Third Party Privacy Preserving Protocol for Secure Web Services

B.Raghuram
Assistant Professor,Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

S.Sandeep
M.Tech.(SE) Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

B.Hanmanthu
Assistant Professor,Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

ABSTRACT
Web services is become major issue in distributed data mining. In the literature we can found a number of proposals of privacy preserving which can be divided into two major categories that is trusted third party and multiparty based privacy protocols. In case of the trusted third party privacy protocol models the conventional asymmetric cryptographic based techniques or algorithms will be used and in case of the multi party based protocols data perturbed technique can be used to ensure no other party to understand original data. In order to improve security features by combining strengths of both the above said models in this work, we propose to use data perturbed techniques in third party privacy preserving protocol to conduct the secure web service. In order to perform web service we propose third party protocol for secure computations. Our results shown that although the data are disguised and decentralized, our method can still achieve literally high accuracy.

Keywords: Distributed Data mining, Vertical Fragmentation, Third Party Privacy Preserving, Data Perturbation, Web Service

1. INTRODUCTION
Web service is a combinations of services published to network for use by other service providers through a standard synchronous system. It provides the opportunity of apparent amalgamation among heterogeneous platforms and applications. The Web service participates in various operations abounding by different providers will be consolidated to achieve a superior range set of services. In this paper, we provide a framework based on Web service, integrating a privacy preserving data mining components, including data preprocessing for privacy requirements, model construction and model deployment for privacy and distributed web service proving. In this framework, geographically distributed software cooperates with each other in secure manner using the pertubated privacy concepts it will execute dynamically according to users preference.

Web services are self explanatory software applications that will be provided, placed and spread across the Internet using a set of standards such as SOAP, WSDL, and UDDI. Because it is a service based and self explainer, SOAP communication can suggest information between services in heterogeneous computing environments without worrying about inter disciplinary protocol problems, there are many other Web Service provisions. Two of them, which are based on XML, are Web Service Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI). WSDL defines a typical method of unfolding a Web Service and its potential, and UDDI defines XML-based-rules for providing Web Service information. Messages are transferred through the SOAP protocol. This allows the data to be seamlessly exchanged not considering of where the client is in the network using world wide web. But same seamless integration will provide certain problem of privacy. So addressing the privacy issues can be the major issues in the area of web service. In this paper we provide a solution for seamless integration of web services for distributed data mining which will be secured by third party privacy preserving based on perturbed technique.

Distributed data mining [1] playing great role in extracting knowledge from geographically distributed sources. The exposures of distributed data mining include sensor networks, mobile ad-hoc communications, context aware computing, weather forecasting, intrusion detection systems and web mining etc. The corporate and financial sectors as they spread over different geographic locations transforming their business intelligence applications to distributed application form conventional centralized data warehouse applications. Even though centralized data warehouse based models are more accurate than distributed applications, due to their easy scalability, communication efficiency and privacy facilities distributed data mining applications attracting lots of researcher's attention.

Data mining approaches on geo graphically distributed data sources can be applied on two ways known as centralized model and distributed model. In the case of centralized model the required data distributed over various sources gathered in to a centralized site where Data mining algorithm will be applied. It provides accurate results but impose huge communication cost and time. Where as in distributed environment, mining will be performed at local sites and results of local sites will be optimized based on feedback from other sites. Even though distributed model will gives less accuracy than centralized model it reduce communication cost, time complexity and makes algorithm easily scalable. In order to develop a model with optimum time and communication cost is concern we adopted distributed environment.

In the literature study we found that many proposals of privacy preserving which can be divided into broadly two major categories out of two the first category is trusted third party and second one multiparty based privacy protocols. In case of the first category trusted third party models the conventional asymmetric cryptographic based techniques or algorithms will be used and in case of multi party based

protocols data perturbed technique can be used so that to make sure no other party to understand original data. In order to maintain enhanced security features by combining strengths of both models in this paper, we propose to use data perturbed techniques in third party privacy preserving protocol to conduct the web services. The remaining part of the paper organized in following manner, in section 2 presents' related works and in section 3 privacy preserving web service model is given finally the performance and evaluation given in section 4.

2. PREVIOUS WORK

The existing work in privacy-preserving data mining has addressed two issues. In one, the main aim is to preserve customer privacy by perturbing the data values [3]. In this scheme random noise data is introduced to distort sensitive values, and the distribution of the random data is used to generate a new data distribution which is close to the original data distribution without revealing the original data values. The estimated original data distribution is used to reconstruct the data, and data mining techniques, such as classifiers and association rules are applied to the reconstructed data set. Later refinement of this approach has tightened estimation of original values based on the distorted data [4]. Perturbation methods and their privacy protection have been criticized because some methods may derive private information from the reconstruction step [5]. One important category is multiplicative perturbation method. In the view of geometric property of the data, multiplying the original data values with a random noise matrix is to rotate the original data matrix, so it is also called rotated based perturbation.

With the greater growth of both the quantity and the diversity of the web services on data mining, an increasing selection of Web services can be brought about the same options on data mining. In the previous decade the process based models to Web services composition has raised considerable impetus and consistency. The lack of the growth in present UDDI models it is that minimizes the service innovation to functional necessities only. It is fact that there may be greater number of Web services accessible that can meet the functional necessities with different excellence of service attributes. Therefore the capability of providing excellence of service into service discovery process and providing privacy becomes very important.

The Web Service has combined the technology with XML (Extended Markup Language), SOAP (Simple Object Access Protocol).UDDI (Universe Description Discovery Integration), WSDL (Web Service Description Language) and others. The basic concept of the Web Service pre order process is to use the greater and recent technology including service account, communication protocols, data mining and others. To our knowledge, very few works focus on mapping or modifying perturbation technique with third party based protocol. In order to provide enhanced security feature we propose trusted third party based web services.

3. PROPOSED WORK

Trusted Third Party Privacy preserving protocol for web services using perturbed data In the distributed third-party setting model as shown in "Fig. 1: Trusted third party based system for Web services" a database set D holds the secret databases D1,D2…..Dn, respectively, each of which can be regarded as a relational table. Each of the databases has same number of rows. All this sub databases of D shares the sub set of variable of same transactions. The ID is common to all that links the rows in distributed environment among all the sites and all subset of transactions also holds the class label to identify transactions which belongs. A trusted third party data miner initiates the web service process. All sites can generate web services process. The locally generated values and private key encrypted attribute vector send to trusted third party. All local sites will repeat the same procedure. The values send to trusted third party which holds public key of all sites to generate web services. Particularly our model aimed to perform web service on distributed data bases where data stored in various shared nothing machine connected over distributed environment. Our model assumes that data is distributed over various sources. In distributive environment we may face two criteria's out of one is the training data distributed among various sources and web service need to generate on one source. In other approach along with training data distribution the web services need to perform on various sources. We proposed our model to handle both of these factors.

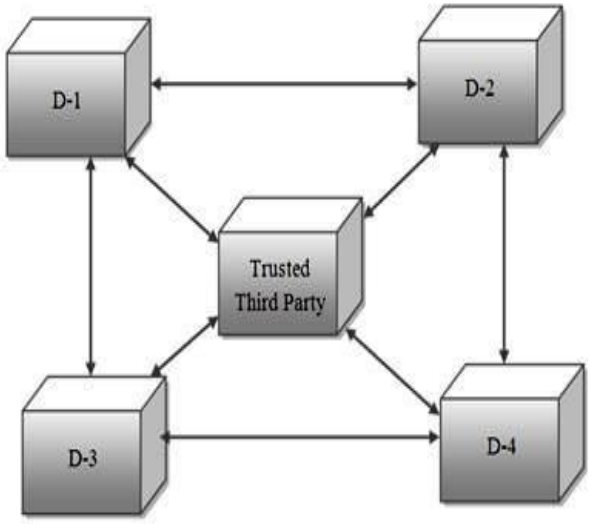

Fig.1: Trusted third party based system for Web services

3.1 Trusted Third party protocol for perturbed data

Step1:The process will be started by the trusted third party by broadcasting web service initiation message.

Step2:All sites which ever interest will respond by sending ready message and their public keys.

Step3:Then the trusted third party will respond by start message.

Step4:Each site generates its perturbed data and calculates individual web service based values

Step5:The individual sites web services will encrypt the required data by private key using RSA Algorithm.

Step6:The individual sites send their encrypted data to trusted third party.

Step7:The trusted third party which hold public key will decrypt data and apply perturbed web services.

1. Process initiates by broadcasting Web service Message
2. Initiation||Interested site Responds by sending start message||Public Key
3. Gives respond by start message
4. Individual sites generates its perturbed data and calculates web service based values
5. Calculated values encrypted using private key and sends to Trusted Third Party
6. After receiving all the encrypted copies of data it will decrypts using their public keys and apply perturbed web service

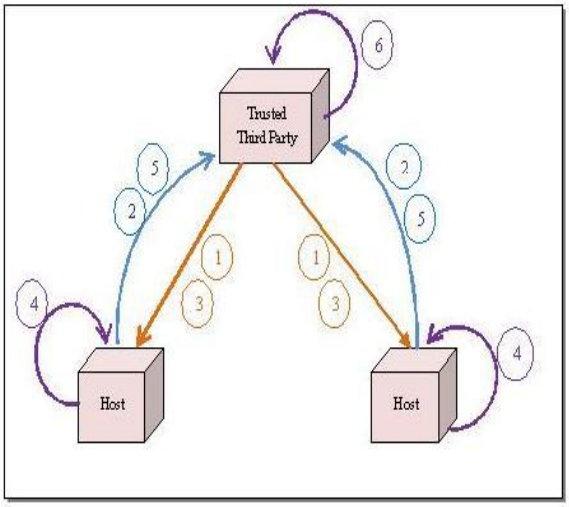

Fig.2 The communication model of trusted third party based system for Web services

The above steps are derived from figure.2 "The communication model of trusted third party based system for Web services". The steps comprises initially the trusted third party initiates the process to enable all the sites by broadcasting a web service message all sites may receives this message but only whoever interested to participate they may sends replay initiation message start message Public keys by replaying that initiation message the trusted third party come to know that they are in group, it also knows interested one and they also sends along with start message their public keys to trusted third party then only who ever given response those sites may receives start message then individual sites generates its perturbed data and calculates web service based values whatever values they have calculated those values encrypted using their private keys and sends the encrypted messages to trusted third party after receiving all such messages at this level it will decrypts the messages using their public keys and applies the same to perturbed web service.

3.2 Web services Construction over Perturbed Data

Web services of each Web data mining can be dynamically integrated through SOAP, WSDL, UDDI and WSFL (Workflow Services Description Language).Web data mining service consists a describe their given that services through the WSDL and information buffer server of Web services about these metaphors. Registration server rationalized directory and bring out them on Internet according to the description of WSDL and UDDI agreement. The Web Services Description Language (WSDL) for a web data mining service provider can be found from local, public, or commercial Universal Description, Discovery, and Integration (UDDI) Registry. The execution sequence for web services is decided and integrated using the activity elements. After finishing the design, the process will be deployed for later execution. In this process of web service providing each requester and service provider the protocols for privacy given in the previous sections. First to initiate the communication the service requester perform key exchange procedure and the perfume communication about web service based data mining using the proposed trusted third party based privacy preserving protocol for web services.

4. PERFORMANCE ANALYSIS

In order to testify the performance of our proposed model, our experiments utilized four P4 2.40GHz PCs with 512Mb main memory and windows XP Operating System. The four PCs are located in 100Mbps LAN out one we used as trusted third party which initiates the process rest of machines stores data sets. We use the credit card and car maintenance data sets obtained from UCI Machine Learning Repository.

4.1 Experiment Setup

At first step of experiment the data sets are processed by discredited quantitative items and partition data sets vertically among three systems connected in distributed environment. For both PIMA and Heart datasets we created unique id for each row and split in to three parts where each part row carry same id and stored in three different systems. In second step proposed web service model implemented on training data sets at local sites and generated results are encrypted with senders public key at receivers side it will be decrypted using receivers private key it ensures the security and all the users keys i.e., private and public keys send to trusted third party for this method we adopted RSA algorithm. In third step combining all local data of individual sites gathered at trusted third party and the same will be used to generate global web service.

4.2 Security Analysis

Our model preserves privacy of results sent to successor site by a site in order to finding global frequent item sets because local frequent values sent in scalar matrix form and item set names sent in encrypted vector form. This is possible because we are using scalar product method to generate frequent item sets using matrixes of two sites. The trusted third party is having certain privileges such as initiation of the mining process, decryption of frequent item sets, finding global frequent item sets and association rules but it can't predict any site's private data because the only the global frequent data came to it after processing at local sites.

5. CONCLUSION

In this paper we proposed improved security feature model by combining strengths of both the models which we discussed above, we also proposed to use data perturbed techniques in third party privacy preserving protocol to conduct the web services. The experiment conducted on UCI data sets proved that our model shown good accuracy measures. We also showed analytically that our model is time, cost and privacy efficient.

6. ACKNOWLEDGMENT

Our thanks to the management members and principal of Kakatiya Institute of Technology and Science-Warangal who have facilitated resources to read and compute in order to develop this model and narrate this article and our sincere

thanks to Head of the Department Prof.P.Niranjan who encouraged us research and publish this paper.

7. REFERENCES

[1] M Zaki. Parallel and Distributed Data Mining: An Introduction. In Large-Scale Parallel Data Mining, pages 1–23. 2000.

[2] J. Vaidya, and C. Clifiton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proceedings of SIGKDD, Canada, 2002.

[3] R. Agrawal and R. Srikant. Privacy-preserving data mining. In SIGMOD Conference, pages 439–450, 2000.

[4] D. Agrawal and C. C. Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In PODS. ACM, 2001.

[5] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar. On the privacy preserving properties of random data perturbation techniques. In ICDM, pages 99–106. IEEE Computer Society, 2003.

[6] W.Du and Z. Zhan. Using randomized response techniques for privacy-preserving data mining. In KDD, pages 505– 510, 2003.

[7] A. V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke.Privacy preserving mining of association rules. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 217–228,2002.

[8] M. Kantarcioglu and C. Clifton. Privately computing a distributed k-nn classifier. In J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, editors, PKDD, volume 3202 of Lecture Notes in Computer Science, pages 279–290. Springer, 2004.

[9] Y. Lindell and B. Pinkas. Privacy preserving data mining. In M. Bellare, editor, CRYPTO, volume 1880 of Lecture Notes in Computer Science, pages 36–54. Springer, 2000.

[10] L. Liu, M. Kantarcioglu, and B. Thuraisingham. The applicability of the perturbation based privacy preserving data mining for real-world data. Data and Knowledge Engineering Journal, 2007.

[11] T. M. Mitchell. Machine Learning. mcgraw-hill, 1997.

[12] B. Rozeberg, and E. Gudes "Association rule mining in vertically partioned databases", Data and Knowledge Engineering, Elsever, pp378-396,2006.

[13] Z.Yang,and N. Wright "Privacy-Preserving Computation of Bayesian Networks on Vertically Partitioned Data", IEEE Transactions on Knowledge and Data Engineering, Vol.18, No.9, pp 1253-1265, 2006.

[14] H. Zheng, and S. R. Kulkarni "Attribute distributed learning: Models,limits, and aglorithms" .IEEE Transactions on Signal processing, Vol.59,no.1, pp386-398,2011.

[15] Y. Sang, H. Shen,H.Tain, "Privacy preserving tuple matching in distributed databases", IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.12, pp 1767-1782, 2009.

[16] A. Delis and V.S. Verykos, and A.Tisonsis, "Data pertubaration approach to sensitive classification rule hiding,"Prociding of ACM SAC 10, New York,USA, pp.605-609, 2010.

[17] B.Hanmanthu, B.RaghuRam, and P.Niranjan, "Third Party Privacy Preserving Protocol for Perturbation Based Classification of Vertically Fragmented Databases," in Proceedings of ELSEVIER, International Conference on Emerging Trends in Electrical, Communication and Information Technologies (ICECIT - 2012), 21-23 December 2012, Anantapur - 515 701, Andhra Pradesh, India.pp.109-113.

[18] B RaghuRam, Jayadev Gyani, B.Hanmanthu "Fuzzy Associative Classifier for Distributed Mining," in Proceedings of IJCA, International Conference and workshop on Emerging Trends in Technology (ICWET 2012),24-25 February 2012, Mumbai,India,pp.431-435. (ISBN:973-93-80864-47-3)

AUTHOR'S DETAILS

B.Raghuram obtained his Bachelor's degree in Computer Science and Engineering from JNT University of India. Then he obtained his Master's degree in Computer Science and Engineering from Pondicherry Central University Pondicherry, and he is also life member of ISTE. He is currently Assistant Professor of Computer Science and Engineering, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal. His specializations include Data mining and Data warehousing, Databases and networking.

S.Sandeep Currently persuing Master of Technology in Computer Science and engineering with specilization in Softwatre Engineering.Computer Science and Engineering Department, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal.A.P.,India.

B.Hanmanthu obtained his Bachelor's degree in Computer Science and Engineering from JNT University of India. Then he obtained his Master's degree in Computer Science and Engineering with specialization Software Engineering from JNT University Hyderabad, and he is also life member of ISTE. He is currently Assistant Professor of Computer Science and Engineering, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal. His specializations include Data mining and Data warehousing, Databases and networking.