Secure Multiparty Computation during Privacy Preserving Data Mining: Inscrutability Aided Protocol for Indian Healthcare Sector

Zulfa Shaikh [1], Poonam Garg[2]

[1]Faculty of Computer Applications,Acropolis Institute of Technology & Research,Indore, India
shaikh.zulfa@gmail.com
[2]Department of Information Technology
Institute of Management Technology,Ghaziabad, India
pgarg@imt.edu@gmail.com

ABSTRACT

Internet today has put up a great challenge on the security for Indian Healthcare Sector. In today's growing environment, most of the computation is jointly computed involving inputs of all the hospitals. Such computations use confidential data of the involved hospitals to compute the result. Each hospital is having confidential data which they would not like to share with other hospitals. Privacy preservation is of great concern as no hospital can be trusted in real scenario. In this paper we have proposed an efficient protocol for computation. This paper is an extension of our previous work in which we have defined and compared single and multi trusted third party protocol. This paper uses multi trusted third party protocol, in which TTPs are selected at runtime from a pool of TTPs and computation is performed by more than one TTP as TTPs can be corrupted and correctness in computation is a major concern. In this paper we proposed a secure protocol that uses encrypted inputs for computation to maintain privacy of inputs and inscrutablizers to make the identity of hospitals ambiguous. Besides this, security analysis is done for the protocol.

KEYWORDS

Secure Multiparty Computation (SMC), Trusted Third Party (TTP), Single TTP, multi TTP, privacy, security, correctness.

1. INTRODUCTION

Joint computation is a need in fast growing Internet world. Most of the applications work on joint computation where large numbers of parties are involved. These parties send their data for computation to TTPs and the computing TTPs announces the result. The first major concern here

Sundarapandian et al. (Eds): CoNeCo,WiMo, NLP, CRYPSIS, ICAIT, ICDIP, ITCSE, CS & IT 07,
pp. 131–140, 2012. © CS & IT-CSCP 2012 DOI : 10.5121/csit.2012.2412

is to maintain privacy of inputs provided by the parties. Security and correctness in the result of computation is the next parameter which has to be maintained in the protocol. This problem is SMC, where n parties send their private inputs *x1, x2,…xn* to TTPs for computation and TTPs announces the result in form of y. The general Secure Multiparty Computation (SMC) model for Indian Healthcare sector is:

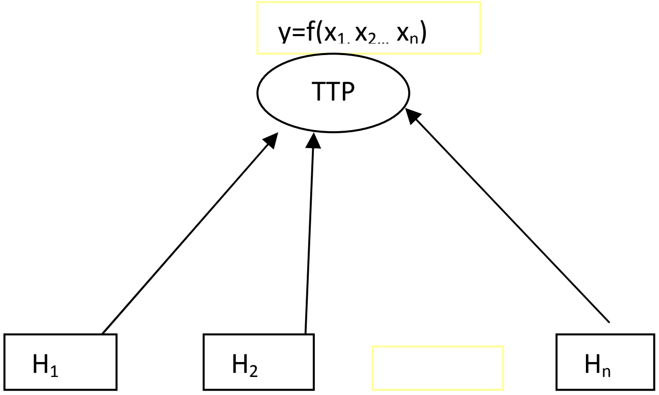

Figure 1. General SMC Model

In figure 1 n hospital send their inputs to TTP for computation. The major concern here is to maintain privacy, security and correctness in the result. Privacy preserving data mining solutions are of great importance in the field of medical research. Consider a case where numbers of different hospitals wish to jointly mine their patient's data for the purpose of medical research and prevention of data is to maintained due to confidentiality of patients record. This problem is an instance of SMC problem. Privacy preserving data mining solutions enable the hospitals to compute desired data mining algorithms on the union of their databases, without ever disclosing the data. The only information learned by different hospitals is the output of data mining algorithm.

1.1 Gaps in the Present Study:

- There is an absence of proper data security and cyber laws which is hindering hospitals and its business prospects. There is also remarkable excitement and a lack of understanding of the problems surrounding security. The most significant security issue is the protection of data. Some of the information security and data privacy challenges that hospitals face include lack of tough data protection laws, use of portable devices such as laptops by employees to store confidential information, rising data security costs due to increased employee background checks, training employees in maintaining data security, ensuring compliance with security policies implemented in the hospitals, and problems through employee activity monitoring procedures.

- To maintain the confidentiality of patients or hospitals information, there is need to implement data security and privacy procedures.

2. BACKGROUND

SMC problem is the problem of n parties to compute a private function of their inputs in a secure method, where security means the correct result computed by the TTPs for maintaining the privacy of the parties as some of the parties may want to misuse the other party's data. We assume that the inputs are *x1,x2,…xn* where xi is the data of party Pi and the TTP will compute a function f(x1,x2,….xn)=y and announce the result y [1]. Security is meant to achieve correctness of the result of computation and keeping the party's input private even if some of the parties are corrupted. In figure 1, trusted third party is used for doing the computation on the inputs provided by the parties. According to [2], the major problem with this approach is that it is difficult to find the third party which is trusted by all the parties providing the inputs and to control the function of adversaries.

The concept of SMC began in 1982 when Yao proposed and gave solution to his millionaire's problem in which two millionaires wanted to know who was richer without disclosing individual wealth to each other [3]. It was a two-party computation protocol for semi honest parties who follow the protocol but also try to know something other than the result. The idea was extended to multiparty computation by many researchers. Goldreich et al. showed the existence of a secure solution of SMC problem. The size of the protocol depends on the number of parties involved in the computation process. [4].They used circuit evaluation protocols for secure computation. Earlier research focused on theoretical studies. Later, some real life applications emerged like Private Information Retrieval (PIR) [5, 6], Privacy-preserving data mining [7, 8], Privacy-preserving geometric computation [9], Privacy-preserving scientific computation [10], Privacy preserving statistical analysis [11] etc. A detailed review of SMC research is provided by Du et al. in [12] where they developed a framework for problem discovery and converting normal problem to SMC problem. A review of SMC with special focus on telecommunication systems is given by Oleshchuk et al. in [13].

Yao's original protocol considered only the case of semi-honest parties; an extension to the case of malicious party was given by Lindell [14]. Ronald Cramer provided the theoretical discussion of complexity constraints on secure multiparty protocols, specifically for the secret sharing problem [15]. Ran Canetti identifies flaws in previous multi-party computation work arising from the introduction of adaptive adversaries, who choose to corrupt involved parties dynamically during the computation. The paper introduced the notion of a semi-honest party, who appears to be honest from an outside perspective, but deviates from the protocol in some way. He presented a secure protocol, using a trusted third party, to avoid the adaptive adversarial pitfalls [16].

Aiming at privacy preserving computing of statistical distribution, which is frequently encountered in statistics, and based on the intractability of computing discrete logarithm and using rigorous logic, they proposed the solution. [17] Presented the protocols allowing the players to securely solve standard computational problems in linear algebra such as determinant of matrices product, rank of a matrix, and determine similarity between matrices. [18] Presented TASTY, a novel tool for automating, i.e., describing, generating, executing, benchmarking, and comparing, efficient secure two-party computation protocols. They used TASTY to compare protocols for secure multiplication based on homomorphic encryption with those based on garbled circuits and highly efficient multiplication. [19] Presented a hybrid-secure MPC protocol that provides an optimal trade-off between IT robustness and computational privacy.

[20] Presented a solution to the Secure Multi-party Computation (SMC) problem in the form of a protocol that ensures zero-hacking. The solution comprises of a protocol with several trusted third parties (TTPs).The protocol selects one TTP among all TTPs in the SMC architecture that owns the responsibility of all the computation in the system. This TTP is called the master TTP and it is different at different times. The procedure of selecting master TTP could be non-deterministic but it is made deterministic by randomization technique. This ensures that no single TTP controls the entire system all the times. At the same time, this also ensures that no TTP knows where the computation is taking place. This approach is having merit over the other one where only one TTP is given the responsibility to hold entire data of the system.

Gaps in [20] are multiple TTPs are given the input but computation is performed by master TTP only. So even if this protocol defines multi TTP environment, but efficiency of TTP in the protocol is not utilized properly. The second gap in [20] is introducing packet layer. The responsibility of packet layer could be handled by parties itself and a virtual party can be used to make identity of party ambiguous. In our multi TTP computation protocol, same computation is performed by multiple TTPs selected at runtime and majority giving the same identical result is considered the right result of computation. Efficient SMC_Multi TTP algorithm designing was our previous work.

3. PROPOSED WORK

3.1 Informal description of the protocol

In this protocol all the hospitals involved in computation split their data into x packets and encrypt data through some pre-decided encryption method. The encrypted data E_{ij} is send to inscrutablizers. This is an untrusted layer whose task is to forward the packets to TTPs selected at runtime for computation. Inscrutablizers cannot store the data, they just forward it. As inscrutablizers are untrusted, so they hold packets of the parties and not the entire data. After computation majority of TTPs giving the same result is considered as the right result of computation as correctness is a major parameter for computation which has been analyzed in previous work.

3 layer architectural framework:

1. n hospitals : H1, H2…Hn with data packets xij

2. Inscrutable layer

3. Multiple TTP layer: TTP1, TTP2…TTPn

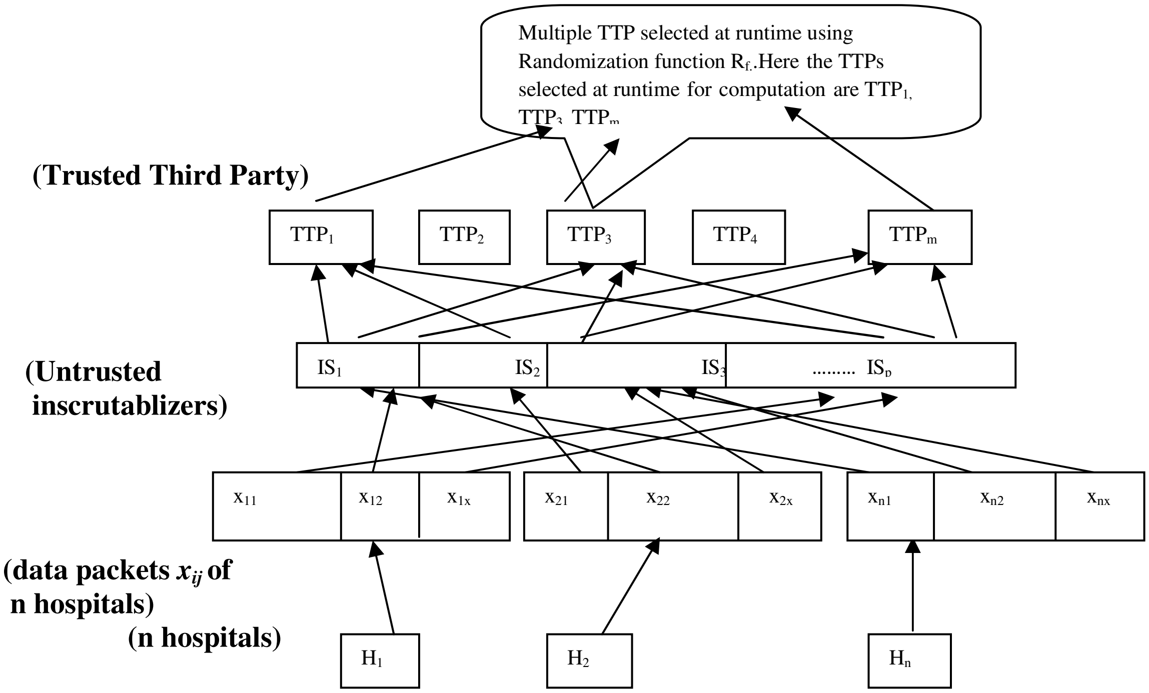

Figure 2. Three layer architectural SMC framework for hospitals using Multi TTP computation

3.2 Formal description

Algorithm: SMC_Split Multi TTP computation

Data Structure

H_i – Hospitals where i ranges from 1 to n

x_{ij} – Data of party Hi where j ranges from 1 to x

R_{ij} – Random data of party Hi where j ranges from 1 to q

D_{ij} – total data including the random and the original data

E_{ij} – Encrypted data associated with party Hi where j ranges from 1 to $x+q$

IS_p– untrusted inscrutablizers, where p ranges from 1 to z

TTP – third party

Algorithm:

- Generate x_{ij} packets for every party *Hi*
- Generate random data *Rij* for every x_{ij}
- Group random data *Rij* with original data x_{ij} to get D_{ij}
- Encrypt data D_{ij} using pre-decided encryption method to get E_{ij}.
- Distribute the encrypted data E_{ij} among the inscrutablizer A_p
- Send the data from un-trusted inscrutablizer A_p to *TTPs*
- Calculate the result at *TTPs* using the encrypted data and the keys.
- Identify the TTPs at runtime for performing computation.
- The result is announced by TTPs
- Majority of TTPs giving same identical result is considered as correct result.

4 layer architectural framework:

1. n hospitals : H1, H2…Hn with data packets xij
2. Untrusted Inscrutable layer (Virtual Party)
3. Trusted Inscrutable layer (Packets are distributed))
4. Multiple TTP layer: TTP1, TTP2…TTPn

The advantage of designing four layer architectural frameworks is to increase the security level of inputs provided by the hospitals. In this framework an untrusted inscrutable layer is added to hide the identities of the hospitals. This layer is inscrutablizer layer. The data from this layer is then send to trusted inscrutablizers who does not have any knowledge about input of the hospitals as the data arrives from virtual layer. In this protocol all the hospitals involved in computation split their data into x packets and encrypt data through some pre-decided encryption method. The encrypted data E_{ij} is send to untrusted inscrutablizers. This is an untrusted layer whose task is to forward the packets to trusted inscrutablizers and then they forward packets to TTPs selected at runtime for computation. Inscrutablizers cannot store the data. Inscrutablizers hold packets of the parties and not the entire data for security and privacy of inputs. After computation majority of TTPs giving the same result is considered as the right result of computation as correctness is a major parameter for computation which has been analyzed in previous work.

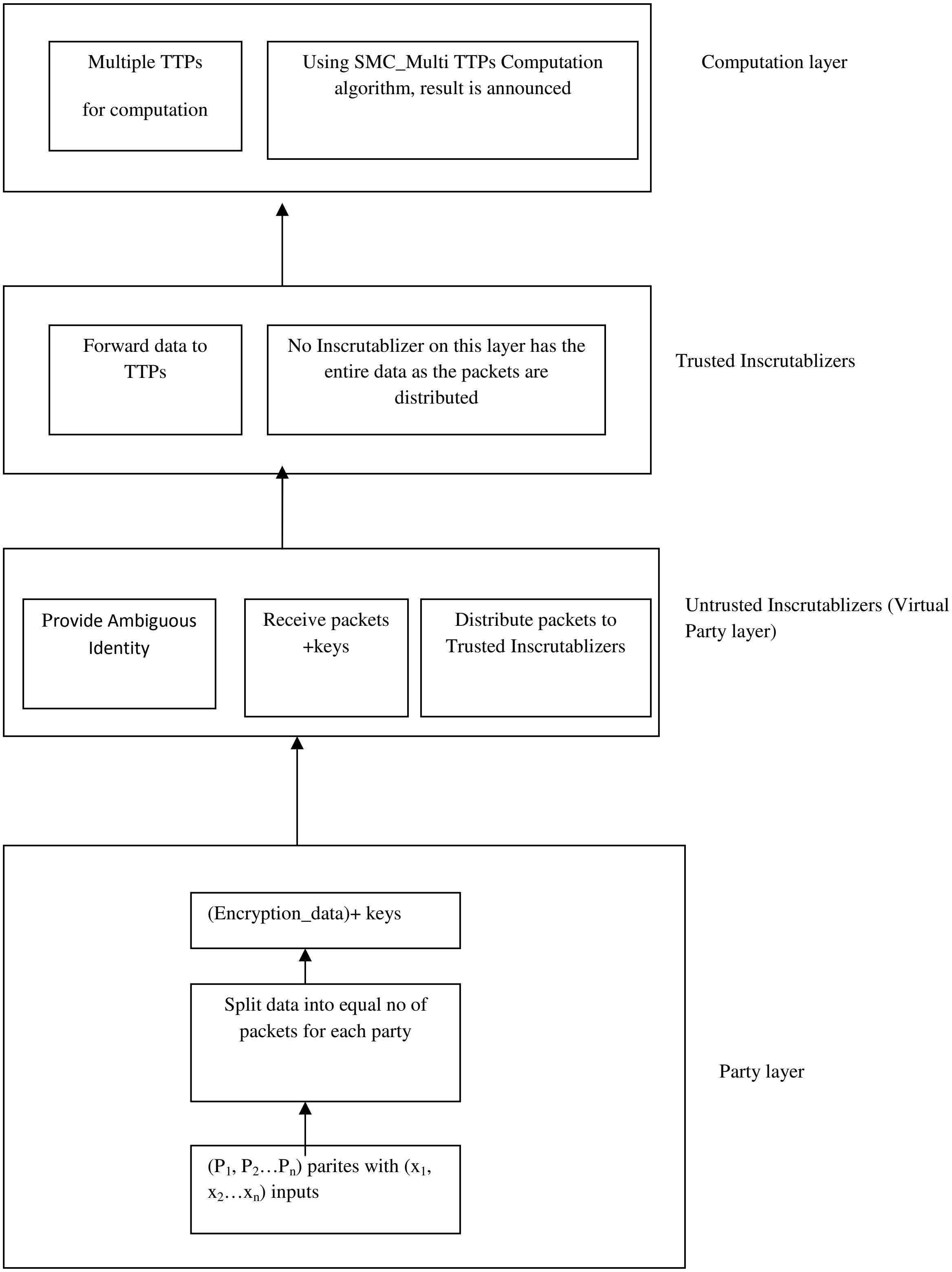
Multiple TTPs
for computation
Using SMC_Multi TTPs Computation algorithm, result is announced
Computation layer
Forward data to TTPs
No Inscrutablizer on this layer has the entire data as the packets are distributed
Trusted Inscrutablizers
Provide Ambiguous Identity
Receive packets +keys
Distribute packets to Trusted Inscrutablizers
Untrusted Inscrutablizers (Virtual Party layer)
(Encryption_data)+ keys
Split data into equal no of packets for each party
(P_1, P_2…P_n) parites with (x_1, x_2…x_n) inputs
Party layer

Figure 3. Responsibiites of each layer of SMC Protocol designed for Indian Healthcare Sect

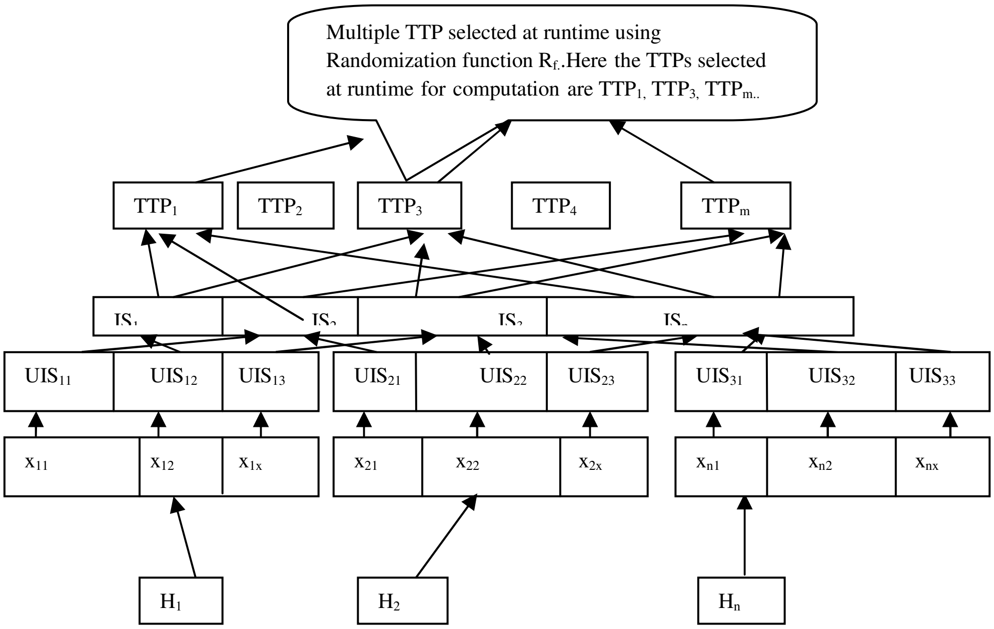
Multiple TTP selected at runtime using Randomization function Rf..Here the TTPs selected at runtime for computation are TTP1, TTP3, TTPm..
TTP1
TTP2
TTP3
TTP4
TTPm
IS1
IS2
IS3
ISp
UIS11
UIS12
UIS13
UIS21
UIS22
UIS23
UIS31
UIS32
UIS33
x11
x12
x1x
x21
x22
x2x
xn1
xn2
xnx
H1
H2
Hn

Figure 4. Four Layer architectural SMC framework for hospitals using Multi TTP computation

3.3 Security Analysis

If the TTP is malicious then it can reveal the identity of the source of data. A set of inscrutablizers from the inscrutable layer will make the source of data ambiguous and will preserve the privacy of individual. The more the number of inscrutablizers in the inscrutable layer the less will be the possibility of hacking the privacy of the data. The inscrutablizers hide the identity of the bank. In the protocol there is one layer of inscrutablizers, consisting of *p* inscrutablizers $IS_1, IS_2, IS_3\ldots, IS_p$. Then the probability of revealing the source of the data at TTP is inversely proportional to the number of parties sending data. We can see that there is more security when there are large numbers of participants.

The probability of hacking the data of a single hospital H_i:

$$\mathrm{P(Hi)} = 1/\mathrm{n} \qquad (1)$$

where n is total number of hospitals involved in computation.

The probability of hacking data of r hospitals:

$$\mathrm{P}(\mathrm{H}_r)=\mathrm{r/n} \qquad (2)$$

Therefore, total Probability for leak of the packets

$$= [r/n] * [{}^{r=1}\textstyle\sum_r X_r)/ ({}^{r=1}\textstyle\sum_n X_r] \qquad (3)$$

where Xr are the packets of r hospitals.

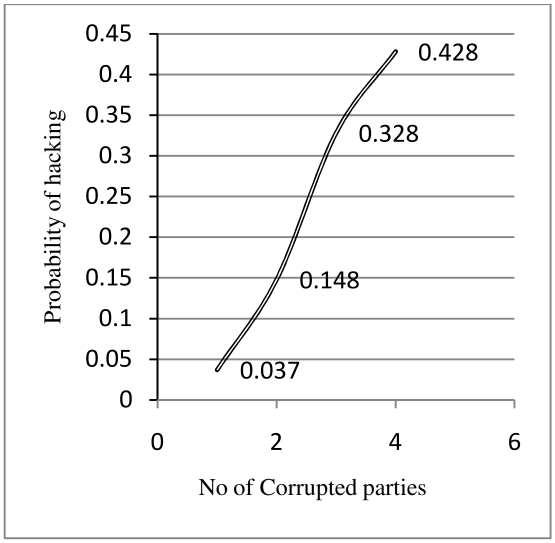

Figure 5. Security analysis with increased number of corrupted parties

4. RESULTS AND CONCLUSION

We know that with the jump into the 21st century, the expansion of the World Wide Web has soared Indian healthcare sector to a new level. Tremendous opportunities for joint transactions have arisen in which multiple hospitals cooperatively conduct some computation. Security and privacy preserving measures are major issues during computation as to maintain confidentiality of inputs. We proposed a secure protocol for multi-party computations in which privacy of individual is preserved.

In this paper the following conclusions are drawn:

- Use of multiple TTPs for computation makes correctness parameter to be more truthful.
- Providing encrypted inputs to inscrutablizers makes the protocol secured.
- Use of inscrutablizer layer makes the identity of hospitals ambiguous as privacy of inputs is a major concern.
- Splitting of inputs into packets and random distribution of it to inscrutablizers increases the security parameter.
- Probability of malicious conduct increases exponentially with increased number of corrupted parties.
- The attractiveness of this protocol is: Multiple TTPs are performing the same computation and announce the result. With the help of such computation, malicious TTP can be easily traced after several round of computation.

Using this protocol and algorithm a wide variety of computations can be optimally performed with enhanced security and privacy.

REFERENCES

[1] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin and Y. Michael. (2002), Tools for privacy preserving distributed data mining , SIGKDD Explorations Volume – 4,Issue – 2, 1-8.

[2] J. Vaidya, and Chris Clifton. (2003), Leveraging the Multi in Secure Multi-Party Computation, in the proceeding of the 2003 ACM workshop on privacy in electronic society, ACM Press.

[3] A.C.Yao. (1982), Protocol for secure computations, in Proc. 23rd IEEE Symposium on the Foundation of Computer Science (FOCS), IEEE, 160-164.

[4] O. Goldreich, S. Micali, and A. Wigderson. (1987), How to play any mental game, in STOC '87: Proceedings of the nineteenth annual ACM conference on Theory of computing, New York, NY, USA: ACM, 218-229.

[5] B.Chor and N.Gilbao. (1997), Computationally Private Information Retrieval (Extended Abstract), in proceedings of 29th annual ACM Symposium on Theory of Computing, El Paso, TX USA.

[6] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. (1995), Private Information Retrieval, in proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science, Milwaukee WI, 41-50.

[7] Y. Lindell and B. Pinkas. (2000), Privacy preserving data mining, in advances in cryptography-Crypto2000, lecture notes in computer science, vol. 1880,2000.

[8] R. Agrawal and R. Srikant. (2000), Privacy-Preserving Data Mining, in proceedings of the 2000 ACM SIGMOD on management of data, Dallas, TX USA, 439-450.

[9] M. J. Atallah and W. Du. (2001), Secure Multiparty Computational Geometry, in proceedings of Seventh International Workshop on Algorithms and Data Structures(WADS2001), Providence, Rhode Island, USA,165-179.

[10] W. Du and M.J. Atallah. (2001), Privacy-Preserving Cooperative Scientific Computations, in 14th IEEE Computer Security Foundations Workshop, Nova Scotia, Canada, pages 273-282, Jun. 11-13 2001.

[11] W. Du and M.J.Atallah. (2001), Privacy-Preserving Statistical Analysis, in proceedings of the 17th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA, 102-110.

[12] W. Du and M.J. Atallah. (2001), Secure Multiparty Computation Problems and Their Applications: A Review and Open Problems, in proceedings of new security paradigm workshop, Cloudcroft, New Maxico, USA, 11-20.

[13] V. Oleshchuk, and V. Zadorozhny. (2007), Secure Multi-Party Computations and Privacy Preservation: Results and Open Problems, Telektronikk: Telenor's Journal of Technology, vol. 103, no.2.

[14] Y. Lindell and B. Pinkas. (2000), Privacy preserving data mining, in advances in cryptography-Crypto2000, lecture notes in computer science, vol. 1880.

[15] R. Cramer, I. Damgard, and S. Dziembowski. (2000), Complexity of verifiable secret sharing and multiparty computation, in Proceedings of the thirty-second annual ACM symposium on Theory of computing, 325-334.

[16] R. Canetti, U. Feige, O. Goldreich, and M. Naor. (1996), Adaptively secure multi-party computation, in Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, 639-648.

[17] Q. Zheng, S. Shan Luo, Y. Xin. (2010), Research on the Secure Multi-Party Computation of some Linear Algebra Problems, Applied Mechanics and Materials, Trans Tech Publication, Switzerland, Vols. 20-23, 265-270.

[18] W. Henecka, S.K. Ogl. (2010), TASTY: tool for automating secure two-party computations, in the Proceedings of the 17th ACM conference on Computer and Communications Security.

[19] C. Lucas, D. Raub, U. Maurer. (2010), Hybrid-secure MPC: trading information-theoretic robustness for computational privacy, PODC '10 Proceeding of the 29th ACM SIGACT-SIGOPS symposium on Principles of distributed computing.

[20] Mishra,D.K.,Chandwani,M. (2008),A zero-hacking protocol for secure multiparty computation using multiple TTP , TENCON 2008 - 2008 IEEE Region 10 Conference.