Efficient Privacy-Preserving Using Novel Based Secure Protocol in SVM

Hamsaveni N[1], Shere banu M[2], Revanth S[3], Selva Prasad J[4], Sashi Rekha K[5]

[1,2,3,4] Student, Computer science and engineering, Dr. N.G.P Institute of Technology, Coimbatore, India
[5] Assistant Professor, Computer science and engineering, Dr. N.G.P Institute of Technology, Coimbatore, India

---***---

Abstract - *Data Mining is the practice that categorizes large datasets to extract the meaningful information. Data are analyzed and segmented in the database that is helpful in identifying the previously hidden patterns of raw data. Support Vector Machine is used for classifying the data in the database Server. Users data will be classified and the results of classification be sent to the user which suffers from data leakage. Given the importance of maintaining security of original data, we proposed a new novel based framework protocol proclaimed as Light Weight Multiparty Random Masking and Polynomial Aggregation Protocol. In our work, the SVM classification result serves more Accuracy, Efficiency and also maintains the confidentiality of user's data.*

Key Words*: *data mining, classification, cloud computing, support vector machine, privacy preserving

1. INTRODUCTION

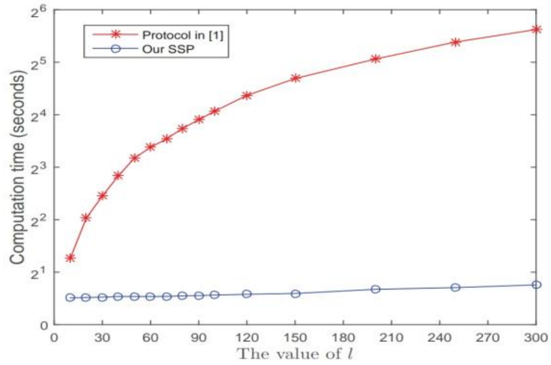
In today's era where the technologies are emerging widely in various Phases and in different fields and forms, an extra pinch is added by data mining. The technological development has marked a point in which a lead taken by data mining to provide a perk to artificial intelligence and machine learning immersing along with it. Also, owing to massive increment of the datasets, retrieving of such data and maintaining is formidable. So, the most Influential and heft technique used for extracting and culling the information from the huge database is the data mining. Data mining is an epitome which involves large scale of datasets and associates to cluster, classify, and detect the anomalies and henceforth summarizing the data. As classifying the data is essential in data mining there is a major issue in maintaining them, this directs to the introduction of support vector machine (SVM).

Considering the significance of mining and decision making, the organization and institution requires support vector machine to provide better classification result of their data. Our System comprises of three entities that include a cloud Computing, server and data provider.

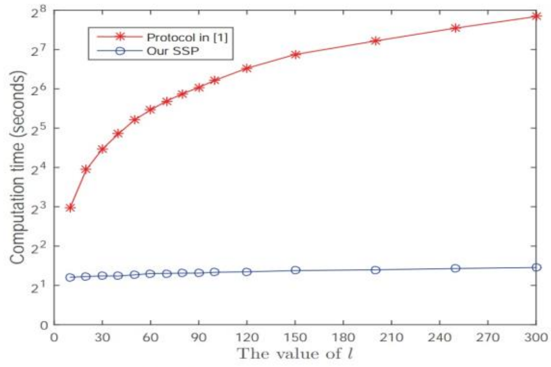
Privacy is the key aspect which encloses sensitive and primitive form of information about the users. Managing of the data in the big organizations is done mostly through the cloud computing technology. All though, cloud computing is very popular for the purpose of the storage; the loop hole here is security. To prevent the Jeopardizing of data, the encryption is done using novel based framework such as Light weight multiparty random masking and Polynomial Aggregation protocol.

2. EXISTING WORK

In existing work, a protocol is used to securely determine the classification result. In addition, existing scheme uses hybrid approach utilizing the combination of homomorphic encryption system and garbled circuit protocol. The existing scheme can also be severed as an efficient basic building in any other application that needs to figure out the sign of a number in encrypted domain. The Secure Sign Protocol (SSP) used in the existing system which can securely and correctly work out the sign of the user. SSP discloses nothing useful about the privacy of any participant, and is probably secure only when it deals with small length of the users keys. With the growth of *l* which is a positive integer used in the user's key, SSP can save more computation time. If the value of *l* is high, the SSP is consuming more time. Thus, our computation time increases, as the key is longer. Nevertheless, the value of *l* has little influence on computation time, since the size of the garbled circuit protocol does not vary with *l*. For a fixed key, the computation time of SSP almost remains the same, as *l* varies. Therefore, in the existing work, SSP is not having efficiency to securely determine the sign in encrypted domain[1].



(a) The bit length of key is 1024.



(b) The bit length of key is 2048.

3. PROPOSED WORK

In proposed work, a novel based approach is used for the data security. A novel based protocol is the combination of two protocols called Light weight Multiparty Random Masking and Polynomial Aggregation Privacy-Preserving protocol. This two protocols works separately to enhance the data security. In proposed scheme, the user's data are encrypted using the Light weight Multiparty Random Masking protocol. Then, data leakage is prevented using the Polynomial Aggregation Privacy-Preserving protocol. This protocol increases the efficiency and enhances the security of data.

3.1 DATA MINING

Data analysis plays major role in data mining. Refinement of results through mining is due to analysis of data in database. Data mining is one of the inter-disciplinary fields which are made up of combination of statistics, technological database and artificial intelligence. The presentable work of mining for the custom data is used for predicting the future and explaining the past by means of analyzing the data in the huge database. Usage of database by business organizations are widely increasing due to increased amount of the data over years and the Users. This leads to the data mining tools and similarly, application ratio of data mining is also estimated to be very high at most of the time. Hence, data mining is used for extracting the most knowledgeable which is also very valuable from the datasets.

The implementation of mining gives the refinement of data from the huge amount of datasets. Data stored in the database is not only involving with single user also comprises of combination of business organizations and institutions. This does not give assurance to the security[2] of data. Sometimes the data reflects the thorough information about any organizations or users private data.

3.2 CLOUD COMPUTING

The cloud computing is used for maintaining the intercommunication which includes the user and the server. In daily life, internet is an important tool either personally or professionally. Due to movement, presence and cheap, the cloud computing is highly known. The main purpose is to extract structure information from the unstructured or semi-structured data resources[3]. Cloud security design is active only when the right gateway implementations are in location. The main intension of this controlling are to minimise the hacking on the cloud computing. In proposed work, the data which needs to be classified for the user send the data to the server. After the data has been classified by using SVM classifier[4], results of classification gets stored in the cloud.

3.3 SUPPORT VECTOR MACHINE

Support vector machine is a supervised machine learning algorithm. SVM fulfills the needs of both classification as well as regression. A SVM is normally used to segregate the data and classify them into datasets .SVM[6] classifies the datasets and eventually labels them according to the user's requirement. In this particular paper the data is sent for a text classification in which the data is segregated and henceforth labeled. The labeled data is taken and using the algorithm it is encrypted. In previous work also SVM was used for purpose of classification. However here SVM plays a major role in the case of segregation.SVM is therefore used not only in technological stuffs it is also used in other fields as well.

4. MECHANISM

In proposed, a new novel framework based on lightweight multiparty random masking and polynomial privacy-preserving is proposed for improving the efficiency and privacy of protocol. Furthermore, the proposed framework can efficiently protect the data privacy as well as ensure confidentiality. Specifically, for a data query from a registered user, the response is directly performed on cipher text at the service provider without decryption, and the prediction result can also only be decrypted by the registered user, meanwhile the query result is consistent with that of un-privacy preserving scheme. The Detailed security analysis demonstrates our proposed protocol shows its security strength and privacy-preserving ability, and extensive experiments are conducted to demonstrate its efficiency.

In this work, connection is established among the client, cloud and server. Then, either the user who wants to classify his/her data or any business/institutions that needs classified data send their unprocessed data to SVM for classification. Initially, the users select the file which contains the primitive form data and then uploaded. The task of file selection and transferring has been takes place in the user's side. These unprocessed data flows from user's side to the server's side for classification. During the transfer stage, there is a chance of hacking the client's private data by intruders. So, our proposed Novel Based Protocol helps in preventing the client's data. Light Weight Multiparty Random Masking protocol is used for encrypting the data and the encrypted data been transferred from user's side to the server's side for classification. Support Vector Machine in the Server side classifies the encrypted data. The classification results then gets stored in the cloud for the retrieval process of user's need. The important task is to prevent the security of data. Here, Polynomial Aggregation Protocol prevents the data leakage and maintaining the privacy data in a secure way.

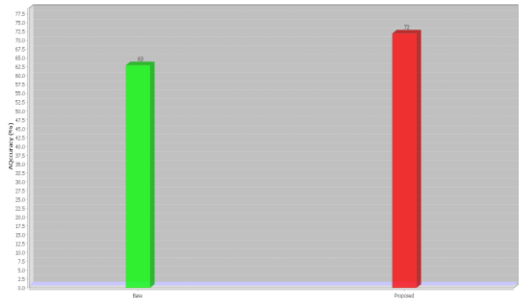
5. RESULT AND ANALYSIS

The performance is tested with two real time data set in terms of accuracy and time. The result shows the proposed privacy preserving method perform efficiently better than the conventional methods.

Accuracy: The accuracy gives true results in proportion (both true positives and true negatives) among the total number of cases examined. Accuracy can be calculated from formula given as follows:

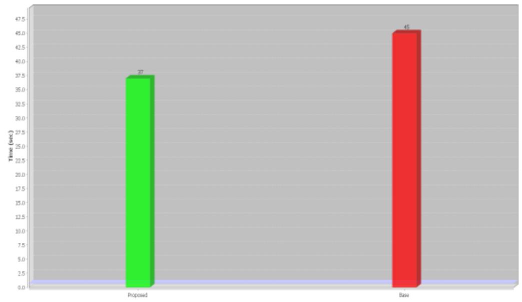
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy



Time: The indefinite continued progress of existence and time taken for classify the encrypted data.

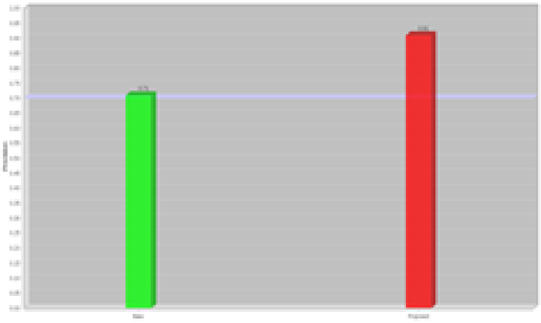
Time



Precision- Precision value is evaluated according to the feature classification at true positive prediction; false positive.It is expressed as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

Precision



Recall- Recall value is evaluated according to the feature classification at true positive prediction, false negative. It is given as,

$$\text{Recall} = \frac{\text{Truepositive}}{(\text{Truepositive} + \text{Falsenegative})}$$

6. CONCLUSION

In the following paper, soundness and leakage was the main issue and apart from that encryption was carried out using secure sign protocol which mainly focused on the matching of the signs to classify and then label the data. In the proposed model we use light weight multiparty random masking protocol which does not require any matching of the key, also classification is easier .Therefore, comparing with the earlier paper we achieve more accuracy ,efficiency and assuring no leakage of the data possible by any means.

REFERENCE

[1] Xingxin Li, Youwen Zhu, Jian Wang, Zhe Liu, Yining Liu, Mingwu Zhang On the Soundness and Security of Privacy-Preserving SVM for Outsourcing Data Classification IEEE Transactions on Dependable and Secure Computing (Volume: PP, Issue: 99) March 2017

[2] Yu, Y., Li, Y., Yang, B., Susilo, W., Yang, G., & Bai, J. (2017). Attribute-Based Cloud Data Integrity Auditing for Secure Outsourced Storage. IEEE Transactions on Emerging Topics in Computing.

[3] Liu, X., Choo, R., Deng, R., Lu, R., & Weng, J. (2016). Efficient and privacy-preserving outsourced calculation of rational numbers. IEEE Transactions on Dependable and Secure Computing.

[4] Malina, L., & Hajny, J. (2013, July). Efficient security solution for privacy-preserving cloud services. In Telecommunications and Signal Processing (TSP), 2013 36th International Conference on (pp. 23-27). IEEE.

[5] Omer, M. Z., Gao, H., & Sayed, F. (2016, November). Privacy Preserving in Distributed SVM Data Mining on Vertical Partitioned Data. In Soft Computing & Machine Intelligence (ISCMI), 2016 3rd International Conference on (pp. 84-89). IEEE.

[6] Rahulamathavan, Y., Phan, R. C. W., Veluru, S., Cumanan, K., & Rajarajan, M. (2014). Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud. IEEE Transactions on Dependable and Secure Computing, 11(5), 467-479.

[7] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[8] I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[9] K. Elissa, “Title of paper if known,” unpublished.

[10] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[11] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[12] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989