Effective Incentive Compatible Model for Privacy Preservation of Information in Secure Data Sharing and Publishing

Mahesh Dumbere, Roshani Talmale
Department of Computer Science and Engineering, TGPCET Nagpur, Maharashtra, 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. For example, different credit card companies and disease control centers may try to build better data sharing or publishing models for privacy protection through privacy preserving data mining techniques (PPDM). A model has been proposed to design the effective Privacy Preserving Mining Framework for secure private information transformation and Publishing. Building this framework depends on Incentive Compatible Model based secure code computation process and PPDM techniques like Association rule mining, Randomization method and Cryptographic technique. An Encryption algorithm is used to identify which data sets need to be encrypted for preserving privacy in data storage publishing. The Incentive Compatible model is very efficient in protecting the sensitive data in privacy preserving data sharing, because it provides the secrecy against not only semi-honest adversary model and also the malicious model.

Keywords : Privacy Preserving, Privacy preserving data mining, Data publishing privacy, secure code computation

I. 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. To describe the data is the first step in data mining. It means that summarize its statistical attributes, visually reviewed it using charts and graphical view, and look for meaningful relationships among variables. The Data Mining Process, which is collected, exploring and selecting the right data are critically important one. To build a better predictive model based on patterns determined from known results, after that testing that model depends on results outside the original sample. A best model never should be critical with reality and utility. To empirically verify the model is the final step in data mining.

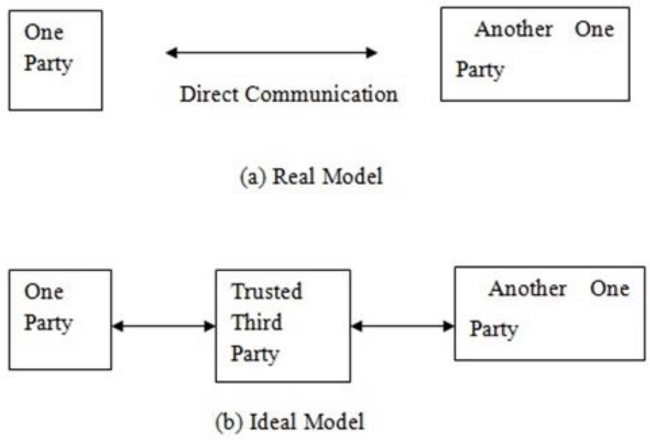
The rest of this paper is organized as follows: The details of privacy preserving in data mining will be discussed in section 2. To discuss about the Non – Cooperative Incentive Compatible Model and its process in section 3. In section 4, to show the implementation results of privacy preserving model.
In section 5 contains the conclusions and future work.

II. PRIVACY PRESERVING IN DATAMINING

Privacy Preservation is a most important area in data mining. Privacy Preservation data mining techniques [1][15] are must use to protect the user's private data

CSEIT183124 | Received : 05 April 2018| Accepted: 20 April 2018 | March-April-2018 [(3)4: 1291-1296]

from unauthorized person. The Privacy Preserving Data Mining Techniques considers the two privacy models like Real and Ideal Model show the figure 2.1. Real Model, which means that Parties run a real protocol with no trusted help and Ideal Model, which means that Parties send inputs to a trusted party, who computes the function for them. Privacy Preserving Data Mining Techniques can be considered in two aspects: Protecting sensitive data values and protecting confidential knowledge in data.



(a) Real Model

(b) Ideal Model

Figure 1: Privacy Preserving Data Mining Models

Secure data sharing, various approaches is used to design the different privacy preserving models. The Privacy-preserving data mining (PPDM), which mainly considered four categories of model parties characteristics like Trusted Third Party Model, Semi - Honest Model, Malicious Model and Other Models – Incentive Compatibility. The Trust Third Party Model, which means that the function only deliver the results of the computation. Semi-honest Model that means every party acts as a semi-honest one. Malicious Model, which means that the participating parties are may or may not follow the privacy preserving rules and Other Models – Incentive Compatibility means that the model motivated the parties to provide the truthful input data. Data mining includes various types of privacy preservation techniques which were classified based on the following dimensions:

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

The first dimension refers to the distribution of data. Some of the privacy preserving approaches have been developed for centralized data and others refer to a distributed data scenario. The second dimension refers to the modification scheme of the data. [3]Data modification is mostly used in order to change the original values of a database that needs to be released to the public and also to ensure high protection of the privacy data.

The third dimension refers to the data mining algorithm. It has been included the problem of hiding data from a combination of data mining algorithms like Association Rule mining [16], Decision tree algorithm and Clustering algorithm[2]. The fourth dimension refers to hiding the rule or the data. The last dimension which is the most important and its refers to the privacy preservation technique used for the selective modification of the data. The important techniques of privacy preserving data mining are Randomization method, Association rule mining and Encryption method.

III. NON – COOPERATIVE INCENTIVE COMPATIBLE MODEL

The main objective of the paper is to maintain the confidentiality of the data in secure transactions [24]. Various types of web service models are used in private data transformation applications. These types of security models are based on the various Privacy Preserving Data Mining (PPDM) techniques such as Randomization method, Secure Code Computation process and Encryption method [6] [9]. In most of the applications, certain PPDA techniques guarantee that nothing other than the final analysis result is exposed, it is impossible to verify whether participating parties are truthful about their private input data in data transformation.

The Incentive Compatible Model has been developed that to provocation the participating parties provide truthful input data [13]. The incentive compatible privacy preserving model has to interact with the participating parties to verify the transaction making use of the user's knowledge. The EShopping is a service oriented application, which provides a user interaction interface that provides more security for individual details transformation compared with the other privacy preserving models.

The figure 3.1 denotes the architecture of the privacy preserving system model. The architecture of a privacy preserving system gives the detailed explanation about the process of the security system in which it allows only the authorized person not others. Suppose, if any fraud user is trying to access the data security system will not allow the user and also the access will be denied for the particular user. Then the appropriate data are retrieved from the database according to the request given by the user.

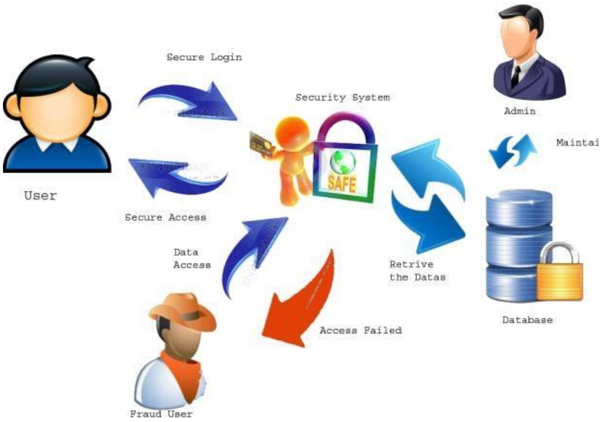


Figure 2: Privacy Preserving System Architecture

The non-cooperative computation (NCC) model is the each party participates in a protocol to learn the output of some, given a function f over the joint data inputs of the parties. In the first step, all participating parties send their private inputs data to a trusted third party (TTP) for secure sharing. And then the second step is TTP computes the function f and sends back the result to every participating party. The Non Cooperative Incentive Compatible model makes the following assumptions:

Correctness: the every participating party is to learn the correct result;
Exclusiveness: the every participating party prefers to learn the correct result exclusively.

Under the correctness and exclusiveness assumptions, the Non Cooperative Incentive Compatible 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 protocol does not limit its generality. The incentive compatible Privacy Model function f over the joint inputs of the parties specified, that is derived from the secure code computation process. Analyze what types of distributed functionalities could be implemented in an incentive compatible fashion. In other words, to 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 computation process considers the following privacy preserving data mining techniques:

A. Horizontal partitioning for the table means to extract the fields that performing extract the collection of incentive data's according to particular customers. Perform these tasks to take the user's private detail as an input to the NCC model. Build the DNCC model with help of Association rules technique and build the Probabilistic NCC model with the help of Randomization technique that provides the incentive data randomly.

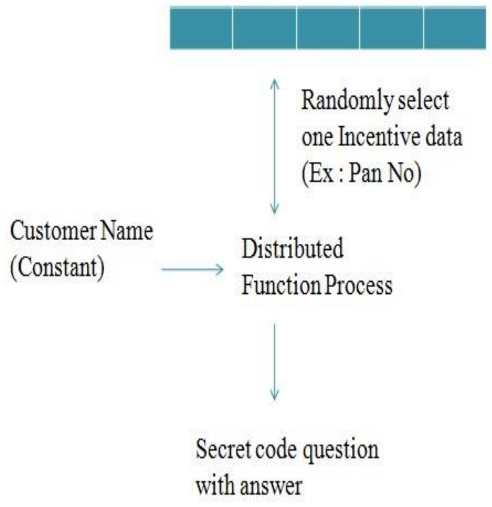

Randomly select one Incentive data (Ex : Pan No)
Customer Name (Constant)
Distributed Function Process
Secret code question with answer

Figure 3: Randomized Incentive Compatible Model – Secure Code Computation Process

B. Secure code computation process is providing incentive compatible secret code question for the Non – Cooperative Privacy Model. The figure 3.2 represents the process of secure code computation for privacy preserving randomized incentive model. The computation process theorem consists of following steps:

Step 1: Select two attributes from customer details from bank database as inputs for the distributed function.
Step 2: Here first attribute is constant and another one attribute is other personal details of customer information that is selected by random.
Step 3: Applied Privacy Preserving techniques such as partitioning data, secure sum and dot protect operation on the selected two attributes.

SMC technology used in distributed privacy preserving data mining areas that mainly consists of a set of secure sub-protocols, such as, secure sum, secure intersection, secure set union, secure comparison, dot product protocol and so on. In the following will be briefly describe the basic idea of two kinds of secure sub-protocols used in horizontally partitioned and vertically partitioned setting.

C. Secure Sum: Secure Sum can securely calculate the sum of the values from different site data. Assume that each site i has some value v_i and all sites want to securely compute $S=v_1+v_2+, \ldots, +v_n$, where v_i is known to be in the range [0.. m]. For example, in horizontally partitioned association rule mining setting can securely calculate the global support count of an item set by the secure sum sub-protocol.

D. Dot Product Protocol: At present, many secure dot product protocols have been proposed. The problem can be defined as follows: Alice has a n-dimensional vector $X=(x_1, x_2,\ldots, x_n)$, while Bob has a n-dimensional vector $Y=(y_1, y_2, \ldots, y_n)$. At the end of the protocol, Alice should get $r_a=X.Y+r_b$ where r_b is a random number chosen from uniform distribution that is known only to Bob, and $X.Y=x_1 y_1+x_2 y_{2+}, \ldots, +x_n y_n$. For Example, using the dot product protocol we can securely calculate the global support count of an item set whose items are located at different sites in vertically partitioned setting.

E. Privacy Preserving Data Publishing: The incentive compatible model is only concentrating on the secure data sharing process and does not consider the data storage publishing. All the users' private information's are stored in the particular database that is more securable one. Using a symmetric encryption algorithm, Triple Des algorithm is used to encrypt the all the users sensitive information in the secure database. The Triple – DES algorithm is a Private Key Algorithm and using the Block cipher with (Electronic Code Book) ECB mode, because (Cipher Block Chaining) CBC mode requires more processing time to compare with ECB mode. The

Electronic Codebook Mode is the basic form of clock cipher where data blocks are encrypted directly to generate its correspondent ciphered blocks that is denoted by the Fig 3.3. The CBC mode requires more processing time to compare with ECB mode.

For Triple-DES Encryption algorithm,

```
**Encrypt** (string strToEncrypt, string strKey)
        {
            ciphertext = $E_{K3}$ ($D_{K2}$ ($E_{K1}$ (plaintext)))
        }
```

For Triple-DES Decryption algorithm,

```
**Decrypt** (string strEncrypted, string strKey)

        {
            plaintext = $D_{K1}$ ($E_{K2}$ ($D_{K3}$ (ciphertext)))
        }
```

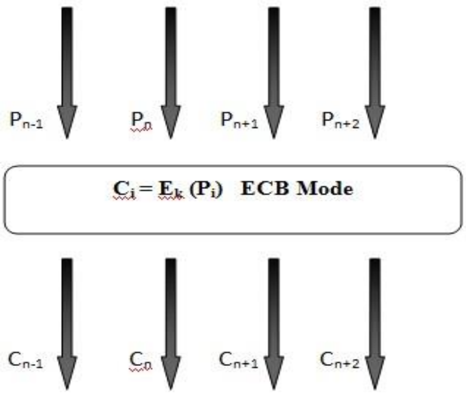


Fig 4: **Working Process for Block Cipher ECB Mode**

The advantages of Triple – DES algorithm is, simple Encryption method, Best public cryptanalysis and required less processing power compare with other encryption algorithms.

Compatible Non Cooperative Computation model in which it maximize the secure data sharing during the information transformation. The privacy preserving data analysis tasks using incentive compatibility under NCC model. The main advantage of this model is that to reduce the number of False Positive transactions. It tries to find any anomalies transaction based on the data analysis model.

IV. CONCLUSION

The probabilistic non cooperative computation model that includes the privacy preserving data mining techniques such as horizontal partitioning, randomization method, association rules classification and secure multiparty computation process. The existing DNCC model, works under the principle of incentive data in some order which can be identified by the anomalies sometimes. To avoid this type of problem, the incentive data's are selected by probabilistically under PNCC Model. The incentive compatible model is helpful for more privacy preserving application approaches that are used to interact with user original knowledge. In future to provide more than two attribute based more securable privacy preserving model can be built. And also using various privacy preserving techniques the efficiency of the privacy preserving model can be improved.

The incentive compatible privacy-preserving data analysis technique has been developed to motivate the participating parties to provide truthful inputs. The privacy preserving data analysis task that provides a new model called Incentive.

V. REFERENCES

[1] 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.

[2] M. Kantarcioglu and O. Kardes, "Privacy-Preserving Data Mining in the Malicious Model," Int'l J. Information and Computer Security, vol. 2, pp. 353-375, Jan. 2009.

[3] M. Kantarcioglu, C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge and Data Engineering, Vol. 16,No. 9, pp. 1026-1037, 2004.

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

[5] Rakesh Agrawal and Evimaria Terzi , "On Honesty in Sovereign Information Sharing," Proc. Int'l Conf. Advances in Database Technology, pp. 240-256, 2006.