© 2014 IJIRT | Volume 1 Issue 5 | ISSN : 2349-6002

CONFIDETIALITY CONSERVE DATA ANALYSIS FOR INCENTIVE COMPATIBLE SYSTEM

Galla Narahari Krishna[1], D. Bheekya[2]
[1]*M.Tech, CSE Dept, MLRIT, Hyderabad*
[2]*M.Tech, Asst. Professor, MLRIT, Hyderabad*

***Abstract*— when compare any contest parties who maintain private data may cooperatively handle privacy preserving distributed data analysis (PPDA) assignments to determine constructive data models or analysis results. For example, different manufacturing companies may try to build better models for the products fraud detection through PPDA tasks. Similarly, contesting companies in the same industry may try to combine their sales data to build models that may predict the future sales. In many of these cases, the competing parties have different incentives. 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. This raises the question of how to design incentive compatible privacy-preserving data analysis techniques that motivate participating parties to provide truthful input data.**

***Index Terms*— Privacy, Secure multi-party computation, Non-cooperative computation.**

I. INTRODUCTION

There is an increasing requirement for sharing information across autonomous entities in such a way that only minimal and necessary information is disclosed. This requirement is being driven by several trends, including end-to-end integration of global supply chains, co-existence of competition and co-operation between enterprises, need-to-know sharing between security agencies, and the emergence of privacy guidelines and legislations. Sovereign information sharing [1, 3] allows autonomous entities to compute queries across their databases such that nothing apart from the result is revealed.

For example, suppose the entity R has a set VR = fb; u; v; yg and the entity S has a set VS = fa; u; v; xg. As the result of sovereign intersection VR \VS, R and S will get to know the result fu; vg, but R will not know that S also has fa; xg, and S will not know that R also has fb; yg. Several protocols have been proposed for computing sovereign relational operations, including [1, 3, 6, 8, 16]. In principle, sovereign information sharing can be implemented using protocols for secure function evaluation (SFE) [7].Given two parties with inputs x and y respectively, SFE computes a function f(x; y) such that the parties learn only the result. The above body of work relies on a crucial assumption, that the participants in the computation are semi-honest. This assumption basically says that the participants follow the protocol properly (with the exception that they may keep a record of the intermediate computations and received messages, and analyze the messages). Specifically, it is assumed that the participants will not maliciously alter the input data to gain additional information. This absence of malice assumption is also present in work in which a trusted-third party is employed to compute sovereign operations.

In a real imperfect world, the participants may behave dishonestly particularly when they can benefit from such a behavior. This benefit can come from learning more than necessary private information of others or preventing others from learning the necessary information. In the sovereign intersection example given in the beginning, R may maliciously add x to VR to learn whether VS contains x. Similarly, S may exclude v from VS to prevent R from learning that it has v.

Progress in bar-code technology has made it possible for retail organizations to collect and store massive amounts of sales data, referred to as the basket data. A record in such data typically consists of the transaction date and the items bought in the transaction. Successful organizations view such databases as important pieces of the marketing infrastructure. They are interested in instituting information-driven marketing processes, managed by database technology, that enable marketers to develop and implement customized marketing programs and strategies [S]. The problem of mining association rules over basket data was introduced in [4]. An example of such a rule might be that 98% of customers that purchase tires and auto accessories also get automotive services done. Finding all such rules is valuable for cross marketing and attached mailing applications. Other applications include catalog design, add-

on sales, store layout, and customer segmentation based on buying patterns. The databases involved in these applications are very large. It is imperative, therefore, to have fast algorithms for this task.

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. Department of Energy supports research on building much more efficient diesel engines [5]. Such an ambitious task requires the collaboration of geographically distributed industries, national laboratories and universities. Those institutions (including the potentially competing industry partners) need to share their private data for building data analysis models that enable them to understand the underlying physical phenomena. Similarly, different pharmaceutical companies may want to combine their private research data to predict the effectiveness of some protein families on certain diseases.

II. PROBLEM STATEMENT

Vampire attacks are not protocol-specific, in that they do not rely on design properties or implementation faults of particular routing protocols, but rather exploit general properties of protocol classes such as link-state, distance-vector, source routing, and geographic and beacon routing. Neither do these attacks rely on flooding the network with large amounts of data, but rather try to transmit as little data as possible to achieve the largest energy drain, preventing a rate limiting solution. Since Vampires use protocol-compliant messages, these attacks are very difficult to detect and prevent.

This paper makes three primary contributions. First, we thoroughly evaluate the vulnerabilities of existing protocols to routing layer battery depletion attacks. We observe that security measures to prevent Vampire attacks are orthogonal to those used to protect routing infrastructure, and so existing secure routing protocols such as Ariadne [9], SAODV [18], and SEAD [8] do not protect against Vampire attacks. Existing work on secure routing attempts to ensure that adversaries cannot cause path discovery to return an invalid network path, but Vampires do not disrupt or alter discovered paths, instead using existing valid network paths and protocol compliant messages. Protocols that maximize power efficiency are also inappropriate, since they rely on cooperative node behavior and cannot optimize out malicious action. Second, we show simulation results quantifying the performance of several representative protocols in the presence of a single Vampire (insider adversary). Third, we modify an existing sensor network routing protocol to provably bound the damage from Vampire attacks during packet forwarding.

III. SYSTEM DEVELOPMENT

Privacy-Preserving Data Analysis:

The privacy preserving data analysis protocols assume that participating parties are truthful about their private input data. The techniques developed in 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.

Non-Cooperative Computation:

In the NCC model, each party participates in a protocol to learn the output of some given function f over the joint inputs of the parties. First, all participating parties send their private inputs securely to a trusted third party (TTP), and then TTP computes f and sends back the result to every participating party. The NCC model makes the following assumptions.

1) Correctness:

The first priority for every participating party is to learn the correct result;

2) Exclusiveness:

If possible, every participating party prefers to learn the correct result exclusively.

Analyzing Data Analysis Tasks Ii The NCC Model:

Combining the two concepts DNCC and SMC, we can analyze privacy preserving data analysis tasks (without utilizing a TTP)that are incentive compatible. We next prove several such important tasks (as function with Boolean output, set operations, linear functions, etc) that either satisfy or do not satisfy the DNCC model. Also, note that the data analysis tasks analyzed next have practical SMC implementations.

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. The Security Code is valid means retrieve the data otherwise you are a fraud user.

IV. RELATED WORK

Secure multi-party computation (MPC) is one of the most surprising computational phenomena known. In fact, the paradigm encompasses not one, but a wide range of phenomena, depending on the MPC task (functionality) in question. However, in another influential work, Canetti [C01] showed that under a more demanding but more realistic model of security, at least one qualitative distinction exists among MPC functionalities, regardless of any computational assumption: the separation between "trivial" and "non-trivial" functionalities. In this paper we show that, under the same intractability assumption needed for the results in [GMW87], the distinction between trivial and non-trivial functionalities is the only qualitative distinction among deterministic 2-party functionalities in Canetti's stronger security framework for MPC.

More formally, we use a natural complexity-theoretic reduction to compare the qualitative "cryptographic complexities" of MPC functionalities. We say that a functionality F reduces to G (written F vPPT G) if there is a secure protocol for F that uses ideal access to G. We use the strong definition of security from the framework of Universal Comparability (UC) [C01]. Under this reduction, there are two natural extremes of cryptographic complexity: we call functionality "trivial" if it can be reduced to every other functionality and "complete" if every functionality can be reduced to it. Stated in these terms, our main result is the following:

The following two statements are equivalent:

Zero-One Law: Every deterministic, finite 2-party functionality is either trivial or complete. sh-OT Assumption: There exists a 2-party protocol that securely realizes the oblivious transfer functionality against semi-honest (a.k.a., passive, honest-but-curious) PPT adversaries.

The zero-one law applies not just to secure function evaluation functionalities, but also to reactive ones that receive input and give output repeatedly over many rounds of interaction, maintaining secret state between rounds. To the best of our knowledge, ours is the first work that considers how to use arbitrary reactive functionalities for other cryptographic purposes. To establish the zero-one law, we advance on two technical fronts in the study of complexity of secure multi-party computation. The first front focuses on understanding distinct non-trivial behavioral components of (possibly reactive) functionalities. We identify a list of four qualitatively distinct such components. For each one we can associate a familiar "canonical" functionality which is non-trivial for only that reason:

- Allowing simultaneous exchange of information, exemplified by the Boolean XOR functionality FXOR.
- Selectively hiding one party's inputs from the other, exemplified by a simple SFE functionality we introduce called simple cut-and-choose, FCC.
- Selectively hiding both party's inputs simultaneously, exemplified by the oblivious transfer functionality FOT.
- Holding meaningful hidden information in internal memory between rounds, exemplified by the commitment functionality FCOM. (This component can appear only in a reactive functionality.)

Notions from game theory are used widely in this paper. Game theory was founded by von Neumann and Morgenstern as a general theory of rational behavior.

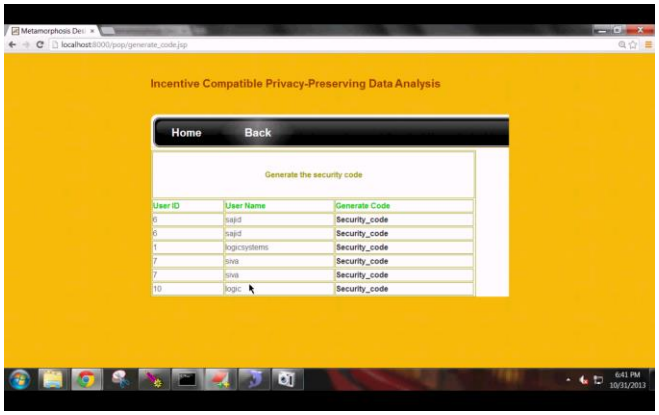
Games related to our work include the interdependent security (IDS) games [10, 13]. They were primarily to model scenarios where a large number of players must make individual investment decisions related to a security – whether physical, financial, medical, or some other type - but in which the ultimate safety of every participant depends on the actions of the entire population. IDS games are closely related to summarization games [11] in which the players' payo is a function of their own actions and the value of a global summarization function that is determined by the joint play of the population. Summarization games themselves are extensions of congestion games [15, 9] in which players compete for some central resources and every player's payo is a decreasing function of the number of players selecting the resources. We have adopted some notions from the IDS games and used them to model information exchange. However, our problem is different from the one presented in [13], while at the same time we are not exploring algorithms for computing the equilibrium of the games as in [10].

Inspection games [4, 5, 14] are also related to our work. These are games repeated for a sequence of iterations. There is an inspector responsible for distributing a given number of inspections over an inspection period. Inspections are done so that possible illegal actions of an inspective can be detected. The question addressed is what are the optimal strategies for the inspector and the inspect in such a game.

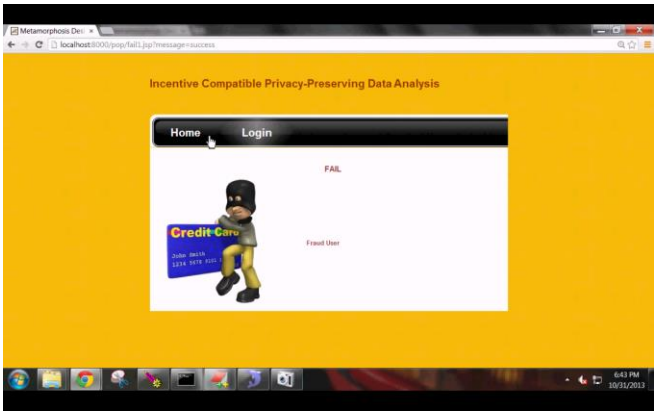
The main difference between these games and the game we have designed is that in the inspection games the inspector is a player of the game. This is not true for our game, where the inspector acts as a referee for the players, helping them (via auditing) to achieve honest collaboration.

In [12], different information-exchange scenarios are considered and the willingness of the participants to share their private information is measured using solution concepts from coalition games. Our study is complementary to this work. We are interested in quantifying when people are willing to participate truthfully in a game, rather than the complementary question of whether they are willing to participate at all. The work presented in [20] models information exchange between a consumer and a web site. Consumers want to interact with web sites, but they also want to keep control of their private information. For the latter, the authors empower the consumers with the ability to test whether a web site meets their privacy requirements. In the proposed games, the web sites signal their privacy policies that the consumers can test at some additional cost. The main conclusion of the study is that such a game leads to cyclic instability. The scenario we are modeling is completely different. Our players are all empowered with the same set of strategies. Our games also admit multiple players.

EXPERIMENTAL RESULTS



The techniques developed in assume that each party has an internal device that can verify whether they are telling the truth or not.



The Security Code is valid means retrieve the data otherwise you are a fraud user.

V. CONCLUSION

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. In this paper, we have investigated what kinds of PPDA tasks are incentives compatible under the NCC model. Based on our findings, there are several important PPDA tasks that are incentive driven. As a future work, we will investigate incentive issues in other data analysis tasks, and extend the proposed theorems under the probabilistic NCC model.

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, Chile, September 12-15 1994. VLDB.

[2] Rakesh Agrawal and Evimaria Terzi. On honesty in sovereign information sharing. In EDBT, pages 240–256, 2006.

[3] 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.

[4] 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.
[5] www.doe.gov, doe news, feb. 16 2005.
[6] 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.
[7] 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.
[8] Keinosuke Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, San Diego, CA, 1990.
[9] 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.
[10] Oded Goldreich. The Foundations of Cryptography, volume 2, chapter General Cryptographic Protocols. Cambridge University Press, 2004.
[11] Joseph Halpern and Vanessa Teague. Rational secret sharing and multiparty computation: extended abstract. In STOC ’04, pages 623–632, New York, NY, USA, 2004. ACM Press.
[12] Standard for privacy of individually identifiable health information. Federal Register, 67(157):53181–53273, August 14 2002.
[13] Geetha Jagannathan and Rebecca N. Wright. Privacy-preserving distributed k-means clustering over arbitrarily partitioned data. In Proceedings of the 2005 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 593– 599, Chicago, IL, August 21-24 2005.
[14] 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.
[15] Xiaodong Lin, Chris Clifton, and Michael Zhu. Privacy preserving clustering with distributed EM mixture modeling. Knowledge and Information Systems, 8(1):68–81, July 2005.
[16] Yehuda Lindell and Benny Pinkas. Privacy preserving data mining. In Advances in Cryptology – CRYPTO 2000, pages 36–54. Springer-Verlag, August 20-24 2000.
[17] Yehuda Lindell and Benny Pinkas. Privacy preserving data mining. Journal of Cryptology, 15(3):177–206, 2002.
[18] Robert McGrew, Ryan Porter, and Yoav Shoham. Towards a general theory of non-cooperative computation (extended abstract). In TARK IX, 2003.
[19] 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.
[20] Noam Nisan and Amir Ronen. Algorithmic mechanism design (extended abstract). In STOC’ 99, pages 129–140, New York, NY, USA, 1999. ACM Press.

AUTHOR DETAILS:

First Author: ***Galla Narahari Krishna*** received B.Tech Information Technology from VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, in the year 2010. He is currently M.Tech student in Computer Science and Engineering Department from Marri Laxman Reddy Institute of Technology. And his research interested areas are in the field of Networking, Information Security and Cloud Computing, Mobile Computing.

Second Author: ***D. Bheekya*** working as an Asst. Professor in Marri Laxman Reddy Institute of Technology. He has completed his M.Tech CSE and he has 6 years of teaching experience. His research interested areas are Data Mining, Network Security and Cloud Computing.