ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

# Privacy Preserving Multiparty Collaborative Data Mining for Multiple Service Providers

**Shrishti Pawar *, Hare Ram Shah**
*Research Scholar, Gyan Ganga Institute of Technology And Science, Jabalpur, M.P, India.
Associate Professor, Gyan Ganga Institute of Technology And Science, Jabalpur, M.P, India
Shrishti.pawar01@gmail.com

## Abstract

The emergence of Application Service Providers hosting Internet-based data mining services is being seen as available alternative for organisations that value their knowledge resources but are constrained by the high cost of data mining software in this paper, we present a new multiple service provider model of operation for the Internet delivery of data mining services. This model has several advantages over the currently predominant approach for delivering data mining, services such as providing clients with a wider variety of options, choice of service providers and the benefits of a more competitive marketplace. In the current modern business environment, its success is defined by collaboration, team efforts and partnership, rather than lonely spectacular individual efforts in isolation. So the collaboration becomes especially important because of the mutual benefit it brings. For this kind of collaboration, data's privacy becomes extremely important: all the parties of the collaboration promise to provide their private data to the collaboration, but neither of them wants each other or any third party to learn much about their private data. One of the major problems that accompany with the huge collection or repository of data is confidentiality. The need for privacy is sometimes due to law or can be motivated by business interests. Performance of privacy preserving collaborative data using secure multiparty computation is evaluated with attack resistance rate measured in terms of time, number of session and participants and memory for privacy preservation.

**Keywords**: Privacy preserving data mining, privacy preserving collaborative data mining, secure multiparty computation.

## Introduction

This research study is concerned with the study and analysis of preserving privacy in collaborative data mining in order to improve the efficiency and effectiveness of privacy preservation among the collaborative parties. Data mining is a process which uses different data analysis tools that discover patterns and relationships in data that can be used to make predictions. 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 and in such a situation, the process is called Privacy Preserving Collaborative Data Mining (PPDM) and the gap between the data mining and data confidentiality can be filled by the privacy preserving data mining. The paper is organized as follows. Section 2 introduces a multiple service provider model for Internet delivery of data mining services. It discusses the comparative differences in operation and interaction protocols between the two models. Section 3 presents the information exchange process for the different models of operation.

### Models of Operation for Data Mining Service Providers

This section presents alternative models of operation for data mining service providers. These models illustrate the context of interaction and communication between "clients" and data mining service providers. We discuss the existing single service provider model – where one ASP host several data mining systems - and present our alternative multiple service provider model – where several ASP's host one or more data mining systems.

### Single Service Provider Model

This model has simpler operational semantics of the two and is the currently predominant approach to providing Internet-based data mining services. A client organization has a single service provider who meets all the data mining needs of the client. The client is well aware of the capabilities of the service provider and there are predefined agreements regarding quality of service, cost and protocols for requesting services. The service provider hosts one or more distributed data mining systems, which support a specified number of mining algorithms. The service provider is aware of the architectural model, specializations,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

features and required computational resources for the operation of the distributed data mining system.

## Multiple service provider models

This model, as illustrated in figure 1, characterized by clients being able to request data mining services from several service providers who host one or more DDM systems.

This approach provides a higher level flexibility for the client and represents the establishment of an open, virtual market place of data mining service providers. The multiple service provider models operate in the form of a "federation". The "federation manager" is a coordinating component in the system that manages the interactions between the client and the data mining service providers. The interaction protocol for this model is as follows:

1. The client requests a service by providing a task specification to the federation manager. It must be noted that the parameters for specifying the task must be well defined and must facilitate the requests to be made at the level of granularity that the client deems appropriate.

2. The federation manager broadcasts the client's requests to the data mining service providers that are registered with it. The federation manager maintains information about each data mining service provider such as the name ,address, contact information, DDM systems hosted, algorithms, architectures and functionality supported by those systems and the computational resources that the service provider has.

3. The data mining service providers evaluate the requested task against the capabilities and functionality of the DDM systems that they host.

4. If they can meet the needs of the requested task, the data mining service providers respond by presenting an estimate of the cost, possible time frame for completion of the task and liabilities for not meeting the targeted response time. This information is presented to the federation manager in a specific, structured format.

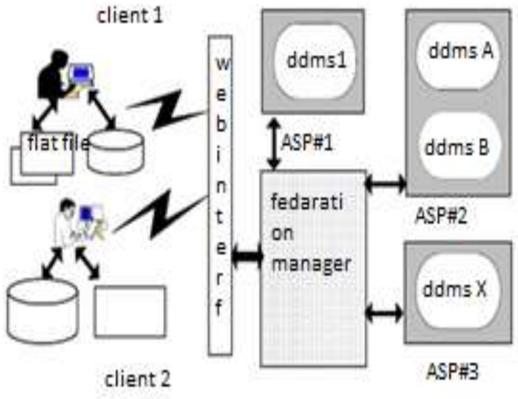

*Figure 1: Multiple service provider model*

5. The federation manager can either present the responses it receives along with the information that it already maintains about the respective service providers to the client or in more sophisticated environment can perform matching of client preferences with the capabilities of the service providers and rank the service providers on that basis and present this to the client. There is also scope for automated negotiation to be incorporated into this stage of the interaction protocol. It must be noted that service-provider ranking, preference matching and automated negotiation are emerging research issues in the e-services domain. We present a brief discussion on these aspects in section 6.3, but it is not the core focus of this paper.

6. The client decides which service provider it deems most appropriate and informs the federation manager and the chosen service provider

7. The service provider gives the client a legal document/contract, which makes the commitment to maintain the confidentiality of the data that is mined and the consequent knowledge that is produced.

8. The client is then required to provide a security deposit in the form of a credit card number to the federation manager. The actual payment is made on completion of the task and provision of results to the client.

9. The client and the data mining service provider exchange information regarding the transfer of data, passwords to access systems and the mode of transfer of results.

10. The data mining service provider processes the task, provides the results to the client (in the agreed format and method) and informs the federation manager of task completion.

11. The client acknowledges the completion of the task to the federation manager and the payment is made to the service provider.

This model overcomes the limitations and restrictions imposed by the previous approach in meeting the needs of the requirements outlined in section 1.

***Information Exchange Process***

There is a significant difference in the manner in which the information exchange takes place between the models involving single and multiple service providers. The principal difference, as illustrated in figure 2, is that information such as access information for the client and access information for the service provider need only be exchanged only once in the single service provider model. This does not hold for the multiple service provider model, where the contractual agreement between the service provider and the client is relatively short-term (and typically limited to a single task). This in turn necessitates exchanging all the information on a per-task basis .Further, in the multiple service provider model, the data mining task description, client access information and service provider access information.

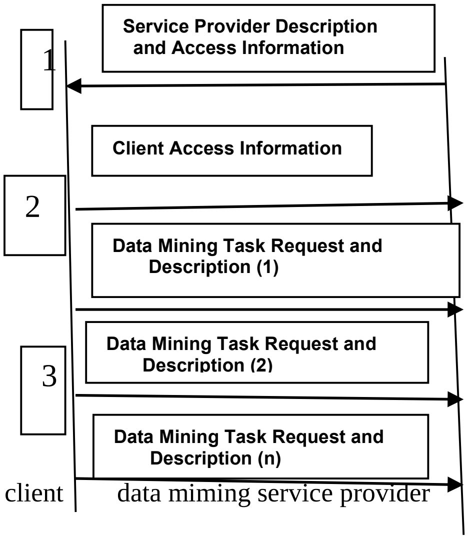
Service Provider Description and Access Information
1
Client Access Information
2
Data Mining Task Request and Description (1)
Data Mining Task Request and Description (2)
3
Data Mining Task Request and Description (n)
client
data miming service provider

*Figure 2: Single service provider model*

After selecting the service provider the information exchange occurs directly between the client and the service provider by passing the intermediary.

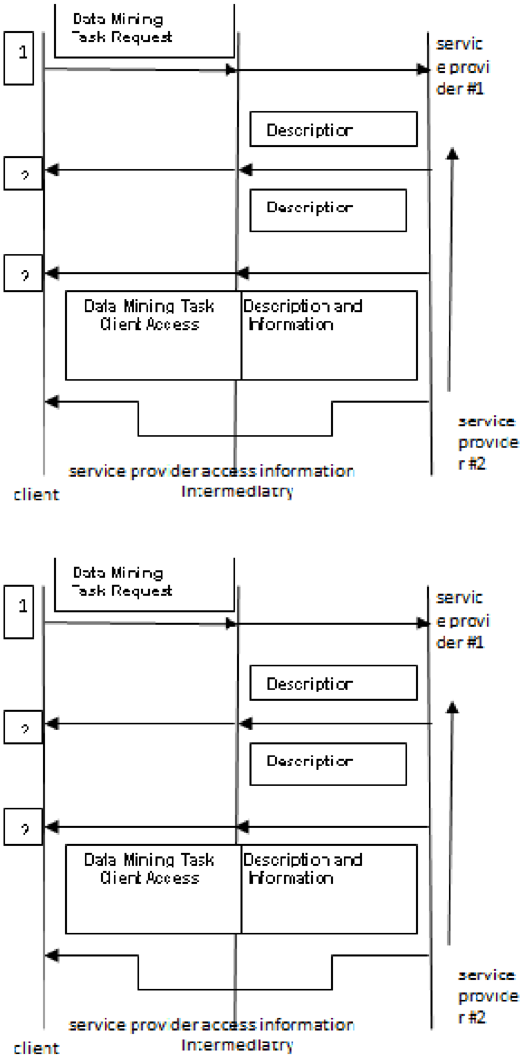

*Figure 3: multiple service provider*

## Materials and methods

The goal of methods for Secure Multi-party Computation (SMC) is to enable parties to jointly compute a function over their inputs, while at the same time keeping these inputs private. SMC refers to computational systems 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. SMC problem deals with computing any probabilistic function on any input, in a distributed network where each participant holds one of the inputs, ensuring independence of the inputs, correctness of the computation and that no more information is revealed to a participant in the computation than can be inferred from that participant's input and output. However success of privacy preserving data mining may depend on the ability to find new definitions that provide both the rigorous security guarantees that are necessary and can be met by highly efficient protocols.

## Apriori Algorithm for multiple service providers

```
$F_1$=(frequent itemset of cardinality);
For (k=1; $f_k$=φ;k++)do begin
$C_{k+1}$=apriore_gen($f_k$);// new candidate for all transaction t E database do begin
$C_t$= subset ($c_{k+1}$, t);// candidate contained in t for all candidate C E $C_t$ do
c.count++;
end
$f_{k+1}$={c E $C_{k+1}$| c.count≥minimum support}
end
```

### Secure multiparty computation

Recent advances in computing, communication have enabled large volumes of data to be accessible remotely across geographical boundaries. There is a great demand on collaborative data mining when compared to the distributed data stores to find the patterns or rules that benefit all of the participants. An important challenge for distributed collaborative data mining is how to protect each participant's sensitive information, while still finding useful data models. The techniques for performing privacy-preserving data mining are drawn from data mining, cryptography and information hiding. The service-oriented infrastructure for collaborative data mining of data distributed has become the most popular solution. Here the data providers are the collaborators who submit their own datasets to the required data mining service provider for discovering and mining the commonly interested models on the pooled data set.

The problem of privacy preserving data mining has become more important in recent years because of the increasing ability to store personal data about users. Number of techniques has been suggested in recent years in order to perform privacy preserving data mining. Multiparty secure computation allows N parties to share a computation, each learning only what can be inferred from their own inputs and the output of the computation. The core idea of the Secure Multiparty Computation (SMC) is secure and at the end of the computation, no participant has the knowledge except its own input and the results. The malicious adversary participant tries to capture the private information of authenticated participant or cause the computation made between the sharable participant's collaborative data to be incorrect.

The main contribution of this study is to securely compute multiparty information with privacy preservation in parallel by maintaining multiple sessions of participants. Each session is validated by trusted third party and the participants involved in it generate individual data independent rules and dependent sharable data rules to work effectively with collaborative data without losing individual's private data. A framework for secured multiparty computation process has been designed. An algorithm has been designed and developed for instance and dynamic rule generation for secure multiparty computation in collaborative data mining.

## Conclusion

The problem of privacy preserving data mining has become more important in recent years because of the increasing ability to store personal data about users. Number of techniques has been suggested in recent years in order to perform privacy preserving data mining. Secure multiparty computation allows N parties to share a computation, each learning only what can be inferred from their own inputs and the output of the computation. The emergence of Internet-based data mining service providers is proving to be a viable means for satisfying the business intelligence needs of knowledge-centric organizations. In this paper we have presented a multiple service provider model as an alternative operational model to the currently predominant approach among data mining service providers The core idea of the Secure Multiparty Computation (SMC) computation is secure and at the end of the computation, no participant has the knowledge except its own input and the results. The malicious adversary participant tries to capture the private information of authenticated participant or cause the computation made between the sharable participant's collaborative data to be incorrect. The privacy rules generated in the proposed collaborative multiparty computation verifies the authenticity of participants and restricts or avoids the internal malicious adversary participant to involve in the mining. we see that the potential benefits of the multiple service provider model.

## References

[1] *Agrawal, D. and C.C. Aggarwal, 2001. On the design and quantification of privacy preserving data mining algorithms. Proceedings of the 20th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 21-23, ACM, Santa Barbara, California, USA., pp: 247-255. DOI: 10.1145/375551.375602*

[2] *Bhuvana, J. and T. Devi, 2011. Performance of secure multiparty computation for preserving privacy in collaborative data mining. Int. J. Res. Rev. Comput. Sci., 2: 463-469*

[3] *R. Agrawal and R. Srikant, "Privacy-preserving datamining," In Proceedings of the ACM SIGMOD Conference onManagement of Data, ACM Press, pp. 439–450*

[4] *D. Agrawal and C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," In Proceedings of the 20th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Santa Barbara, CA, pp. 247–255, May 21–23*

[5] *Privacy-Preserving Multiparty Collaborative Mining with Geometric Data Perturbation Keke Chen, Member, IEEE, and Ling Liu, Senior Member, IEEE vol 20 no.12 December2009*

[6] *A New Scheme to Privacy-Preserving Collaborative Data Mining Jianming ZhuSchool of Information, Central University of Finance and Economics, Beijing, China E-mail:tyzjm65@163.com*

[7] *Jie Wang, Jun Zhang. Addressing accuracy issues inprivacy preserving data mining through matrix factorization,Intelligence and Security Informatics, 2007 IEEE, 23-24 May 2007 Page(s):217 – 2*

[8] *Feng Li, Jin Ma, Jian-hua Li. An adaptive privacypreserving data mining model under distributed environment,Signal-Image IEEE Conference Technologies and Internet-Based System, 2007.SITIS '07. Third International on, 16-18 Dec. 2007 Page(s):60 – 68*