ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Effective and Efficient Performance of Session Based Security Multiparty Collaborative Data Mining

J Bhuvana*
Department of Computer Application,
Bharathiar University,
Coimbatore, India
bhuvi.jayabalan@gmail.com

T Devi
Department of Computer Application,
Bharathiar University,
Coimbatore-46, India

M Balamurugan
Department of Information Technology,
Selvam College of Technology,
Namakkal-03, India

***Abstract*:** With the advancements in E-commerce and E-governance where more personal data gets exchanged over internet, data privacy has become a key issue. Data mining is often defined as the process of discovering meaningful, new correlation patterns and trends through non-trivial extraction of implicit, previously unknown information from large amount of data stored in repositories and data mining is often carried out on internet based data and there is a need to protect private knowledge during data mining process. The process of preserving private knowledge during data mining is called as privacy preserving data mining (PPDM) [Yasien A.H, 2007]. The privacy preserving data mining is designed to fill up the gap between data mining and data confidentiality. When common users are involved in data mining, all users need to send their data to trusted common centre to conduct the mining; however, in situations with privacy concerns, it is very difficult for a user to trust the other users such a situation is called privacy preserving collaborative data mining. The proposed privacy preservation research work follows secure multiparty computation where multiple parties collaboratively get valid data mining results while disclosing no private data to each other or to any party who is not involved in the collaborative computations. The experiments are conducted to evaluate the privacy performance in terms of share participant, adversary resistance rate, communication round, number of participants and the session participants. The results clearly show that an improvement of nearly 12% of communication execution efficiency and 10% adversary resistance rate to the share participant effectiveness when compared to that of classical key models. This paper describes multiparty computing, (privacy preservation in collaborative data mining) cryptographic method for secured computation. The efficiency is achieved in terms of obtaining combined mining results from the participant private databases without disclosing individual's private data.

***Keywords*:** Privacy Preserving Data Mining, Privacy Preserving Collaborative Data Mining, Secure Multiparty Computation

I. INTRODUCTION

The trusted partners share information between themselves, by maintaining their privacy, with the secure cooperative computations. The share information sent by the participant to a remote database also contains privacy data inferences which should not be disclosed to the receiving participant. In order to maintain privacy of individual's, when sharing data in public domain, needs to be secured. Few constraints and computations are required to maintain in the public domain during the information sharing happen between the participants.

One must know inputs from all the participants to conduct the constrained based security computations. However if nobody can be trusted enough to know all the inputs, privacy will become a primary concern. Secure Multiparty Computation (SMC) based on cryptographic functionality plays a major role in the context of privacy preserving data between different participants in sharing authorized data. SMC is a computational system in which the value based on individually held secret bits of information that compute multiple parties wish to jointly.

With exponential increase in usage of public network information sharing (internet), need for cooperative computation becomes more demanding along with the security of the user's private data. These secure cooperative computations could occur between trusted partners, between partially trusted partners, or even between competitors. Customers might send to a remote database the queries that contain private information, two competing financial organizations might jointly invest in a project that must satisfy both organizations' private and valuable constraints, and so on. Usually, to conduct these computations, one must know inputs from all the participants, however if nobody can be trusted enough to know all the inputs, privacy will become a primary concern. Secure Multiparty computation (SMC) based on cryptographic functionality is initially suggested by Andrew C. Yao[1]. SMC is a computational system in which multiple parties wish to jointly compute some value based on individually held secret bits of information, but do not wish to reveal their secrets to one another in the process. Secure multiparty computation provides solutions to various real-life problems such as distributed voting, private bidding and auctions, sharing of signature or decryption functions, private information retrieval, etc.

The field of privacy has seen rapid advances in recent years because of the increases in the ability to store data. Advances in data mining field along with collaborative mining, 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. The primary motivation for studying methods of secure computation is to design systems that allow for maximum utility of information without compromising user privacy.

Privacy preservation done through role based access control, data perturbation/randomization technique, but the leakage of private information ratio is not healthy. When choosing a cryptographic algorithm, there are many critical questions to ask, such as how the structure should be represented? (ideal/real), which category of adversaries to be focused? (Semi-honest, Malicious), and how can increase the participants in the collaborative task with high confidentiality? [17]. In order to answer these questions, the following special requirements that are to be considered are:

a. Finding a suitable structure for run and use their own secure multiparty protocols without the need for trusted third party. The reliability of trusted third party is doubted and so, uses additional security and other protocol mechanism to enforce greater security.
b. Not all the parties are adversaries, some parties may be adversary, which can be semi-honest (honest but curious) parties or malicious parties (Deviates from the protocol in arbitrary ways).
c. The problem is very challenging, when there is no trusted third party available and parties can misbehave arbitrarily (i.e they can send wrong message or fail to send message at all).
d. The existing algorithms are complex, accuracy, and scalability provides less efficient solutions.

These issues have motivated this paper work. The proposal present in this paper formulate an efficient, privacy preserving collaborative data mining technique and secure multiparty computation method to provide accurate, effective and scalable algorithms that address the challenges of privacy preservation. The proposed multiparty computation presented in this work is based on session framework which comprises of dependent and independent data objects, work cohesively to fetch sharable data in a collaborative data mining. The security context introduced in this work preserves the private data of the involving participant during their distributed computing tasks data, even an internal malicious adversary try to attack the collaborative data.

The secured multiparty computation party process, do not wish to reveal their secrets to another one. Secure multiparty computation provides solutions to various real-life problems such as distributed voting, sharing of signature or decryption functions, private bidding and auctions, private information retrieval, etc. In this paper researchers presented an efficient and effective secured multiparty computation mechanism which preserves the privacy during mining collaborative data mining.

II. LITERATURE REVIEW

The privacy-preserving data mining has two different problem problems, proposed by Lindell and Agrawal. The first problem was that two parties, each having a private database, want to jointly conduct a data mining operation on the union of their two databases. In this scenario they wanted to reveal certain data and hide certain other data. On cooperative mining of two database, there is a possibility of private data of one participant being disclosed to the other. The second problem is that, if the number of participants increased in the cooperative mining, constraints applicable for the cooperative mining goes complex. Lindell and Pinkas find the solution to these two similar problems Lindel [3], [4] used secure multiparty computation protocols, while Agrawal [2] uses the data perturbation method.

The computation method for solving privacy preservation in mining the data by [1] considers only two parties where as in this paper considers multiple participants (number of people/parties) involve in data mining is more than two. The basic idea of SMC[1] is that a secured computation happens during the data mining process with the effect that, no party knows anything except its own input and the resultant data obtained after mining. One way to view this is to imagine a trusted third party; everyone gives their input to the trusted third party, who performs the computation and sends the results to the participants.

The concept of SMC was first introduced by Yao [1] and it has been proved that for any function, there is a SMC solution [13]. The approach used is the constraint function to be computed is first represented as a combinatorial proposition, and then the participants run their private function in their machines on the network of public accessible domain. Every participant gets corresponding input values and output results in concurrence with privacy maintenance of every other participant in the public domain. SMC [8] not only preserves individual privacy, it also preserves leakage of any information other than the final result. However the existing work on the privacy preserving data mining [9] works on less number of participant (10 to 20) with SMC.

In this paper, more number of participants (nearly 50 to 80) involved in the cooperative data mining of privacy data collectively using secured multiparty computation mechanism. In the proposed modified SMC mechanism, new element added are session participants and mine collaborative data of all the participants collective with rule constraints for preserving the privacy of each individual participant.

III. PROBLEM FORMULATION OF PRIVACY PRESERVING COLLABORATIVE DATA

The usage of information without compromising user privacy is termed as the primary motivation for studying methods of secure computation to design an efficient and effective collaborative data mining. It should also do the work of elegantly sharing data between the participants on agreed upon conditions. The two participants in a public domain want to mine data collaboratively in the union of their two different databases. However each private database contains privacy information which need not be disclosed to each other. For successful evaluation of the collaborative mining process joint computation function is defined as per the rule constraints provided by the two participants. The combined computational function in any means did not reveal the privacy; however produce results on the collaborative mining as per each individual's requirements and constraints. The result of one participant will not be disclosed to the other.

In the multiparty computation, with n number of participants each has a private data. The combined computation function work on variable constraints at different session of communication, computes the values required by respective participants in their computational node. The multiparty computation (MPC) protocol is accepted as secured one, if no participant understands more from the description of the combined computation function except their input and results concerned to them only.

The result of the collaborative data mining functional calculation to each participants work under particular conditions depending on the session of communication happened in the public domain.

The rule constraint depends on the security requirement posed by the each individual participant of the secured multiparty computation in their respective session. However internal or external adversary seems to break the computational or conditional rule constraints. The adverse effect can be identified by visualizing some unauthorized warning message during collaborative data mining task. The participants use a synchronized network to identify the adversaries actively or passively in the secured multiparty computation mechanism with the help of session maintenance. Information, theoretically secure multiparty computation is related to the problem of secret sharing, and more specifically verifiable secret sharing.

IV. MULTIPARTY COMPUTATION SCHEME FOR PRESERVING PRIVACY IN COLLABORATIVE DATA MINING

In the collaborative data mining, implicit user authentication confidentiality is only possible for legitimate users and mutual authentication is achieved only after secure communication using the session key start. The proposed multiparty computation is executed purely in the secured channel. The participant's privacy preservation in collaborative data assumes that every participant shares a secret key with the trusted centre in advance either by direct contact or by other ways.

The multiple participants in the data mining domain collectively prune the united data sets of the communication partners to obtain the requisite results. To keep the private data of each individual participant being disclosed to others, session authorization is done by the trusted centre. In the privacy preservation, multiple participants communicate between them in single groups or in parallel multiple groups. The key distribution phase assumes that the TC has been notified to start the multiparty computation between the participants permitted to undergo mining in the collaborative data. Trusted centre and the participants have to perform the computation scheme for secured transaction with the preservation of their private data.

The trusted centre governs the parties involved in specific session of computing the hidden data from the union data sets. The entities (participant, session participants, private and sharable data, and sessions) involved in the communication of participants are independent of each other and in the evaluation of mining process, the task of computation makes all the entities interdependent. In this multiparty computation the major contribution is maintaining the rule of data independency along with the activity of interdependent entities. This makes the privacy preservation more effective in terms of fast mining results obtained during the computation of multiple parties in a more complex environment.

V. SECURITY OF PRIVATE DATA IN MULTIPARTY COMPUTATION

Trusted centrr collects all the inputs from the players, computes the function F and announces the result. The trusted party is linked to all participants via, private and authenticated channels. The trusted party, after receiving an input from the party, runs the machine. If there is an output, the same is sent to the party. Secure computation can be used to solve any distributed data mining problem. Honest adversaries act according to their prescribed actions in the protocol. Figures 1 clearly show the privacy is preserved for the user in a collaborative manner.

The security mechanism provided for the private data in multiparty computation for collaborative data mining is shown in Figure 1. The security mechanism assumes that the participants have privacy data available in his database which is union among other participant database to collectively undergo mining process to obtain their required results. In the process of data privacy maintenance in the multiparty computation comprises of participants, private and sharable data, constrained rule set and the task based session.

The participants are the parties involved in the process of mining the collaborative data to obtain requisite results as per their conditions and input data. The participant maintains a private database, which contains his private data as well as the public data to be shared among different participants. In addition the participant, provide some constrained rule set to access their private database during the collaborative data mining process. The rule set may vary for different tasks as maintained by the session participant involvement or the authorization provided to the accessibility of the private databases by respective individuals. In this framework of secured multiparty computation, task is the process of extracting / mining the required results on a defined session, in which the communicative participants are specified (i.e., authorization of sharing at the particular time instance).

The collaborative data is the union of the entire participant's private database which comprises of private and sharable data. To accomplish a task of pruning the required resultant data from the collaborative data, a session is initiated by the trusted centre for the authorized participants during the communication. With the validated session key the communicating participant start the mining process from their private database to obtain respective resultant outcome for their input conditions and data. The session participant block presented in the framework (Fig.2) provides the exact communicative users in any given session along with the constraints applied for mining the collaborative data.

The adversaries (either external or internal) involved in disrupting the authorized mining of the union databases to get their required information which they are not allowed to access. The security framework presented in this paper, sends the warning message to the trusted centre that the privacy concern of respective participant is in the verge of

violation. Then the trusted centre invalidates the session authentication key issued initially and a caution message is sent to the respective targeted participants (in lieu of maintaining its privacy).

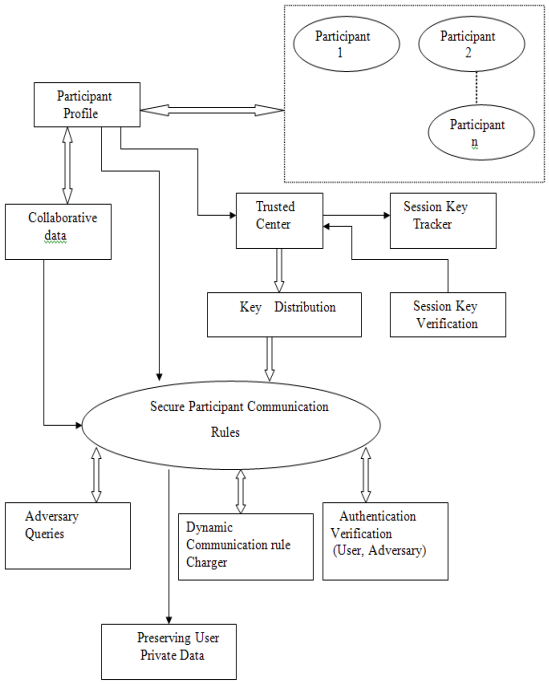

Figure 1. Complete Framework

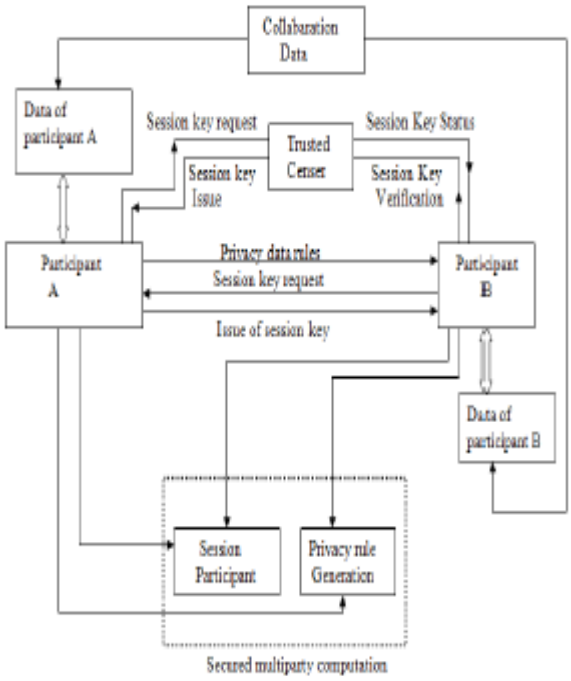

Figure 2. Framework for Secured Multi-Party Computation Process of Collaborative Data

In case of internal adversary being detected, those participants are marked down by the trusted centre for scrupulous activities. The adversary participant also receives a warning message from the trusted centre to regularize the working member of the collaborative data.

The secured multiparty computation is supported to take part in a fixed nonempty set of legitimate participants and the trusted centre. The instances of various sessions are correlated in distinctive and concurrent executions of security computation channel. At the time of session key issue, the pre shared secret key and the session participant are precisely specified for efficient secured communication to obtain their respective results in their node terminals. To compute a session key, participant instance accepts when it gains sufficient information and it should be noted that the state of acceptance only appears in the respective session participant instances. New participant can be accepted at any time with valid authorization provided from the trusted centre for intimation. The session identifier is used to maintain the participant's unique name of his proceeding session. The session identifier for instance based task is the key element of security provider in the execution of secured multiparty computation. The participant identifier names the participant with which shared session key is associated. With this unique session maintenance the participant involvement and their conditional rule set for collaborative data mining are conserved for the effective security deployment of multiparty computation.

Session Based Secured Multiparty Collaboration Data Communication Algorithm:

Step 1 : Data Initialize
Step 2 : Initialize the participant
Step 3 : Session Generation
Step 4 : Session Participant
Step 5 : Malicious adversaries invocation
Step 6 : Resistance of malicious adversary

VI. EXPERIMENTAL EVALUATION OF SECURED MULTIPARTY COLLABORATIVE DATA PRIVACY

In the experimental evaluation process the secured multiparty computation combines the sharing of public data between multiple participants with the key element of maintaining their private data not being disclosed to others in collaborative data stream. The secured multiparty computation, with implicit user authentication is designed with effect of preserving respective participant's private data. It executes multiparty computation purely in the secured communication channel. Every participant shares a secret key with the trusted centre in either by direct contact or by other ways. The experiment is conducted by introducing the participants in one session and union the private database of all the participants. The trusted centre issues session key to the participants in lieu of maintaining the authorization during the computation of requisite resultant data. The participant mine the union data with their input data share required for others.

Table 1. Performance results of secured multiparty computation on preserving privacy in collaborative data

Performance Parameters	Secure Multiparty Computation for Preserving Private Data	Classical Key Security Model
Adversary Resistance Rate	91%	83%
Communication Round	3	4
Execution time for authenticated data sharing (Hundreds of	2 ms	6 ms

Keys)		
Share Key Length	12	16

The experiment is conducted with two modes of privacy preservation with classical key model, and secured multiparty computation key model. The classical model is the traditional cryptographic RSA implementation. The secured multiparty computation allows explicit mutual authentication and its results are listed in Table.1. The adversary resistance rate indicates the effect of secured multiparty computation in maintaining the privacy of the participants involved in mining the complex collaborative data. In this integrated classical and quantum key model shows better resistance rate compared to that of the classical key model. The communication round required among the three models shows that the less number of round is enough for the secured multiparty computation for preserving the private data of individual participants compared to that of other two models.

The efficiency of the secured multiparty computation model is shown by its minimum execution time required to mine the collaborative data as per the rule constraints set by individual session participants concurrently.

In addition the secured multiparty computation needs smaller key size in the communication computation of session maintenance between various tasks. The secured multiparty computation model requires that the trusted centre and each participant pre-share a sequence of shared key pairs. Shared key pairs are measured and need to be reconstructed by the trusted centre and a participant after each session of execution.

The potential effort in mining the collaborative data is its privacy concern between multiple participants along with the sharable data. The efficiency of the session based secured multiparty collaborative data mining model along with preservation of the privacy of all participants is tabulated in Table 2. The number of participants involved in the mining domain versus communication time for sharing the data shows the efficiency factor. Table 2 shows that the secured multiparty communication model has minimal time compared to that of classical key model in graphical representation. In addition, as the number of participants increase, the execution communication time also increases, but this model shows better result than the classical key model.

Table 2 Efficiency (Participants Data Exchange Communication Time) of secured multiparty communication in preserving the privacy compared to classical key model

Number Of Participants	**Communication Time (ms)**	
	session based secured multiparty collaborative data	**classical RSA key model**
10	2.6	5.2
20	3	5.7
30	3.32	5.99
40	3.6	6.32
50	3.8	6.57
60	4.14	6.88
70	4.3	7.21
80	4.65	7.5
90	4.9	7.775
100	5.2	7.9

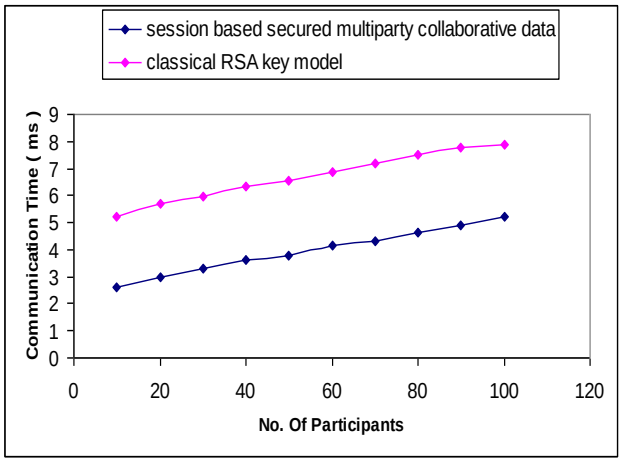

Figure 3 Efficiency of SSMDCM compared to classical key model

The efficiency of the proposed session based multiparty security communication model is also measured in terms of session participant. All the participants may not involve in session communication at a time. Sometimes, the participant in one session may involve in another session with another participant. These possibilities made the researcher to evaluate the efficiency in terms of session participants as well. Table 3 depicts the results of communication time between the session participants to complete in any specific privacy data mining requirements. The performance of secured multiparty computation model in terms of communication time for the session participants is better than the classical RSA key model.

Table 3 Efficiency of session based secured multiparty collaborative data number of session participants Vs communication time

Session Participants	**Communication Time (ms)**	
	session based secured multiparty collaborative data	**classical RSA key model**
10	4.2	5.9
20	4.4	6.55
30	4.6	6.99
40	4.7	7.18
50	4.8	7.37
60	5	7.49
70	5.2	7.73
80	5.4	7.88
90	5.6	8.1
100	5.8	8.1

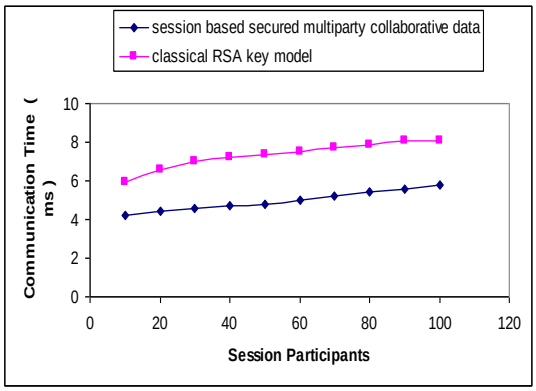

Figure 4 Number of session participants with communication time

A. Effectiveness of Secured Multiparty Computation in Collaborative Data Mining:

The effectiveness of the session based secured multiparty computation is depicted in table-4 and Figure-5 tabulated by means of computing the session participant complexity against the adversary resistance rate. The session participant involved in the mining of collaborative data indicates the number of participants available in a specified session. It also contains the rule constraints applied by participants to maintain their respective data privacy. The adversary resistance rate is obtained during the collaborative data mining and it is measured as the rate of resistance against the internal or external adversaries to maintain the privacy element of its participant in mining the union of private databases.

Table 4 Effectiveness of session based secured multiparty collaborative data against classical key model

Number Of Participants	Adversary Rate (ms)	
	session based secured multiparty collaborative data	classical RSA key model
10	11.4	10.1
40	9.2	6.9
60	8.34	5.4
80	8.18	5.08
100	8.02	5.02

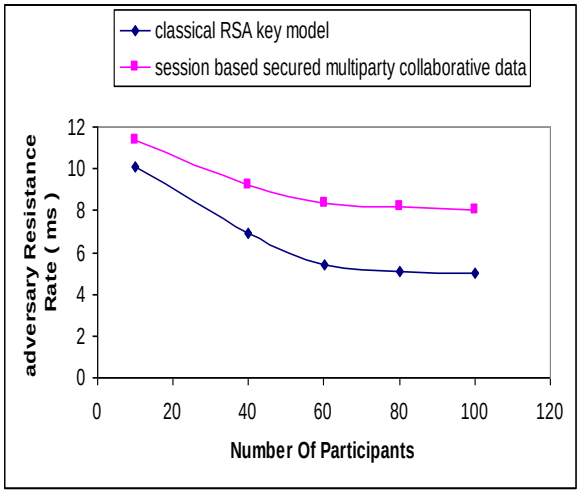

Figure 5 Effectiveness of SSMDCM again

VII. CONCLUSION

The research work presented in this paper demonstrates the multiparty computation for preserving the private data in collaborative data mining. By combining the advantages of classical cryptography with share key cryptography, this work presents a new direction in designing secured multiparty computation for privacy preserving in collaborative data. The design and implementation of preserving the security of data and evaluation with respect to scalability, authentication provision, response time to respond to the adversary's queries and in the collaborative data mining environment have been detailed. Since a fully distributed solution is complex, design divides responsibility between a small numbers of well-known, independent parties. The functionality of preserving private data and finding results can be easily distributed to all the users with greater scalability and reliability. Compared with classical key model, the proposed multiparty computation easily resists replay and passive attacks. Compared with other key distribution schemes, the proposed scheme efficiently achieves key verification and user authentication and preserves a long term secret key between the trusted centre and every user. Experiments with this prototype implementation show that the system performs well under increasing numbers of keys and participants. The multiparty computation for preserving the privacy of participants in collaborative data shows an improvement of 13% throughput for data exchange when compared with traditional cryptography key models. The performance is well within the requirements of motivating collaborative data mining based applications.

The performance analysis of proposed session based secured multiparty collaborative data mining model and comparative classical key models, is made in the above sub sections, show the efficiency in terms of communication time between participants involved in various sessions and the effectiveness of number of participant and session participant against malicious adversaries of internal active and passive participants. With improved performance of our proposed model, we can easily and effectively deploy our session based secured multiparty collaborative data communication model for various social networking communication (i.e., face book, twitter etc.,) where privacy is also essential.

VIII. ACKNOWLEDGEMENTS

I record my sincere thanks to Bharathiar University and Selvam college of technology for providing necessary financial assistance to carry out my research.

IX. REFERENCES

[1] Agrawal R. and Srikant, R. "Privacy-Preserving data mining", Proceedings of the ACM SIGMOD Conference on Management of Data, ACM Press, pp. 439–450, May 2000.

[2] Ahmed HajYasien, "Preserving Privacy in Association Rule Mining", Ph.D Thesis, University of Griffith, June 2007.

[3] Clifton C., "What is privacy? critical steps for Privacy-Preserving data mining", In IEEE ICDM Workshop on Security and Privacy Ascepts of Data Mining, Houston, Texas, USA, November, pp. 27-30, 2005.

[4] Cramer, R., and Damgard, I, " On the amortized complexity of zero-knowledge protocols", Advances in Cryptology - CRYPTO - Aug. 2009.

[5] Dachman-Soled, D., Malkin, T., Raykova, M., and Yung, M. "Efficient robust Private set intersection", Proceedings of the ACNS 2009.

[6] Dingledine, R., Mathewson, N., and Syverson, P. Tor, "The second-generation Onion router", Proceedings of the 13th Usenix security symposium, Aug 2004.

[7] Dolev, D. Dwork, and M. Naor, "Non-malleable cryptography", In Proceedings of the twenty-third annual ACM symposium on Theory of computing, New Orleans, Louisiana, United States, pp: 542 – 552, 1991.

[8] Lindell Y. and Pinkas B, "Secure Multiparty Computation for Privacy Preserving Data Mining", Journal of Privacy and Confidentiality, Number 1, pp. 59-98, 2009.

[9] Lindell Y. and Pinkas B., "Privacy preserving data mining", Advances in Cryptology – Crypto 2000, Lecture Notes in Computer Science, vol. 1880, 2000.

[10] Martin Geisler, "Cryptographic Protocols: Theory and Implementation", Ph.D Thesis, University of Aarhus-Denmark, February 2010.

[11] Tzonelih Hwang, Kuo-Chang Lee et.al, "Provably Secure Three-Party Authenticated Quantum Key Distribution Protocols", IEEE Transactions on Dependable and Secure Computing, Vol:4, pp: 71-80, January 2007.

[12] Goldreich O., "The foundations of cryptography",Vol. 2, Cambridge University Press, 2004.

[13] Golreich O., "Secure Multiparty Computation (working-draft) "http://www.wisdom.weizmann.ac.il/home/oded/public html/foc.html, 1998.

[14] Paillier ,"Public-key cryptosystems based on composite degree residuosity classes", Advances in Cryptography - EUROCRYPT '99, Prague, Czech Republic, pp 223-238, 1999.

[15] Sourav S. Bhowmick, et.al., "PRIVATE-IYE: A Framework for Privacy Preserving Data Integration", In Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06), 2006.

[16] Yao. A.C., "Protocols for secure computations", Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, pp.160-164, 1982.

[17] Zhang N., Wang S., Zhao W. "A New Scheme on Privacy-Preserving Data Classification", In Proceedings of the SIGKDD, 2005.