes, within the two areas realm, by considering secret sharing some players honest and a majority of players rational. In this work seeks a broader realm of computation, and which build their computation model on a secret sharing model that attempts to combine game theoretic and cryptographic methodologies, many of which are surveyed. If many of these rational secure computation systems could be used to ensure privacy in our mechanism like other secure, they make computation systems no guarantees about the truthfulness of the inputs. More closely related to this paper, several works have attempted to enforce honest behavior among the participants in a data sharing protocol on this paper builds on the work of who present a model which enforces honesty in data sharing through use of auditing mechanisms. Are it can presents strategies which enforce honesty in a distributed computation, without relying on a mediator integrate the auditing mechanism, to convert with secure computation existing protocols into rationally secure mechanism protocols are-based framework for regression learning using risk minimization is says nothing, and solely focuses on regression learning to this work is analyzes each step of a multi-party computation process in terms of focus preventing cheating within the process, and removing coalitions from each game play of these deals with the problem of ensuring truthfulness in data mining each one requires the ability to verify the data after our mechanisms calculation have no such requirement.

There is one work, by which does not make use of an auditing mechanism to encourage truthfulness this work does not actually encourage truthful sharing by all parties game theoretic strategies proposed for a non-malicious player actually encourage the player data, although not completely, in the face of a malicious adversary strategy results are reduced accuracy, interestingly enough, but greater privacy malicious adversary strategy presented has no incentive to change this input does not consider parties to be malicious or otherwise. Our work only assumes parties are rational focus on data integration rather than data mining value has been applied to many things, from fair division to power cost allocation, but has not been applied in this way to data sharing.

III. GAME THEORETIC BACKGROUND

Game theory is the study of competitive behavior among multiple parties is to contain a game four basic elements: players, actions, payoffs, and information. And it has players actions which they can perform at designated times as a result of the actions in the game, players receive payoffs. In this players have different pieces, on which the information payoffs, and may depend it is the responsibility of the player to use a profitable strategy to increase his or her payout a player who acts in such a way as to maximize his or her payout is termed rational take games many forms, and vary in the four attributes mentioned above, but all games deal with them specific game we describe in this paper is a single round, incomplete information game, finite player, with payouts based on the final result of players' simultaneous actions. We can proceed with a discussion of mechanism design; it is convenient to define a common notation used within the literature and within this paper.

A. Mechanism Design for Non-Cooperative Games

Mechanism design is a sub-field and deals with the construction of games for the purpose of some achieving goal, when players act rationally. A mechanism is defined, for our purposes1, as:

Definition 1: Given a set of n players, and a set of outcomes, A, let V_i be the set of possible valuation functions of the form v_i (a) which player i could have for an outcome $a \in A$. We then define a mechanism as a function f: $V_1 \times V_2 \times ... \times V_n \rightarrow A$, which given the valuations claimed by the players, selects an outcome, and n payment functions, $p_1, p_2, ..., p_n$, where p_i: $V_1 \times V_2 \times ... \times$

$V_n \rightarrow \Re$ that is, given the valuations claimed by the players, selects an amount for player i to pay. Thus, the overall payout to a player in this mechanism is his valuation on the outcome, v_i (a), minus the amount he is required to pay, p_i (v_i, v−i). A mechanism is said to be incentive compatible if rational players would prefer to give the true valuation rather than any false valuation is the more formally:

Definition 2: If, for every player i, every $v_1 \in V_1$, $v_2 \in V_2$,..., $v_n \in V_n$, and every $v'_i \in V_i$, where a = f(v_i, v−i) and a′ = f(v'_i, v−i), then v_i(a)−p_i(v_i, v−i) ≥ v_i(a′)−p_i(v'_i , v−i), then the mechanism in question is incentive compatible. Thus, a player would prefer to reveal his true valuation rather, we are assuming all other valuation than the other players are truthful. Other important terms of an intuitively whether a individual rationality, which is player would desire to participate in a game in the first place in the utility a player receives in the event that they choose not to participate is called the reservation utility is the order for strategy to be considered equilibrium, for all players, it must be individually rational and incentive compatible. The specific mechanism used in our data mining is the Vickrey-Clarke-Groves (VCG) mechanism, seeks to maximize the social welfare of all participants in a game is the social welfare can be defined as the sum of the valuations of all VCG players wishes to cause rational players to act in such a way that the sum of the valuations each player has of the maximized outcome is mathematical notation, this is where the outcome chosen is argmax $_{a \in A} \sum_i v_i(a)$, where *A* is the set of possible actions, and v_i is the valuation function for player *i*. The VCG mechanism is defined as follows:

Definition 3: A mechanism, consisting of payment functions are p_1, p_2, ..., p_n and a function f, for a game with outcome set A, is a Vickrey-Clarke-Groves mechanism if

$$f(v_1, v_2, ..., v_n) = \text{argmax}_{a \in A} \sum_i v_i(a) \quad (3)$$

(f maximizes the social welfare) and for some functions are h_1, h_2, ..., h_n, where h_i: $V_{-i1} \rightarrow \Re$(h_i does not depend on v_i), for all (v_1, v_2, ..., v_n) ∈ V, p_i(v_1, v_2, ..., v_n) = h(v−i) − $\sum_{j \neq i}$ =v_j (f(v_1, v_2, ..., v_n)). Since p_i is the paid amount by player i, this ensures that each player is paid an amount equal to the valuation of all other players that means each player would have incentive to make actions to maximize the social welfare formal proof that the VCG mechanism is incentive compatible can be found.

B. Cooperative Game Theory

Cooperative games, first formalized by von Neumann and Morgenstern use a different setup than the standard non-cooperative game scenarios consist of a set of players N (usually called the grand coalition) and a valuation function v which maps subsets of N to the amount the subset of players can gain by cooperating, with v($\emptyset$) = 0. A non-cooperative game can be translated into the cooperative scenarios are, assuming in a few ways that coalitions can enforce coordinated behavior is the most common methods are to associate with each coalition the max-min or min-max sum of the gains its members can guarantee by cooperating. One important mechanism designed for use in cooperative games is the Shapley value, which is defined for each player *i* as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)) \quad (4)$$

This function can also be defined as:

$$\phi_i = \frac{1}{|N|!} \sum_R v(P_i^R \cup \{i\}) - v(P_i^R) \quad (5)$$

Where *R* is taken over the possible orderings of *N*, and P^R_i is defined as the elements of *R* which precede *i* in *R*. Informally, this value is formed by taking the contribution brought to the coalition by the player at each possible time the player could have been added to the coalition. This overall sum gives a "fair" value for the player's contribution to the grand coalition. The Shapley value is considered will choose individually rational, that is, players to join the coalition if offered their Shapley value, if the game is super additive. In a super additive game, for any disjoint coalitions S, T ⊆ N, we have:

$$V(S \cup T) \geq v(S) + v(T) \quad (6)$$

The Shapley value is defined, for other games, but not necessarily individually rational.

IV. CONCLUSION AND FUTURE WORK

Even though privacy-preserving data analysis techniques guarantee that nothing other than the final result are disclosed, whether or not participating parties provide truthful input data in this paper cannot be verified, we have investigated what kinds of PPDA tasks is incentive compatible under the NCC model can be based on the findings, there are several important PPDA tasks that are incentive driven classifies the common data analysis tasks studied in this paper into DNCC or Non-DNCC categories. Is most often, data partition schemes can making difference in determining DNCC or Non-DNCC classifications? And any functions, are providing a general way to determine if a function is in DNCC. In addition, Claim 5.1 can be used to analyze if a composite function is in DNCC, and it also provides a method to design a PPDA protocol that guarantees to be incentive compatible under the DNCC definition. For instance, a PPDA task can have many variations, and one common variation is to place a filter at the last step of the task to make the PPDA

protocols more secure (e.g., secure similar document detection vs. its threshold based variation). According to Claim 5.1, as long as the last step in a PPDA task is in DNCC, it is always possible to make the entire PPDA task satisfying the DNCC model. Therefore, when designing a PPDA protocol, it is in our best interests to make the last step of the PPDA task incentive-compatible whenever possible. As a part of future research direction, we will investigate incentive issues in other data analysis tasks, and extend the proposed theorems under the probabilistic NCC model. Another important direction that we would like to pursue is to create more efficient Secure Multi-party Computation techniques tailored towards implementing the data analysis tasks that are in DNCC.

V. REFERENCES

[1] Murat Kantarcioglu, U. of Texas at Dallas, Wei Jiang, Missouri S, **"**Incentive Compatible Privacy-Preserving Data Analysis", IEEE Transactions on Knowledge and Data Engineering.

[2] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. In Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing, pages 53–62. ACM New York, NY, USA, 2006.

[3] R. Agrawal and E. Terzi. On honesty in sovereign information sharing. Lecture Notes in Computer Science, 3896:240, 2006.

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

[5] I. Ashlagi, A. Klinger, and M. Tenneholtz. K-NCC: Stability against Group Deviations in Non-Cooperative Computation. LECTURE NOTES IN COMPUTER SCIENCE, 4858:564, 2007.

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

[7] B. Chor and E. Kushilevitz. A zero-one law for Boolean privacy. In STOC '89, pages 62–72, New York, NY, USA, 1989. ACM Press.

[8] www.doe.gov, doe news, Feb. 16 2005.

[9] Wenliang Du and Zhijun Zhan. Building decision tree classifier on private data. In Chris Clifton and Vladimir Estivill-Castro, editors, IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining, volume 14, pages 1–8, Maebashi City, Japan, December 9 2002. Australian Computer Society.

[10] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Communities, No I.(281):31–50, October 24 1995.

[11] Keinosuke Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, San Diego, CA, 1990.

[12] 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.