International Journal of Advanced Research in

Computer Science Engineering and Information Technology

Volume: 2 Issue: 1 08-Feb-2014,ISSN_NO: 2321-3337

Secure Sharing Of Private Data Using Privacy-Preserving Data Analysis

D.S.Deepika[1], R.Nandhini[2], L.Nandhini[3], A.Sangeeth Sharmili[4]

Asst.Professor, Dept. of Computer Science, Velammal Institute of Technology, India[1]

Dept. of Computer Science, Velammal Institute of Technology, India[2 3 4]

deepi.soorasamharan@gmail.com[1]

ravinandhini@ymail.com[2],nandhinilatha@ymail.com[3],sangeethsharmili@yahoo.in[4]

Abstract **--** *Generally, colliding parties who have private data may conduct privacy-preserving data analysis (PPDA) tasks to learn beneficial data models in a distributed manner. The field of privacy has seen rapid advances in recent years because of the increases in the ability to store data. In particular, recent advances in the data mining field have lead to increased concerns about privacy. While the topic of privacy has been traditionally studied in the context of cryptography and information-hiding, recent emphasis on data mining has lead to renewed interest in the field. In this paper, we will introduce the topic of privacy-preserving data mining. It is often highly valuable for organizations to have their data analyzed by external agents. However, any program that computes on potentially sensitive data may lead to risks leaking information through its output. Differential privacy provides a theoretical framework for processing data while protecting the privacy of individual records in a dataset. Unfortunately, it has seen limited adoption because of the loss in output accuracy, the difficulty in making programs differentially private, lack of mechanisms to describe the privacy budget in a programmer's utilitarian terms. So, in this paper we have proposed how to share private data securely.*

Keywords - Privacy, secure multiparty computation, noncooperative computation

1, INTRODUCTION

Technical advancements had lead to difficulties in preserving and securing data. Data analysis requires data sharing among several sources. But such data sharing also involves the sharing of private data. Sharing of private data may become harmful in the case of misuse of such data. European Community privacy standards [5], U.S. health-care laws [15], and California SB1386 are some of the laws formulated for the protection of confidential data. Yet, these laws involve real cost which include expenditures for security. The main aim is to fulfill the purpose of data sharing for data analysis without any explicit sharing or letting out of data.

This serves the purpose by providing security and at the same time enabling the benefits of a global data source. Secure multiparty computation (SMC) [7], is a latest solution to this problem. Nowadays, data management applications have evolved from pure storage and retrieval of information to finding interesting patterns and associations from large amounts of data. With the advancement of Internet and networking technologies, more and more computing applications, including data mining programs, are required to be conducted among multiple data sources that scattered around different spots, and to jointly conduct the computation to reach a common result. However, due to legal constraints and competition edges, privacy issues arise in the area of distributed data mining, thus leading to the interests from research community of both data mining. According to SMC, the parties which participate will learn only the final outcome and what is deduced from their own private inputs. Some applications of SMC protocol are forecasting [9], decision tree analysis [19] and auctions [16] among others. Yet, SMC does not provide guarantee of data i.e the data furnished by the parties may be truthful or not. Data analysis is usually carried out among parties with adverse interests. In SMC, the general assumption is that the inputs provided by the parties are truthful. The main interest of all participating parties is to learn the results.SMC requires expensive computations and so parties find it difficult to participate. One major detriment of the SMC protocol is it is not possible to verify whether the parties are truthful about their private data which is given as input.

2, RELATED WORKS

A function is non-cooperative computable [NCC] if honest agents can compute it by reporting truthfully their private inputs, while unilateral deviations by the players are not beneficial. If a deviation from truth revelation can mislead other agents, then the deviator might end up with a wrong result. Previous work provided full characterization of the Boolean functions which are non-cooperatively computable.By [20], it extends the study of NCC functions to the context of group deviations. According to [11] we analyze how to build a decision tree classifier under the following scenario: a database is vertically partitioned into two pieces, with one piece owned by Alice and the other piece owned by Bob. Alice and Bob want to build a decision tree classifier based on such a database, but due to the privacy constraints, neither of them wants to disclose their private pieces to the other party or to any third party. We present a protocol that allows Alice and Bob to conduct such a classifier building without having to compromise their privacy.

3, SYSTEM ANALYSIS

3.1 Existing System

In the existing system, we use SMC. Secure multi-party computation (also known as secure computation or multi-party computation (MPC)) is a subfield of cryptography. The goal of this

field is to create methods that enable parties to jointly compute a function over their inputs, while at the same time keeping these inputs private. The concept is important in the field of cryptography and is closely related to the idea of zero knowledge. . For example, two individuals who each possess some secret information—x and y, respectively—may wish to jointly compute some function $f(x,y)$ without revealing any information about x and y other than what can be reasonably deduced by knowing the actual value of $f(x,y)$, where "reasonably deduced" is often interpreted as equivalent to computation within polynomial time. The primary motivation for studying methods of secure computation is to design systems that allow for maximum utility of information without compromising user privacy. Secure computation was formally introduced in 1982 by A. Yao.

3.2 Proposed System

Multiple parties can participate using PPDA .This is in contrast to SMC protocol which is used in the existing system. Each party participates in PPDA to learn the output of some function f over the joint inputs of the parties. First, all participating parties send their private inputs securely to a trusted third party (TTP), then TTP computes f and sends back the result to every participating party. Here TTP will checks either participant's parties give truthful inputs or not. In addition to the final result, we analyse the incentive issues and display the list of the organisations in which the individual takes part. The techniques used in the proposed system are discussed below.

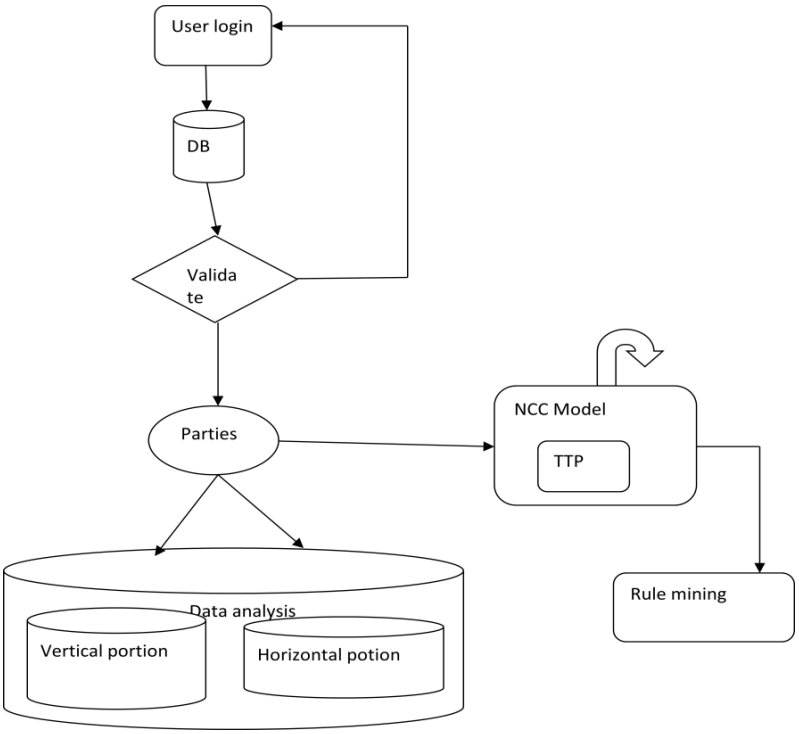

Figure.1 System Architecture

3.2.1 Noncooperative Computation:

This model works by collecting the private inputs from the participating parties and jointly computing the correct function results. We introduce the concept of non-cooperative computation (NCC), which is the co-operative computation of a function by self-motivated agents, where each of the agents possesses one of the inputs to the function. In NCC the agents communicate their input (truthfully or not) to a trusted center, which performs a commonly-known computation and distributes the results to the agents. The question is whether the agents can be incented to communicate their true input to the center, allowing all agents to compute the function correctly. NCC is a game theoretic concept.

3.2.2 Association Rule Mining:

Proposed by Agrawal et al in 1993. It is an important data mining model studied extensively by the database and data mining community.

$I = \{i_1, i_2, \ldots, i_m\}$: a set of *items*.
T: a set of transactions
$T = \{t_1, t_2, \ldots, t_n\}$.
Computation result is correct with probability one, 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.

ALGORITHM

- Either $\exists v_{-i} \in D_{-i},\ g_i(f(t_i(v_i), v_{-i}), v_i) \neq f(v_i, v_{-i})$
- Or $\forall v_{-i} \in D_{-i},\ f(t_i(v_i), v_{-i}) = f(v_i, v_{-i})$

The above definition simply states what function could be computed in NCC setting deterministically (i.e., computation result is correct with probability one), 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. In other words, if a party **i** replaces its true input **vi** with **v_ i** and if **f(v_ i, v−i) _= f(vi, v−i**), then party **i** should not be able to calculate the correct **f(vi, v−i)** from **f(v_ i, v−i**). And **vi.** Note that strategy **(ti, gi)** means that the way the input is modified, denoted by **ti,** and the way the output is calculated, denoted by **gi**. In **ti** can be considered as choosing a value different from the actual input, and **gi** can be considered as the ways the correct **μ** and **s2** are computed. Another implication of the above definition is that for any **ti,** the corresponding **gi** should be deterministic, because each party want to exactly compute the "correct" result. A two-party protocol is proposed to securely compute JC. The protocol consists of two stages

- **Stage 1** - Computing Random Shares of $|D_1 \cap D_2|$:
  At the end, P_1 has a random number α_1 and P_2 has a random number α_2, such that $\alpha_1 + \alpha_2 = |D_1 \cap D_2|$.
- **Stage 2** - Computing JC Score:
  P_1 sets $\beta_1 = |D_1| - \alpha_1$ and P_2 sets $\beta_2 = |D_2| - \alpha_2$.
  Both parties securely compute $\frac{\alpha_1+\alpha_2}{\beta_1+\beta_2}$.

4, SPECIFICATION

4.1 Functional Specification

A functional requirement defines a function of a software-system or its component. A function is described as a set of inputs, the behavior, and outputs. The proposed system is achieved by creating a network coding between virtual servers.

4.2 Non-Functional Specification

4.2.1 Efficiency

Providing a network coding between virtual servers and improve the client downloading speed and decrease the sink workload.

5, APPLICATION

To deal with the multi Nash equilibrium problem of non cooperative game based spectrum sharing in cognitive radio networks, we use the variation of utility of cognitive users to judge the stability after several iterations. We limit our study to simple single-level master-worker platforms and to the case where each scheduler is in charge of a single application consisting of a large number of independent tasks

6, FUTURE ENHANCEMENTS

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.

7, CONCLUSION

Through this paper we have investigated privacy issues and proposed a solution using the game theoretic concept NCC .Further we would analyze the incentive issues and enhance the system using DNCC in future.

8, REFERENCES

[1] 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.

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

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

[4] Shuguo Han and Wee Keong Ng. Preemptive measures against malicious party in privacy-preserving data mining. In SDM, pages 375–386, 2008.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

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

[11] 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.

[12]Oded Goldreich. The Foundations of Cryptography, volume 2, chapter General Cryptographic Protocols. Cambridge University Press, 2004.

[13] S.D. Gordon and J. Katz. Rational secret sharing, revisited. Lecture Notes in Computer Science, 4116:229, 2006.

[14] 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.

[15] Standard for privacy of individually identifiable health information. Federal Register, 67(157):53181–53273, August 14 2002.

[16] M. Naor, B. Pinkas, and R. Sumner, "Privacy Preserving Auctions and Mechanism Design," Proc. First ACM Conf. Electronic Commerce, 1999.

[17] J. Han and M. Kamber. Data mining: concepts and techniques. The Morgan Kaufmann series in data management systems. Elsevier, 2006.

[18] S. Izmalkov, S. Micali, and M. Lepinski. Rational secure computation and ideal mechanism design. In Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on, pages 585–594, 2005.

[19]Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, vol. 15, no. 3, pp. 177-206, 2002.

[20] Itai Ashlagi, Andrey Klinger, and Moshe Tennenholtz Technion–Israel, K-NCC: Stability against Group Deviations in Non-Cooperative Computation, 2007.

[21] Y. Shoham and M. Tennenholtz, "Non-Cooperative Computation: Boolean Functions with Correctness and Exclusivity," Theoretical Computer Science, vol. 343, nos. 1/2, pp. 97-113, 2005.

[22] 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.