Dhanalakshmi.M *et al*, International Journal of Computer Science and Mobile Applications,

Vol.2 Issue. 1, January- 2014, pg. 170-179 **ISSN: 2321-8363**

EFFICIENT INCENTIVE COMPATIBLE MODEL FOR SECURE DATA SHARING

Ms. Dhanalakshmi.M
PG Student, CSE
Government College of Engineering,
Tirunelveli, India

Mrs.Siva Sankari.E
Assistant Professor, CSE
Government College of Engineering,
Tirunelveli, India

ABSTRACT

Privacy preserving is one of the most important research topics in the data security field and it has become a serious concern in the secure transformation of personal data in recent years. A number of algorithmic techniques have been designed for Privacy Preserving Data Mining (PPDM). For example, different credit card companies may try to build better data sharing models for credit card fraud detection through PPDA tasks. 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. This raises the question of how to design incentive compatible privacy-preserving data analysis techniques that motivate participating parties to provide truthful input data. A model has been proposed to design the effective incentive compatible model for secure private information transformation. Every user need to transform the private data in many applications (E-Shopping), that must be protected from the hackers. Building this model depends on many privacy preserving data mining techniques like Cryptographic techniques, Privacy preserving association rule and Function Evaluation Theorem. The Incentive model is very efficient for privacy preserving data mining, because it provides the protocols against not only semi-honest adversary model and also the malicious model.

Keywords: Privacy Preserving, Data sharing, Privacy preserving techniques, secure multi-party computation, Non-cooperative computation.

INTRODUCTION

Data mining is a process that uses a variety of data analysis tools to deliver the patterns and relationships in data sets that may be used to make valid predictions. The first step in data mining is to describe the data .It means that summarize its statistical attributes, visually reviewed it using charts and graphs, and look for meaningful links among variables .In the Data Mining Process, collecting, exploring and selecting the right data are critically important one. To build a predictive model based on patterns determined from known results, then testing that model depends on results outside the original sample. A good model should never be critical with reality. The final step is to empirically verify the model.

The main objective of the paper is to maintain the confidentiality of the data in data transaction. Various types of web services models are used in private information transformation applications. These security models are based on the various types of Privacy Preserving Data Mining (PPDM) techniques such as Randomization method, Anonymization method and Encryption method. In many applications , certain PPDA techniques guarantee that nothing other than the final analysis result is revealed, it is impossible to check whether participating parties are truthful about their private input data in data sharing. The incentive compatible privacy-preserving data analysis techniques have been developed that motivate participating parties to provide truthful inputs.

The incentive compatible privacy preserving model has to interact with the users to verify the transaction making use of the users knowledge. The E-Shopping is an application which provides a user interaction interface that provides more security for sensitive (personal) information transformation compared with the other privacy preserving models. Privacy Preservation is most important area in data mining. Privacy Preservation techniques are used to protect the users private data from unauthorized person.

Data mining includes various types of privacy preservation techniques:

1. Data distribution
2. Data modification
3. Data mining algorithm
4. Data or rule hiding
5. Privacy preservation

Secure data sharing, various approaches are used to design the different privacy preserving models. The Privacy-preserving data mining (PPDM), which mainly considered four categories of models such as Trust Third Party Model , Semi-honest Model , Malicious Model , Other Models-Incentive Compatibility.

The data modification technique also includes four types of methods such as Perturbation, which is accomplished by the alteration of an attribute value with a new value (i.e., changing a 1-value of a 0-value, or adding noise).*B*locking is the replacement of an existing attribute value with an aggregation or merging which is the combination of several values. Swapping that refers to interchanging values of individual records, and Sampling, which is referred to delivering the data for only a sample of a population.

SYSTEM ARCHITECTURE

A better data analysis model is generated which has the ability to compute the desired "beneficial outcome" of data sharing for analyzing without having to actually share or disclose data. 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. For example, different credit card companies may try to build better models for credit card fraud detection through PPDA tasks. The figure 1 represents the model for Privacy Preserving System Architecture.

The system architecture of privacy preserving gives the detailed explanation about the function of the security system in which it allows only the authorized person. Incase, if any fraud user is trying to access the data security system will not allow user the access will be denied for the particular user . And the datas are retrived from the database according to the request given by user.

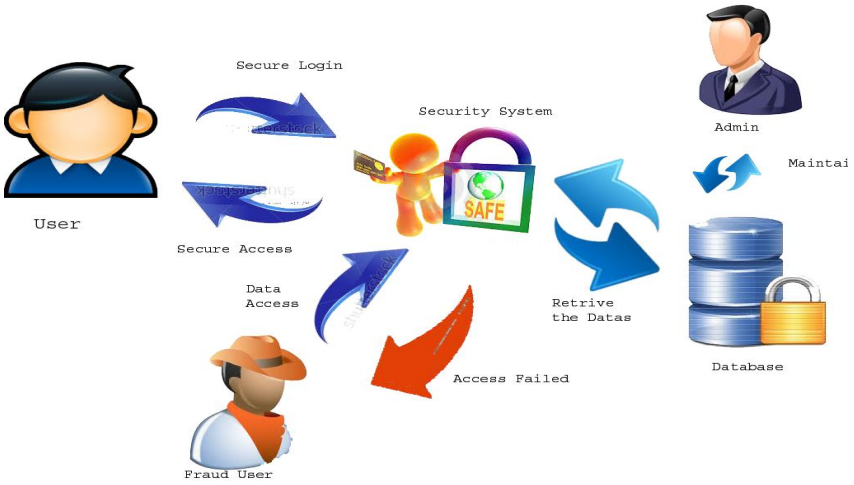

Fig 1 : Privacy Preserving System Architecture

The incentive compatible privacy preserving model that provides more security for sensitive (personal) information transformation compared with the other privacy models. It does not require fraud signatures and yet is able to detect frauds by considering the data analysis model's result. Another important advantage of our project is a drastic reduction in the number of False Positive transactions. It tries to find any anomalies in the transaction based on the data analysis model.

Non-Cooperative Computation Model

The non-cooperative computation (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), then TTP

computes f and sends back the result to every participating party. The NCC model makes the following assumptions:

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

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

Under the *correctness* and *exclusiveness* assumptions, the NCC model is formally defined as follows: Given a set of n parties, for a party i,

1) Each party i sends v′ i (not necessarily the correct private input) to a TTP.

2) The TTP computes f (v′) = f (v′ 1 , . . . , 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 vi.

Considering the above simple protocol does not limit its generality. Under the literature of SMC, the TTP can be replaced such that the required functionality (represented by f) is still computable without violating privacy regarding each participating party's private input. The parties are expected to provide their true inputs to correctly evaluate a function that satisfies the NCC model. Therefore, any functionality that satisfies the NCC model is inherently incentive compatible under the assumption that participating parties prefer to learn the function result correctly, and if possible exclusively. In the NCC model considers the two types of model,

1. D-NCC (Deterministic Non-Cooperative Computation)
2. P-NCC (Probablistic Non-Cooperative Computation)

The D-NCC model, the function f is in the DNCC because the proof needs to consider all possible ti and gi pairs. The strategy ti defines a way to change the input, and the strategy gi defines a method to reconstruct the actual result based on the true input, modified input and the result computed based on the modified input and other parties' input data.

The Incentive compatible model is a web service model for online shopping, online recharge and many online applications. The figure 2 represents the flowchart for incentive compatible system model. The system is used to protect the user details in data sharing such as

payment processing. This model is built using data mining techniques such as association rules, Horizontal partitioning of the table and Vertical partition of the data.

The system considers a distributed database like bank database that is used to construct the incentive marking. The incentive data are used to check the user knowledge that is the processing user is correct person or not.

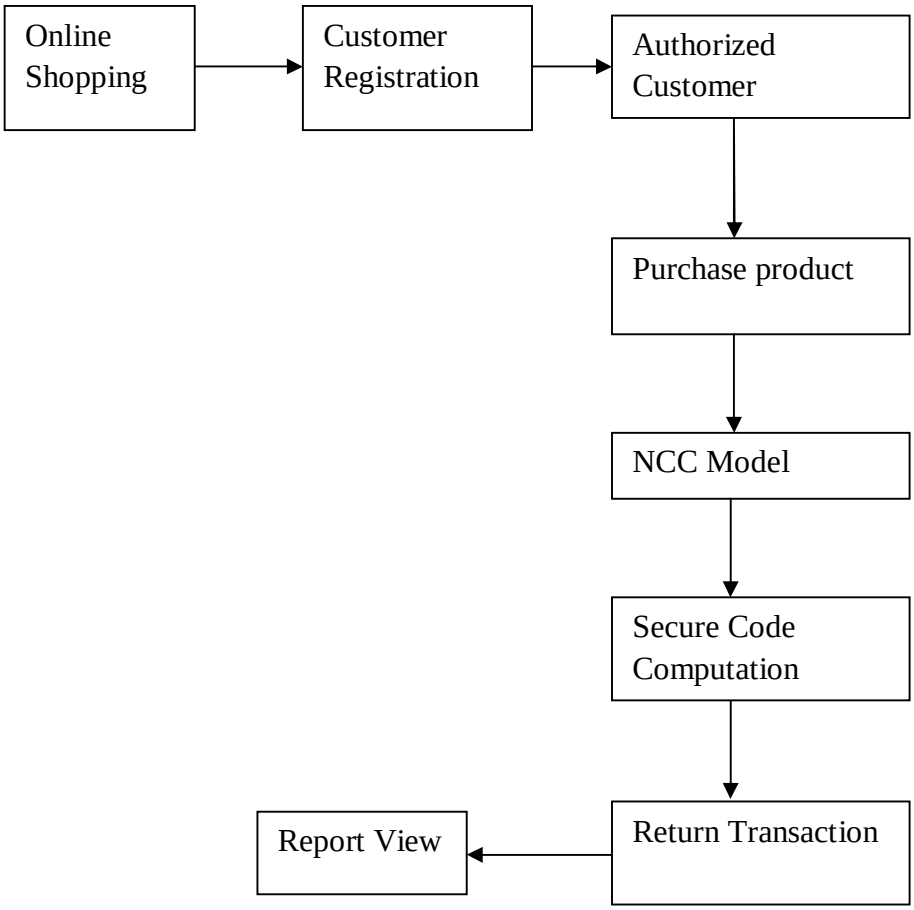

Fig 2 : Flowchart for Incentive Compatible System Model

Incentive Marking

Analyze what types of distributed functionalities could be implemented in an incentive compatible fashion. In other words, exploring 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. The secure multiparty computation process includes the following privacy preserving data mining techniques:

Horizontal partitioning for the table means to extract the fields that performing extract the collection of incentive data's according to particular customer. Perform this task to take the user's private detail as an input to the NCC model. Built the DNCC model with help of Association rules technique that provides information of this type in the form of "if-then" statements.

Secure code computation process , is providing incentive compatible secret code question for the NCC Model. The computation process theorem consists of following steps:

Step1: Select two fields from customer details from bank database as input for secure code computation process.

Step2: Here first field is constant and another one field is other information of customer details. For example (one field is username that is constant , another one field is other information like dob , accno , emailid ,etc..).

Step3: Apply vertical partition on the first field data and attaching second field in the middle of partition data using Secure Sum Process technique.

CONCLUSION

The incentive compatible privacy-preserving data analysis technique have been developed to motivate the participating parties to provide truthful inputs. The privacy preserving data analysis provide a new model called Non Cooperative Computation model in which it maximize the secure data sharing during the information transformation. Privacy Preserving data analysis the task using incentive compatibility under NCC model. The main advantage is that it reduces the number of False Positive transactions. It tries to find any anomalies in the transaction based on the data analysis model.

The deterministic non cooperative computation model includes privacy preserving data mining techniques such as horizontal partitioning, association rules classification and secure multiparty computation process. The DNCC model, works under the principle of incentive data in some order which can be identified by the anomalies easily . To avoid these type of problem the incentive data's are placed in probabilistically under PNCC Model.An efficient incentive compatible model used for many privacy preserving application approaches to interact with user original knowledge and develop the new model in the future.

FUTURE ENHANCEMENT

1) To provide more securable Randomized personalized privacy preservation NCC model (P-NCC).
2) To improve the efficiency of implementation and ensure availability of the result in order to meet the various privacy preserving techniques.
3) To analyze the Deterministic NCC model with the Probabilistic NCC model.

REFERENCES

[1] G. Loukides, J.H. Shao, “An Efficient Clustering Algorithm for k-Anonymisation”, International Journal of Computer Science And Technology, Vol. 23, no. 2, pp. 188-202, 2008.

[2] Jaideep Vaidya, Chris Clifton, "Privacy-Preserving Data Mining: Why, How, and When," IEEE Security and Privacy, vol. 2, no. 6, pp. 19-27, November-December, 2004.

[3] Lindell, Yehuda, Pinkas, “Privacy preserving data mining”, In Proceedings of the Advances in Cryptology–CRYPTO, pp. 36–54,2000.

[4] Li Liu , Murat Kantarcioglu and Bhavani Thuraisingham “Privacy Preserving Decision Tree Mining from Perturbed Data”,In Proceedings of the 42nd Hawaii International Conference on System Sciences – 2009.

[5] Murat Kantarcioglu and Wei Jiang, “Incentive Compatible Privacy-Preserving Data Analysis”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 6, JUNE 2013.

[6] Wei , J.L. Lin and M.C., “Genetic Algorithm-Based Clustering Approach for k-Anonymization”, International Journal of Expert Systems with Applications, Vol. 36, no. 6, pp. 9784-9792, 2009.

[7] W. Jiang, C. Clifton, and M. Kantarco_glu, “Transforming Semi-Honest Protocols to Ensure Accountability,” Data and Knowledge Eng., vol. 65, no.1, pp. 57-74, 2008.

[8] W. Jiang and B.K. Samanthula, "A Secure and Distributed Framework to Identify and Share Needed Information," Proc. IEEE Int'l Conf. Privacy, Security, Risk and Trust (PASSAT '11), Oct.2011.

[9] X.K. Xiao, Y.F. Tao, "Personalized Privacy Preservation", In Proceedings of the ACM Conference on Management of Data (SIGMOD), pp. 229-240, 2006.

[10] X. Lin, C. Clifton, and M. Zhu, "Privacy Preserving Clustering with Distributed EM Mixture Modeling," Knowledge and Information Systems, vol. 8, no. 1, pp. 68-81, July 2005.

[11] Z. Huang, W. Du, B. Chen, "Deriving Private Information from Randomized Data", In Proceedings of the ACM SIGMOD Conference on Management of Data, Baltimore, Maryland, USA, pp. 37-48, 2005.

