IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Application of Secure Multiparty Computation in Privacy Preserving Data Mining

Karan Saxena [1], Saurabh Satpute[2], Aditya Gupta[3], Varun G[4]

B.Tech Student, Computer Science and Engineering, SRM University, Chennai, India[1, 2, 3, 4]

Abstract: Privacy preserving data mining is an area of research concerned with the issues of privacy thus providing a solution to minimize privacy threats in data mining. PPDM also helps in maximizing analysis outcome and also helps in minimizing the disclosure of individual or organizational private data. The existing system comprises of several privacy preserving techniques but all these techniques lacked in the parameters of Input privacy and correctness. With the use of Secure Multiparty Computation (SMC) more focus is given on the parameters of Input privacy and correctness with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. SMC helps in performing global computations on the private data with the help of several trusted third parties (TTP) so that there is no loss on data and privacy is maintained. The main aim is to implement this technique of SMC in the online transaction processes so as to make the transactions happening across the world as safe and secure as possible. The overall online transaction system developed must be user friendly and the privacy or confidentiality of the users have to be preserved so that in the near future the users taking part in the process do not hesitate in providing their details and the confidentiality for each user detail is maintained.

Keywords: Secure Multiparty Computation(SMC), trusted third party (TTP), PPDM.

I. INTRODUCTION

Secure multiparty computation provides numerous possibilities for collaborative and joint computations. SMC is a privacy preserving data mining technique which helps in performing joint computations in several networked environment. It also helps in providing computations among several diverse organizations in a safe and secure manner. With SMC, global computation can be performed on the private data of several participating parties so that there is no loss on data and privacy is maintained. End-to-end secure multiparty protocol development can also be implemented through SMC.

With the increase in use of data mining tools in both the public and private sectors raises concerns about the potentially sensitive nature of the data which is being mined. All the utility which is being gained from widespread data mining seems to come in direct conflict with an individual's need and right to privacy. Privacy preserving data mining solutions helps us to achieve the paradoxical property of enabling a data mining algorithm so as to use data without ever actually "seeing" it.

A. SMC

Secure multi-party computation (also known as secure computation or multi-party computation/MPC) is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. [1]

B. Trusted third party (TTP)

A trusted third party (TTP) is an entity which facilitates interactions between two parties who both trust the third party; the Third Party reviews all critical transaction communications between the parties, based on the ease of creating fraudulent digital content. In TTP models, the relying parties use this trust to secure their own interactions. TTPs are common in any number of commercial transactions and in cryptographic digital transactions as well as cryptographic protocols, for example, a certificate authority (CA) would issue a digital identity certificate to one of the two parties in the next example. The CA then becomes the Trusted-Third-Party to that certificates issuance. Likewise transactions that need a third party recordation would also need a third-party repository service of some kind or another. [2]

C. PPDM

Privacy preserving data mining (PPDM) is a novel research direction in Data Mining (DM), where DM algorithms are analysed for the side-effects they incur in data privacy. The main objective of PPDM is to develop algorithms for modifying the original data in some way, so that the private data and private knowledge remain private even after the mining process. [3]

LITERATURE SURVEY

The aim of a secure multiparty computation task is for all the participating parties to securely compute some function on their inputs which are either distributed or private. One of the key question that arises here is what do you mean by secure computation? One way to approach this particular question is by collecting the list of security properties that should be preserved. The first such property that comes to mind is the property of privacy or confidentiality. A naïve attempt to formalize privacy can be done by learning about each individual

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

party and check whether its data is malicious or not. However how confidential a data may be the defined output of the computation typically reveals some information on other parties' inputs. Therefore, the privacy requirement is always formalized by saying that the only information learned by the participating parties in the computation (again, even by those who behave maliciously) is that specified by the function output. Although privacy is a primary property of security, it rarely suffices. Another property that comes to our mind is that of correctness. The correctness states that the parties' output should only be defined by that function (if the correctness is not guaranteed, then a malicious party may be able to receive the specified decision tree while the honest party receives a tree that is modified to provide misleading information). A key question that always arises in the process of defining the security properties is: when will the list of properties be complete? This question is application-dependent and it means that for every new problem, the process of deciding which security properties are required and after the decision is being made then it must be re-valued. Deciding which security properties to be used in our privacy model is a very difficult process and may take years of research on this particular issue.

Yao initially presented the concept of SMC in the form of "Two Party Computations" [4]. Later on this was generalized to multiparty computation problems by Goldreich, who is one of the prominent researcher who contributed a lot to SMC in the form of secure solutions for any functionality [5]. Besides this, Agrawal et al provided fast and secure algorithms for mining association rules [8]; Atallah et al contributed to secure multiparty computation geometry, which are a base for routing and other network related problems [6]. Lindall et al provided cryptographic techniques and solutions for SMC [7]. Rebecca Wright provided some solutions to SMC and Privacy Preserving data mining through its PORTIA project [6]. Several problems and protocols to solve them have also been proposed by various eminent researchers which provide a clear view of SMC, their problems and solutions.

II. BACKGROUND OF SMC

Figure 1 depicts the simple architecture of multiparty computation. The first layer as seen in the figure above is defined as the input layer. The input layer consists of all

The participating parties in order of p1,p2,p3…..p7 and hence combination of all these parties gives us the input layer. As all the parties are interconnected with each other the data packets are distributed among them. One of the parties from the input layer transfers the data packet to layer 2 (Computation layer). The third party in the second layer has no idea which participating party has transferred the data packets as all the parties are interconnected with each other in the first layer. The second layer then collects the data packets and computes it using some pool of functions. The confidentiality of individual parties are always maintained. [9]

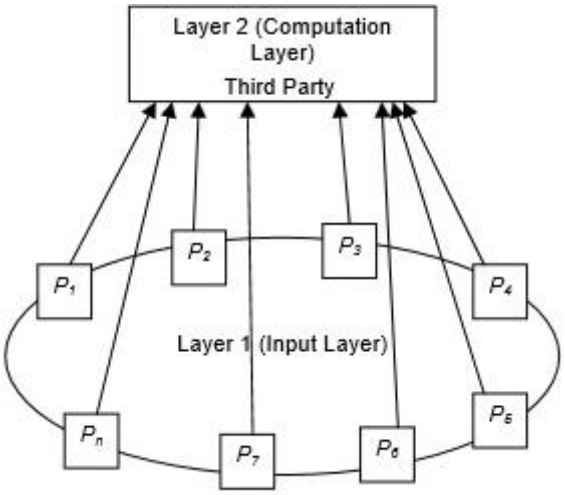

Fig. 1:The Existing architecture

By comparing this architecture with the figure 1 the extended architecture consists of certain modifications. The extended architecture of our protocol. Instead of using a single TTP, the computational layer (2nd layer) now consists of several TTP's. Each TTP consists of same pool of functions that were defined in the earlier architecture. These pool of functions are used to encrypt the data packet so as to maintain the overall privacy .out the many TTP takin part one of the TTP is chosen at run time and all data packets are forwarded to that particular TTP and makes him responsible for putting data packets into data blocks and then can perform some calculations on it.

III. PROPOSED TECHNOLOGY

Secure multiparty allows all the parties which are participating in the computation having similar background to compute the results on their input data. With the increase in the activities of data mining the privacy of the data shared between individual parties or organization are coming under threat and this needs to minimize. Day by day dependency on online transaction is increasing significantly and thus it also leads to a large threat of privacy details getting leaked. Our proposed technology also deals with the online transactions as the whole of world is dependent on online transaction so one cannot ignore this issue. In this paper we have combined the techniques of SMC, Key generation algorithm and Masking techniques so that the online transaction process can work smoothly and people do not hesitate in sharing their details with anyone. All these techniques combined will give us secure methods to make these process as secure as possible for the customers involved in the business of transactions.

A. Key generation

Key generation is a process of generating keys in the cryptographic process. With the use of proper public and private key one can encrypt or decrypt the data which is being encrypted/decrypted. The public key can be made available to anyone sometimes by the means of digital certificate. The user who is sending the data encrypts the

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

data with the public key and only the one who is holding the data can decrypt the data. The simplest method with which encrypted data can be read without actually decrypting is known as a brute force attack—simply going through every number one by one until the maximum length of the number. In this paper Pseudo random number generator which comes under the key generation algorithm has been implemented to generate the random sequences of number.

B. Pseudorandom number generator (PRNG)

A pseudorandom number generator (PRNG) can also be known as the deterministic random bit generator (DRBG), is a frequently used algorithm for generating a random sequences of numbers whose properties approximate the properties of other sequences of random numbers. The numbers generated through PRNG is not truly random, because they are determined by a relatively small set of initial values, called as the PRNG's seed . the PRNG seed contains the value which may be true Although close sequences of random can be can be generated using hardware random number generators. Their speed in generation of numbers and reproducibility are one of the advantages of pseudorandom number generators and this is the reason why they are so widely used than other key generators.

PRNGs are used in wide variety of applications but they are more central in applications such as the procedural generations, simulations and cryptography. Cryptographic applications require that whatever output is generated it should not to be predictable from earlier outputs. Good statistical properties are a must to generate random numbers efficiently and with ease.

1. All the cryptographic primitives of PRNGs must be studied properly so that we get a better understanding of how the PRNGs work and how its limitation can be reduced.
2. Careful selection of good algorithms and protocols must be made as the PRNG can be a single point of failure for many real-world cryptosystems.
3. The PRNGs must be used in a proper way because some system use it in a bad way which only leads to their downfall.

C. Mathematical model

Suppose the number of parties are P1, P2, ..., Pn out of that each party have some private input x1, x2, ...xn and they wish to perform some polynomial-time computation

$$f(x1, x2, ...xn, R) = (y1, y2, ..., yn)$$

Where R gives us the randomness and (y1, y2, ..., yn) are defined as the private output values for each party participating. The protocol (π) for this computation is defined as:

• Correct - π should allow the participating parties to compute f correctly.
• Privacy – for each party P1, P2, ..., Pn, each player's input should always remain private. That is, no one can learn about anyone about their input than can be deduced from the output.
• Output delivery - this protocol doesn't end until everyone receives an output, the output can also define failure.
• Fairness - if one of the party gets the answer, so does everyone else.

The goal is to show that anything that A can learn by running the protocol π, the ideal adversary S can also learn the same thing by interacting with the functionality Definition: If π is an n party protocol. Then π t-securely computes f if:

$$\forall A \exists S \text{ such that } \forall I \subseteq [n] \text{ with } |n| \leq t \; \pi,A,I \approx f,S,I$$

Where I is defined as a set of indices of the corrupted parties taking part. For example, if more than one or two parties are found to be corrupted, then we are not able to achieve all four informal security goals.

IV. SYSTEM ARCHITECTURE

The overall process starts with the user as our project focuses on user based methodology. The user can be of two types one is existing the other one can be new user. If the user who is logging in is an existing type of user then that particular user can get the details from the database which is maintained by the admin. As the logging process continues the user get the detail from the database by putting a request to the admin, the admin acknowledges the request of the particular user and grants them to access to the database where the that particular user details are kept. The user after getting that particular detail logs into the system and proceeds with the work which the user wants to finish. The other type of user are the new ones who have never logged into the system. These user start by entering their details which are saved and stored in the database and any changes to be made hereafter in details of the user have to make by requesting to the admin, if the admin acknowledges the request then only that particular detail of the user can be changed or removed. The admin has full control over the list of all the users taking part in the transaction process and at any moment the admin can retrieve the data about any user that is taking part in the overall transaction process.

The security system helps in establishing secure login for each of the user taking part in the transaction process. Verification of each user takes place through the help of security system. The security system acts as an interface between the users taking part in the process and the admin. The two security questions which were given to the user at time of providing details into the database is verified in this security system and when the two security question are verified then only that particular user is provided with the secure access to overall transaction process. The security system also helps in detecting any fraud entry into the overall process. As soon as the user who wanting to involve itself in the process enters the security system it automatically asks for the two security question which

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

were provided to the user at the start of the detail filling process. When the two question verified by the security system turns out to be wrong the security system flags it as a fraud entry and denies the access to it database and does not provide the user who is entering as a fraud the secure key which being generated by the security system. The next step after detecting the whether the user is fraud is not is providing the trusted users with the secure key through which they can complete the overall transaction which is pending.

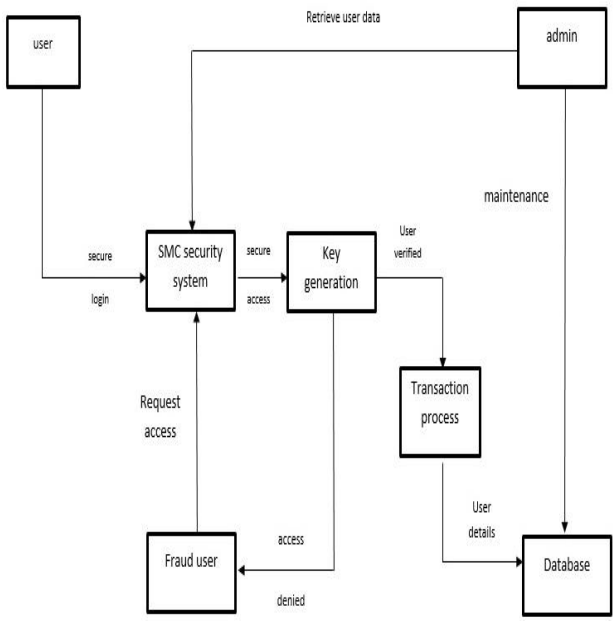

Fig. 2: System Architecture

V. MASKING TECHNIQUES

Masking technique is a process where the original data are transformed in a way so as to produce new data that are valid for statistical analysis and also helps in preserving the confidentiality of respondents. Masking techniques can be further classified as:

A. Non-perturbative

There is no change in the original data but some of the data are suppressed and/or some of the minute details are removed; Non-perturbative techniques helps in providing protected micro data by eliminating some details from the original micro data. Some of the non-perturbative techniques are Sampling, Local suppression, Global recoding, Top coding, Generalization and Bottom-coding.

B. Perturbative

Modification is done on the original data and with the use of perturbative techniques, the micro data table is also modified for publication. Modifications helps in making unique combinations of values in the original table so that original data disappear as well as introduce new combinations. Lossy compression, Rounding, PRAM, MASSC, Resampling, Random noise, Swapping, Micro-aggregation and Rank swapping are some of the perturbative Techniques of masking.

C. Synthetic data generation techniques

The original sets of data in a micro data table is replaced with a new set of data generated in such a way so that the key statistical properties of the original data is preserved .All the generation process is usually based on a statistical properties and the key statistical models that are not included in the model will not be necessarily replaced by the synthetic data. Since the micro data which are released table contains synthetic data, the re-identification risks are minimized. The released micro data table can be mixed with the original data or entirely synthetic.

VII. CONCLUSION AND FUTURE WORKS

This paper brings the concept of SMC and their solutions to light such as database queries, intrusion detection, geometric computation, Scientific Statistical Computation. Researchers are still underway to get efficient solutions for all the SMC problems and as the scope of SMC gets wider and wider, this area will gain a lot of interest and attention. With widespread use of computers in today's world, proliferation of the sensitive and private data is considered as a very important thing. The aim is to make good use of SMC and make the transaction process as safe and secure as possible. In the future work the implementation of FKN protocol will be undertaken. After generating the shared random coins, each party sends a single message to the server receives the encoding of the output created. The parties then individually use their local coins to recover their outputs. For P1 and P2, this protocol is significantly more efficient than using a standard secure two-party computation protocol. None of the parties (including the server) have to perform public-key operations, except performing the coin-tossing which is only performed once. This is in contrast to standard MPC where public-key operations are a necessary.

REFERENCES

[1] https://en.wikipedia.org/wiki/Secure_multi-party_computation
[2] https://en.wikipedia.org/wiki/Trusted_third_party.
[3] AlbertoTrombetta and Wei Jiang (2011), 'Privacy-Preserving Updates to Anonymous and Confidential Databases', IEEE Transactions on Knowledge and Data Engineering, Vol. 22, pp. 578-568.
[4] Y.C.Yao, "How Generate and Exchange Secrets". In proceedings of the IEEE Symposium on Foundation of Computer Science IEEE, 1986, Pages 162-167.
[5] O.Goldreich, "Secure Multiparty Computation", September 1998 (Working draft) Online available on: http://www.wisdom.weizmann.ac.il/~oded/pp.html.
[6] Wenliang Du and Mikhail J. Atallah,"Secure Multiparty Computation Problems and their Applications: A review and Open Problems," Tech. Report CERIAS Tech Report 2001-51, Center for Education and Research in Information Assurance and Security and Department of Computer Sciences, Purdue University, West Lafayette, IN 47906, 2001.
[7] Y.Lindell and B. Pinkas, "Privacy Preserving Data Mining". In advances in Cryptography-CRYPTO-2000, pp 36-54, SpringerVerlag, August 24 2000.
[8] Dr. Durgesh Kumar Mishra et al /International Journal on Computer Science and Engineering Vol.1(3), 2009, 171-175 "A Glance at Secure Multiparty Computation for Privacy Preserving Data Mining".
[9] U. Maurer, "Secure Multi-Party Computation made Simple." Security in Computational Network (SCN'02), G. Persiano (Ed.), Lecture notes in Computer Science, Springer- Verlag, Vol. 2576, 2003, pp 14-28.