A Study of Privacy Preserving Technique Using K-nn Classification

Miss Rashmi S. Kodane[1], Prof. Karuna G. Bagde[2]
*[1]PG Student, Department of Computer Science & Information Technology,
H.V.P.M.C.O.E.T. Amravati, India*
*[2]Associate Professor, Department of Computer Science,
H.V.P.M.C.O.E.T. Amravati, India*

---***---

Abstract - *Due to the increase in sharing sensitive data through networks among businesses, governments and other parties, privacy preserving has become an important issue in data mining and knowledge discovery. Privacy concerns may prevent the parties from directly sharing the data and some types of information about the data. This paper proposes a solution for privately computing data mining classification algorithm for database without disclosing any information about the sources or the data. The proposed method (PPDM) combines the advantages of AES cryptosystem encryption scheme. PPDM method is robust in terms of privacy, accuracy, and efficiency. Data mining has been a popular research area for more than a decade due to its vast spectrum of applications. However, the popularity and wide availability of data mining tools also raised concerns about the privacy of individuals. The aim of privacy preserving data mining researchers is to develop data mining techniques that could be applied on databases without violating the privacy of individuals. To the best of our knowledge, our work is the first to develop a Privacy preserving data mining over encrypted data using AES algorithm.*

Keywords: **Privacy preserving, Security, AES algorithm, Encryption, PPDM**

1. Introduction

Rapid growth of online services has increased the opportunities to store private or confidential information. Data mining enables to exploit valuable knowledge from data collections while data miners are forced to treat such private datasets under prudent control in order to prevent the leakage or misuse. . Privacy concerns may prevent the parties from directly sharing the data and some types of information about the data. This data needs to be mine for various real time and other type of applications like banking, medicine, scientific research and among government agencies. Data Mining is a wider area now days due to the necessity of knowledge discovery from a different perspective on a very large scale. One of the commonly used tasks in data mining applications is the Classification.

Considering such situations, privacy preserving data mining (PPDM) provides a secure way to compute the output of a particular algorithm to distributed datasets without sharing of cloud computing, users now have the opportunity to outsource their data, in encrypted form, as well as the data mining tasks to the cloud. Since the data on the cloud is in encrypted form, existing privacy-preserving classification techniques are not applicable. This paper aims to solve the classification problem on encrypted data by using Cryptography based PPDM (Privacy Preserving Data Mining). To provide a better security, we propose AES algorithm.

2. Literature Review

Privacy Preserving Data Mining (PPDM) is defined as the process of extracting/deriving the knowledge about data without compromising the privacy of data. [4] In the past decade, many privacy-preserving classification techniques have been proposed in the literature in order to protect user privacy. Agrawal and Srikant introduced the notion of privacy-preserving under data mining applications. In particular to privacy preserving classification, the goal is to build a classifier in order to predict the class label of input data record based on the distributed training dataset without compromising the privacy of data.

Privacy preserving [6] has originated as an important concern with reference to the success of the data mining. Privacy preserving data mining (PPDM) deals with protecting the privacy of individual data or sensitive knowledge without sacrificing the utility of the data. People have become well aware of the privacy intrusions on their personal data and are very reluctant to share their sensitive information. This may lead to the inadvertent results of the data mining. Within the constraints of privacy, several methods have been proposed but still this branch of research is in its infancy.

In Data Perturbation methods, values of individual data records are perturbed by adding random noise in such way that the distribution of perturbed data look very different from that of actual data. After such a transformation, the perturbed data is sent to the miner to perform the desired data mining tasks. Agrawal and Srikant proposed the first data perturbation technique to build a decision-tree classifier. Since then many other randomization based methods have been proposed in the literature such as [5].

Data Distribution methods assume the dataset is partitioned either horizontally or vertically and distributed across different parties. The parties later can collaborate to securely mine the combined data and learn the global data mining results. During this process, data owned by individual parties is not revealed to other parties. This approach was first introduced by Lindell and Pinkas who proposed a decision tree classifier under two-party setting. Since then much work has been published using secure multiparty computation techniques [1].

3. Proposed system design

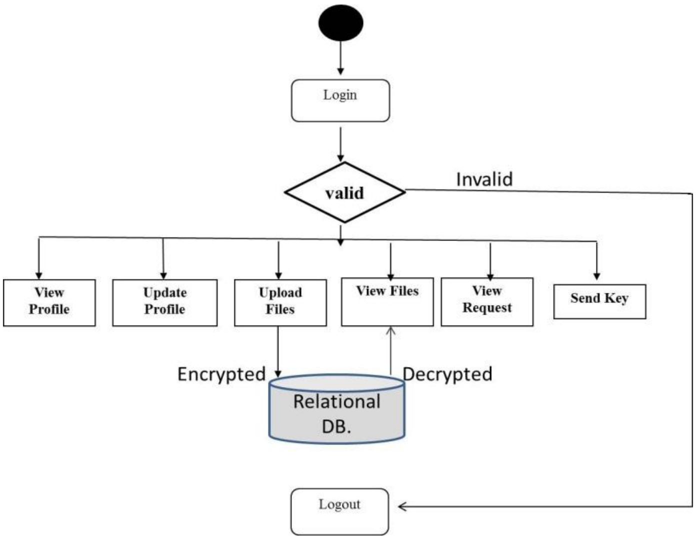
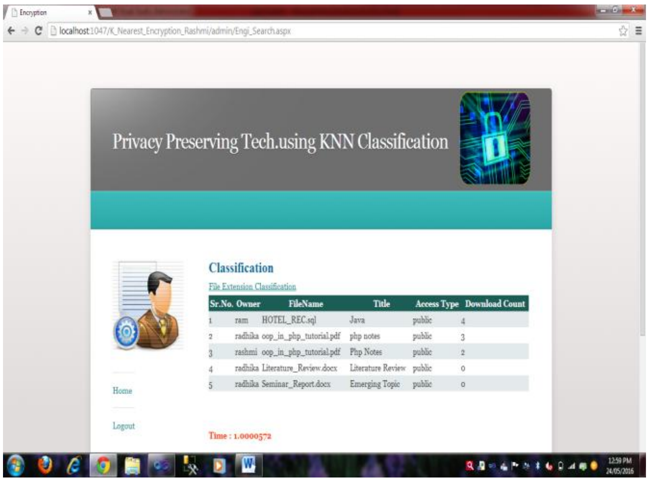


fig.1 DFD for Data owner

In this data owner first register here and with the help of user id and password they can login. If login is valid they can view their profile, update profiles, upload file. If the login is invalid then the user wants to log out. The users from the different fields such as finance, register here. The password in encrypted format is generated. Once the users are login they can view their profile, upload file, and search the file. A key is generated at the time when user uploads a file. When new user search the file uploaded by another user he will see the file name but when he want to download it the notification will send to the data owner and also the key request is sent.

4. Result

We compare existing method K-NN (K Nearest Neighbor) with the proposed method PPDM (Privacy preserving data mining) by using time as a parameter. In the PPDM method for data time require is 1.0000572 second and K-NN method requires 1.5000858.So, here time require is large as compared to proposed method. From this we conclude our method is best


Encryption
localhost:1047/K_Nearest_Encryption_Rashmi/admin/Engi_Search.aspx
Privacy Preserving Tech.using KNN Classification
Home
Logout

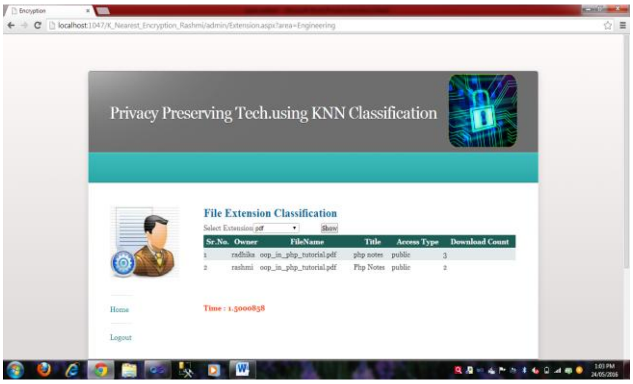
Classification

File Extension Classification

Sr.No.	Owner	FileName	Title	Access Type	Download Count
1	ram	HOTEL_REC.sql	Java	public	4
2	radhika	oop_in_php_tutorial.pdf	php notes	public	3
3	rashmi	oop_in_php_tutorial.pdf	Php Notes	public	2
4	radhika	Literature_Review.docx	Literature Review	public	0
5	radhika	Seminar_Report.docx	Emerging Topic	public	0

Time : 1.0000572

Fig.2 classification of engineering field using using PPDM


Encryption
localhost:1047/K_Nearest_Encryption_Rashmi/admin/Extension.aspx?area=Engineering
Privacy Preserving Tech.using KNN Classification
Home
Logout

File Extension Classification

Select Extension pdf Show

Sr.No.	Owner	FileName	Title	Access Type	Download Count
1	radhika	oop_in_php_tutorial.pdf	php notes	public	3
2	rashmi	oop_in_php_tutorial.pdf	Php Notes	public	2

Time : 1.5000858

Fig.3 classification of engineering data by k-nn

5. Conclusion

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. The existing techniques are not applicable to outsourced database environments where the data resides in encrypted form on cloud. To meet the security goals for the outsourcing data it should be outsourced as encrypted. Classification is an important task in many data mining applications. To protect user privacy, various privacy-preserving classification techniques have been proposed for the past decade. Nevertheless, the existing techniques are not applicable in outsourced database environment where the data resides in encrypted form on a third-party server. In this paper, a privacy-preserving data mining technique has been presented. Proposed method based on AES cryptographic algorithm which is semantically secured. This technique includes secure multiparty computation where a computation is secure if at the completion of the computation, no one can know anything except its own input and the results. Experimental results show that PPDM has good capability of privacy preserving, accuracy and efficiency, and relatively comparable to classical approach. Also, we will investigate and extend our research to other classification algorithm.

References:

[1] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techn.: Adv. Cryptol., 2011, pp. 129–148.

[2] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing private queries over untrusted data cloud through privacy homomorphism," in Proc. IEEE 27th Int. Conf. Data Eng., 2011, pp. 601–612.

[3] Y. Huang, D. Evans, J. Katz, and L. Malka, "Faster secure twoparty computation using garbled circuits," in Proc. 20th USENIX Conf. Security, 2011, pp. 35–35.

[4] R. Agrawal and R. Srikant. Privacy-preserving data mining. In ACM Sigmod Record, volume 29, pages 439–450. ACM, 2000.

[5] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.

[6] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," eprint arXiv:1403.5001, 2014.

[7] B. Yao, F. Li, and X. Xiao. Secure nearest neighbor revisited. In Proceedings of 29th IEEE International Conference on Data Engineering (ICDE), Brisbane, Australia, April 2013.

[8] M. Li, S. Yu, W. Lou, and Y. T. Hou. Toward privacy-assured cloud data services with flexible search functionalities. In 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), pages 466–470. IEEE, 2012.

[9] Y. Lindell and B. Pinkas. Secure multiparty computation for privacy-preserving data mining. Journal of Privacy and Confidentiality, 1(1):5, 2009.