
ISSN: 2319-345X
IJMRBS
International Journal of
Management Research and
Business Strategy
www.ijmrbs.org
E-mail
editor@ijmrbs.org
editor.ijmrbs@gmail.com

International Journal of Management Research and Business Strategy

ISSN 2319-345X www.ijmrbs.com

Vol. 9, Issuse. 4, Dec 2021

Classification o f Semantically Secure Encrypted Relational Data Usin gk-NN: A Case Study Using Drive HQ

VamshiRayabarapu1,NayaniSateesh2

Abstract: Applications for data mining can be found in a wide range of fields, including medicine, banking, education, and scientific research. In a variety of data mining applications, classification is the most commonly utilized data mining task. Various security methods have been proposed to solve the categorization challenge as a result of the rise in privacy concerns. When data in the cloud is encrypted, existing classification procedures that protect privacy are no longer applicable. A k-NN classifier is used instead of privacy-preserving approaches in this work to solve the classification problem over encrypted data.

Keywords: Encryption, outsourced databases, and k-NN classifiers are all aspects of security.

INTRODUCTION

Innovative methods of data storage and dissemination are supported by today's digital infrastructure [1]. In fact, we can store our data in remote servers, utilize dependable and efficient services provided by third parties, and leverage computer power available in different locations across the network. When it comes to the cloud, many companies are interested in its flexibility, cost effectiveness, and ability to unload administrative burdens. Protection and security are important concerns for enterprises despite the enormous benefits provided by the cloud. When the information is highly confidential, it should be encrypted before being sent to the cloud. Data mining tasks become increasingly difficult when information is encrypted.

An insurance company may have outsourced its encrypted client database to a cloud. A categorization technique can be used by an agent of the business to determine the risk level of a possible new client. As a first step, we need to establish an information record q (let's say) for the customer that contains specific personal information about the client. A class label can be assigned to the record q later on if it is delivered to the cloud. Data in q should be encrypted before being uploaded to the cloud since it contains confidential information. For the sake of illustration, we'll utilize a cloud-based example of data mining over encrypted data (abbreviated DMED) to demonstrate the importance of safeguarding individual records when used in data mining. If the

1PGScholar,CSEDepartment,CVRCollegeofEngineering,Hyderabad.
2AssistantProfessor,ITDepartment,CVRCollegeofEngineering,Hyderabad.

information is encrypted, however, the cloud can still deduce relevant and secret information about the data itself by watching the information access patterns. A cloud-based solution to the DMED problem requires three levels of security and privacy: confidentiality of encrypted data, protection of user queries, and concealment of information access patterns..

An insurance company may have outsourced its encrypted client database to a cloud. A categorization technique can be used by an agent of the business to determine the risk level of a possible new client. As a first step, we need to establish an information record q (let's say) for the customer that contains specific personal information about the client. A class label can be assigned to the record q later on if it is delivered to the cloud. Data in q should be encrypted before being uploaded to the cloud since it contains confidential information. For the sake of illustration, we'll utilize a cloud-based example of data mining over encrypted data (abbreviated DMED) to demonstrate the importance of safeguarding individual records when used in data mining. If the information is encrypted, however, the cloud can still deduce relevant and secret information about the data itself by watching the information access patterns. A cloud-based solution to the DMED problem requires three levels of security and privacy: confidentiality of encrypted data, protection of user queries, and concealment of information access patterns..

1. **RELATEDWORK**

De Capitani di Vimercati [2] outlined the different aspects of security in developing contexts. Additionally, the risks, arrangements, and unresolved difficulties linked to ensuring the security of clients accessing cloud services or assets, sensitive data stored at external gatherings, and the access to such data, were represented.

In a novel realistic oblivious data access protocol, P. Williams, R. Sion, and B. Carbunar [3] demonstrated the accuracy. Key findings include novel constructions and sophisticated reshuffle protocols that result in practical computational complexity of order O (log n log n), as well as storage overheads of order O (n). Additionally, a first practical implementation was presented that enables orders of magnitude faster query processing on many databases, with full computational privacy and correctness for many queries per second than existing approaches.

In [9], Bharath K. Samanthula [9] introduced a new PPkNN protocol, a secure k-NN classifier for semantically secure encrypted data. Privacy-protection measures that are both effective and efficient must be in place in order for cloud computing to be widely adopted. However, addressing the issue of privacy in the cloud is anything but straightforward, and it's a huge undertaking for everyone involved in the field. As a result, the author suggested a few solutions for dealing with the security issues that arise in a cloud computing environment.

C. Encryption schemes that are completely homomorphic can be used to search encrypted data, according to Gentry [10]. It is possible for users to analyze circuits over encrypted data, even if the server is unable to decrypt the files on its own.

There are three novel generalization approaches for k-anonymity introduced in [11] by R. J. Bayardo and R. Agrawal. They supply the new generalization schemes with enumeration algorithms, pruning criteria, and strategies for locating the best anonymization based on the discernibility metric, among other things.

Two-party decision tree classifiers proposed by Lindell and Pinkas [12] were the first to assume data distribution between them when they proposed their decision tree classifier. A great deal of research has been done since then employing SMC methods. Due to the fact that our data is encrypted and not transferred in plaintext across various parties, we cannot use data distribution techniques to tackle the PPkNN problem

2. **FRAMEWORK**

1. AuthorUploadthefiles
2. Initializetheindextermandthevalueofk

3. Togetthenearestneighborstocalculatet heEuclideandistancebyusingtheIndexIDoftrainingdata.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Wherek=numberofnearestneighbors

4. Sortthedistancesinascendingorder.
5. Getthek-nearestneighbors.

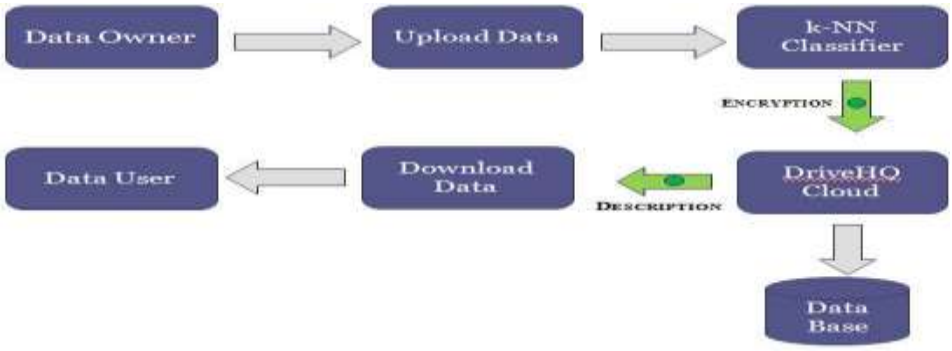

Fig.1:SystemArchitecture

The data owner is responsible for uploading the files, which will include file properties such as file name, index term, and description. The owner of the data will encrypt the file before transferring it to the cloud. The file is encrypted using AES (Advanced Encryption Standard). The files will be categorized and saved in the cloud based on the index term's similarity measurement. After a file is uploaded to the cloud, the owner of the data is no longer involved in any further calculation. As a result, the data holder will not be privy to any information. The index term of the file and the required number of nearest neighbors will be used by the data user to conduct the search. Based on the query data, the user will be presented with a list of the most relevant files in the database. In addition, the following privacy standards are met by this document: Information about the contents of Data or any intermediate outcomes should be kept private in the cloud. No one should be able to see what a data user is trying to do in the cloud. The data owner and the cloud service provider should not be made aware of access patterns, such as the records belonging to the k-next neighbors (to prevent any inference attacks).

The rest of the paper is laid out in this way. In Section 2, we go over the previously published related work. Section 3 lays forth the overall structure of the paper. Section 4 features the design elements. Section 5 presents the findings of the experiments. Lastly, in Section 6, we wrap up our report and discuss our plans for the future.

4.DESIGNELEMENTS

Thispapercontainsfollowingbasicelements,

i. Dataowner
ii. Data user
iii. Cloud
iv. Verifier

i. **Dataowner**

Dataownerwilluploadthefileintothecloud.Beforeuploadingthefile,dataownerwillencryptthedata forconfidentiality.Toaccessthedatafromthecloud,datauserrequiresthedataowner'spermission.

ii. **Datauser**

To use the cloud environment, data users must be registered in the cloud. The secret key1 [OTP1] will be sent to the user's email address once registration has been completed

successfully. Data users can access the cloud only by entering this secret key1. The owner of the data must give the data user permission to access it in the cloud. As a result, the data user will ask the data owner for access to the files. A user with secretkey2 [OTP2] will then be granted access to the data by the data owner through email. The file's index word will be used by the user to find and download the contents in the file. Afterwards, KNN was used to locate the appropriate cloud files. Next, the user can use Secretkey2 to select the desired file and download it from the cloud as needed.

iii. **Cloud(DriveHQ)**

Thecloud willstorethefiles,whichareuploadedbythedata owner.Tousethecloud environmentData users should be registered in the cloud.

iv. **Verifier**

Theverifierwillverifythedatabeforeuploadingintothecloud.

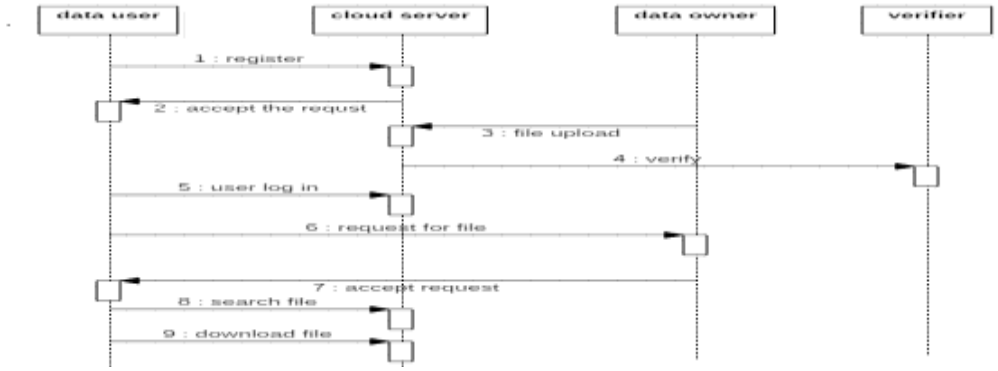

Fig.2:Sequencediagramoftheprocess

5. EXPERIMENTALRESULTS

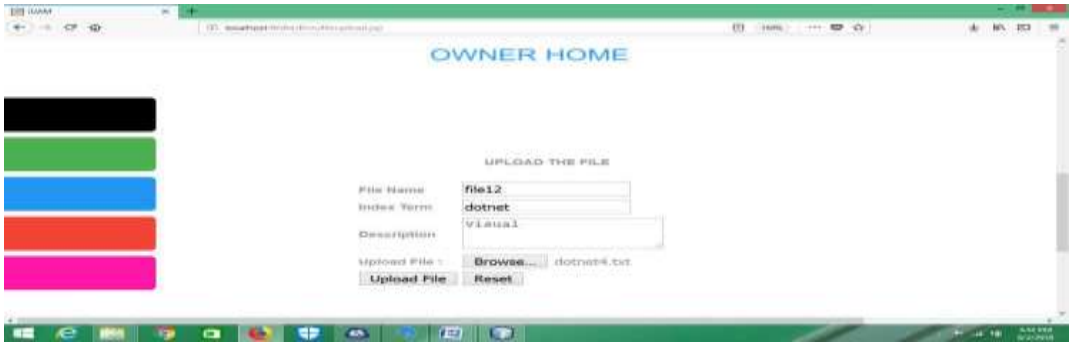

Fig.3:Fileupload

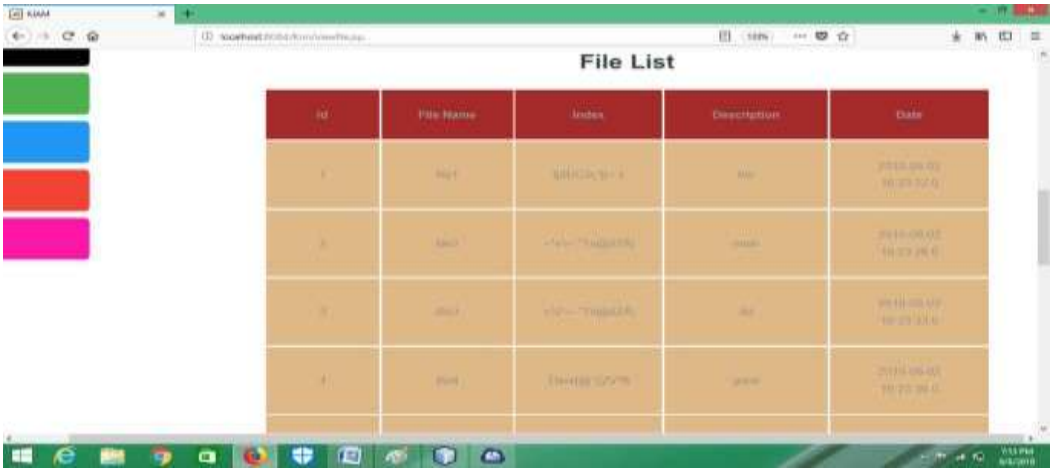

Fig.4:ListofFiles
Fig.5:search

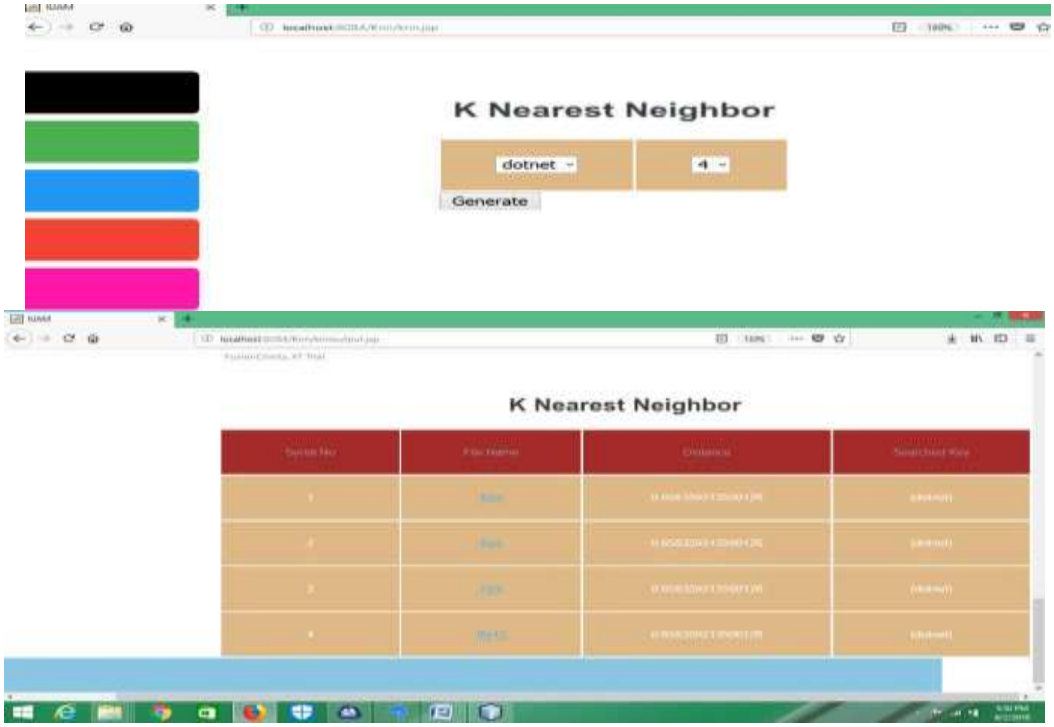

Fig.6:searchresult

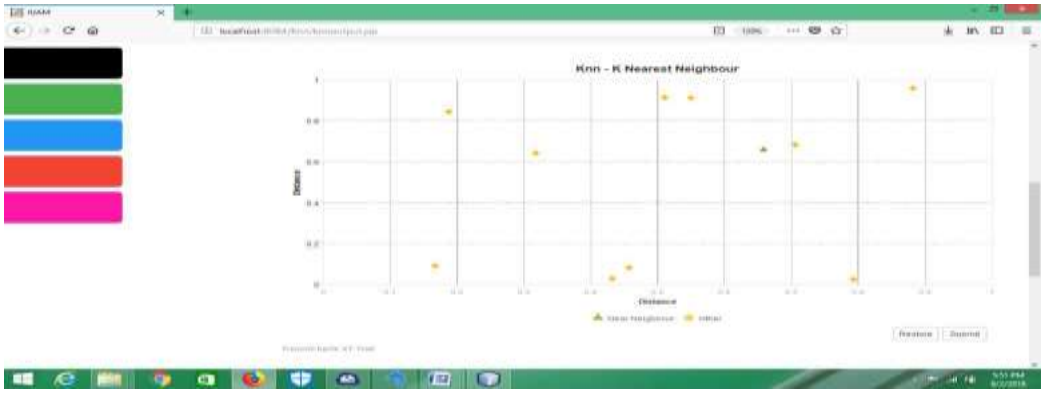

Fig.7:Distancemeasurementgraph

6. CONCLUSIONS

Several privacy-preserving methods have been presented in the last decade to protect the privacy of users. Outsourced database conditions, where information is encrypted and kept in the cloud, exclude the use of the current procedures. Each method of classification has its own advantages and disadvantages. Semantically secure encrypted relational data is the primary emphasis of this paper's classifier. Protects the user's privacy by obscuring data access patterns as well as data confidentiality. Other classification

approaches on semantically safe encrypted relational data can be used in the future.

7. ACKNOWLEDGEMENTS

Theauthorsthanktoallthepeoplewhoinspiredfortheinitiativeofthisworkandtothepeople who directly or indirectly supported all the times.

REFERENCES

[1] Cloud computing is defined by NIST Special Publication 800, volume 145, p. 145 by P. Mell and T. Grance.

[2] Management and access to data in the cloud: privacy hazards and approaches," in Proceedings of the 7th International Conference on Risk and Security of the Internet and Systems, pp. 1–9, 2012. S. De Capitani di Vimercati, P. Samarati, and S. Foresti.

[3] To build a castle from mud: Practical access pattern privacy and correctness on untrusted storage, P. Williams, R. Sion and B. Carbunar, Proceedings of the 15th ACM Conference on Computer and Communications Security, p.139–148, 2012. [3]

[4] There are several ways to protect the privacy of your data, but one of the most common is to use naive Bayes classification, which is based on the assumption that all information is equal.

[5] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," Information Systems, volume 29, pp. 343–364, 2004. [5].

[6] Privacy homomorphism is used to process private queries over untrusted data clouds in Proceedings of the IEEE 27th International Conference on Data Engineering, pp. 601–612, 2011. Hu, J. Xu, C. Ren, and B. Choi

[7] In Proceedings of 8th European Conference on Principles Practice Knowledge Discovery in Databases, pp. 279–290, 2004, M. Kantarcioglu and C. Clifton present "Privately constructing a distributed k-nnclassifier."L. Xiong, S. Chitti, and L. Liu, “K nearest neighbor classification across multiple private databases,” inproceedingsofthe 15th ACM international conference on Information Knowledge Management, pp. 840–841, 2006.

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

[9] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in Proceedings of 41st Annual ACM Symposium Theory of Computing,pp. 169–178, 2012.

[10] R. J. Bayardo and R. Agrawal, “Data privacy through optimal k-anonymization,” in Proceedings of IEEE 21st International Conference on Data Engineering pp. 217–228, 1979.

[11] Y.LindellandB.Pinkas,“Privacypreservingdatamining,”inProceedingsof20thAnnualInternationalCryptology Conference pp. 36–54, 2000.