Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.367 – 371

RESEARCH ARTICLE

Privacy Preserving Incentive Compatible Way

Shweta, B.E, (M.Tech), **C.H.Kiran**, B.Tech, M.Tech
Computer Science and Engineering, TITS, JNTU, Hyderabad, TS, India
shwetamadakattimtech@gmail.com , kiran.30.aug@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. Most often, 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 participating parties are truthful about their private input data. Unless proper incentives are set, 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 inputs. In this paper, we first develop key theorems, and then base on these theorems, we 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.

I. INTRODUCTION

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. Such an ambitious task requires the collaboration of geographically distributed industries, national laboratories, and universities. 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 growing problem of identity theft. To prevent misuse of data, there is a recent surge in laws mandating protection of confidential data, such as the European Community privacy standards, U.S. health-care laws and California SB1386. However, this protection comes with a real cost through both added security expenditure and penalties and costs associated with disclosure. What we need is the ability to compute the desired "beneficial outcome" of data sharing for analyzing without having to actually share or disclose data. This would maintain the security provided by separation of control while still obtaining the benefits of a global data source.

Secure multiparty computation (SMC) 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 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, decision tree analysis and auctions among others. Nevertheless, the SMC model does not guarantee that data provided by participating parties are truthful. In many real-life situations, data needed for building data analysis models are distributed among multiple parties with potentially conflicting interests. For 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 peers. An engine design company may want to exclusively learn the data analysis models that may enable it to build much more efficient diesel engines. 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. This assumption is usually justified by the fact that learning the correct data analysis models or results is in the best interest of all participating parties. Since SMC-based protocols require participating parties to
Perform expensive computations, if any party does not want to learn data models and analysis results, the party should not participate in the protocol. Still, this assumption does not guarantee the truthfulness of the private input data when participating parties want to learn the final result exclusively. For example, a drug company may lie about its private data so that it can exclusively learn the data analysis model. Although SMC protocols guarantee that nothing other than the final data 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, current SMC techniques cannot prevent input modification by participating parties. To better illustrate this problem, we consider a case from management where competing companies (e.g., Texas Instruments, IBM and Intel) establish a consortium (e.g., Semiconductor Manufacturing Technology). The companies send the consortium their sales data, and key manufacturing costs and times. Then, the consortium analyzes the data and statistically summarizes them in a report of industry trends, which is made available back to consortium members. In this case, it is in the interest of companies to learn true industry trends while revealing their private data as little as possible. Even though SMC protocols can prevent the revelation of the private data, they do not guarantee that companies send their true sales data and other required information. Assume that n companies would like to learn the sample mean and variance of the sales data for a particular type of product.

II. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into account for developing the proposed system.

III. EXISTING SYSTEM

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.

IV. OBJECTIVES

- In this paper, we study user profile matching with privacy-preservation in mobile social networks (MSNs) and introduce a family of novel profile matching protocols.
- The common goal of these works is to enable the handshake between two encountered users if both users satisfy each other's requirement while eliminating the unnecessary information disclosure if they are not.
- We analyze the communication overhead and the anonymity strength of the protocols.

V. PROPOSED SYSTEM

In design incentive compatible privacy-preserving data analysis techniques that motivate participating parties to provide truthful input data. 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. Secure multi-party computation (SMC) has recently emerged as an answer to this problem.

VI. SYSTEM ANALYSIS

Feasibility study

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

Technical feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Economical feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized

Operational feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

VII. Data Flow Diagram (DFD's)

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system.

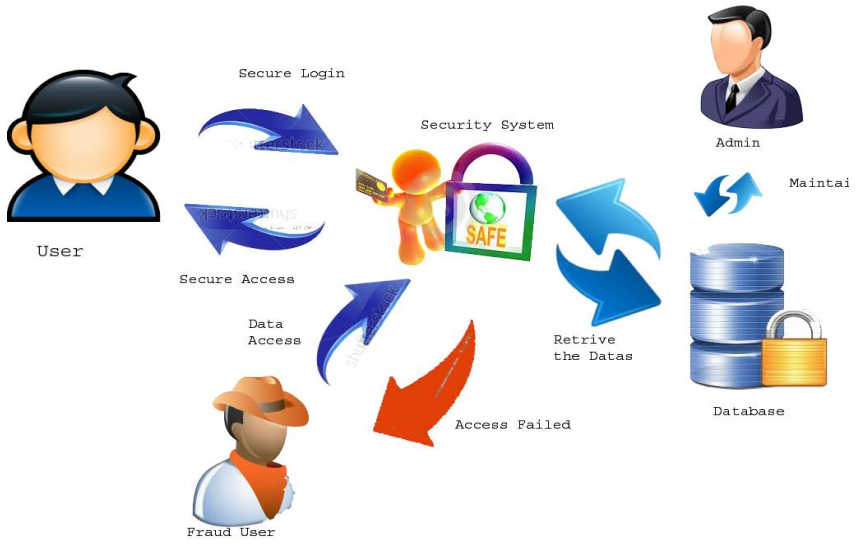

Fig (a):

VIII. 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 is incentive 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] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization:A geometric framework for learning from labelled and unlabeled examples. *Journal of Machine Learning Research*, 7(Nov):2399–2434, 2006.
[2] J. Blitzer, R. Mcdonald, and F. Pereira. Domain adaptation with structural correspondence learning. In *EMNLP '06: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Pro-cessing*, pages 120–128. Association for Computational Linguistics, July 2006.
[3] C. J. C. Burges, R. Ragno, and Q. V. Le. Learning to rank with nonsmooth cost functions. In *NIPS '06: Advances in Neural Infor-mation Processing System*, pages 193–200. MIT Press, Cambridge, MA, 2006.
[4] C. J. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22th International Conference on Machine Learning*, 2005.
[5] Z. Cao and T. yan Liu. Learning to rank: From pairwise approach to listwise approach. In *ICML '07: Proceedings of the 24th Interna-tional Conference on Machine Learning*, pages 129–136, 2007.
[6] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In *ACM Multimedia*, pages 729–732, 2008.
[7] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu. Boosting for transfer learning. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 193–200, 2007.
[8] H. Daume, III and D. Marcu. Domain adaptation for statistical classifiers. *Journal of Artficial Intelligence Research*, 26:101–126, 2006.
[9] Y. Freund, R. Iyer, R. E. Schapire, Y. Singer, and G. Dietterich. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.
[10] B. Geng, L. Yang, C. Xu, and X.-S. Hua. Ranking model adaptation for domain-specific search. In *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*, pages 197–206, 2009.

Sites Referred:
http://java.sun.com
http://www.roseindia.com/
http://www.java2s.com/

