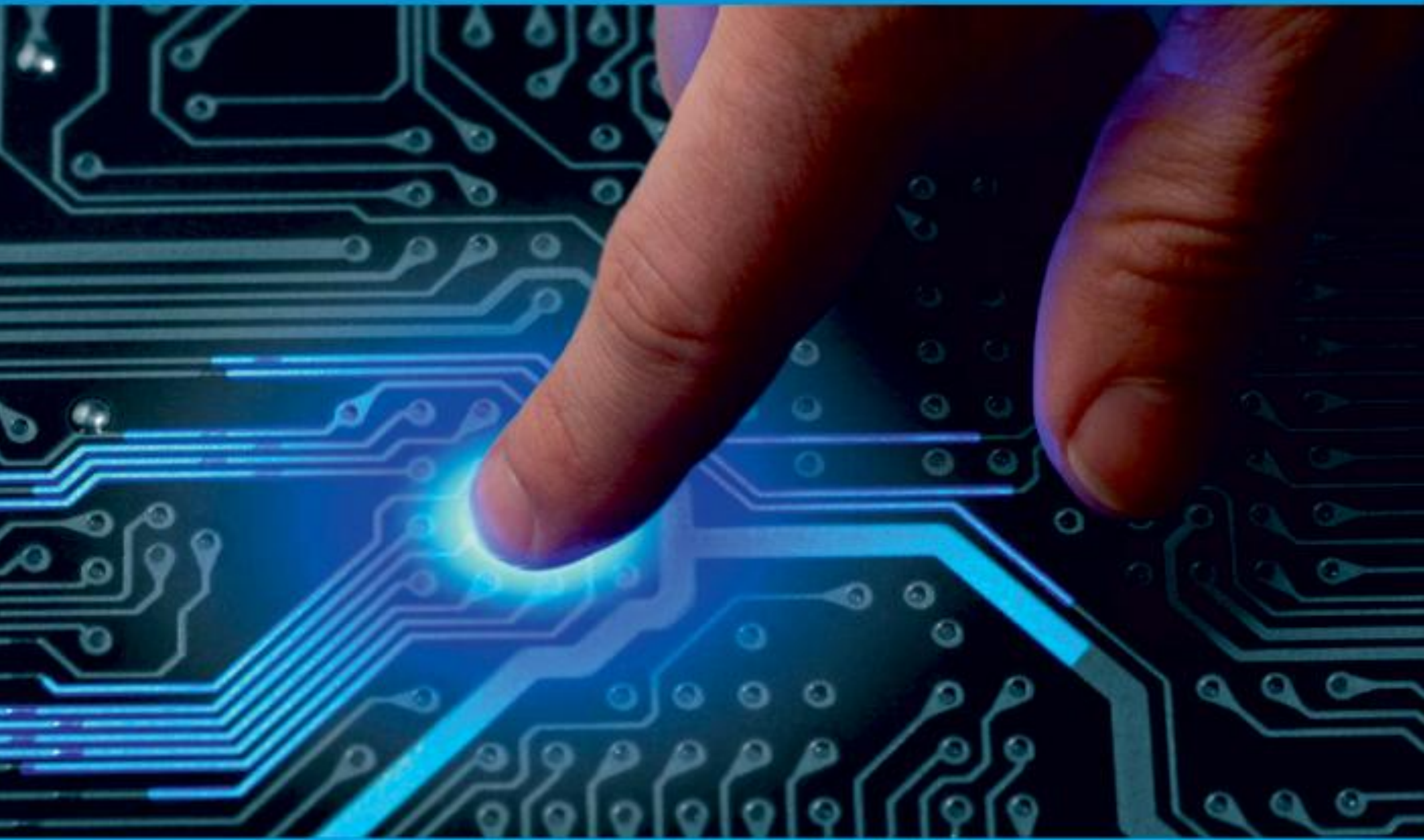
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542

9940 572 462 6381 907 438 ijircce@gmail.com www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906296 |

K-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data

Nutan S. Shelke[1], Prof. Monika D. Rokade[2]

PG Student, Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India

Assistant Professor (ME Co-ordinator), Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India

ABSTRACT: Data mining has a wide range of uses in a variety of industries, including finance, medical, scientific research, and government agencies. One of the most common jobs in data mining applications is classification. Many theoretical and practical solutions to the categorization problem have been presented under various security models over the last decade as a result of the growth of various privacy issues. With the recent rise in popularity of cloud computing, customers can now outsource their data, in encrypted form, as well as data mining jobs to the cloud. Existing privacy-preserving classification approaches aren't relevant because the data on the cloud is encrypted. The classification problem over encrypted data is the topic of this paper. We present a secure k-NN classifier for encrypted data in the cloud, in particular. The suggested k-NN protocol safeguards the data's confidentiality, as well as the user's query and data access patterns. To our knowledge, this is the first time a safe k-NN classifier has been developed over encrypted data using the standard semi-honest model. In addition, we use a variety of studies to test the efficacy of our approach.

KEYWORDS: Security, k-NN Classifier, Outsourced Databases, Encryption, privacy preserving.

I.INTRODUCTION

The cloud computing paradigm has recently revolutionised how businesses manage their data, particularly in terms of data storage, access, and processing. Many firms are actively considering cloud computing as an emerging computing paradigm because of its cost-efficiency, flexibility, and offloading of administrative burden. Organizations frequently delegate their computing activities to the cloud in addition to their data. Despite the numerous benefits that the cloud provides, corporations are unable to take use of those benefits due to privacy and security concerns. When data is extremely sensitive, it must be encrypted before being sent to the cloud. However, when data is encrypted, regardless of the underlying encryption strategy, doing any data mining operations without first decrypting the data becomes extremely difficult. The data owner outsources his or her database and DBMS operations (e.g., kNN query) to an untrustworthy external service provider, who handles the data on the data owner's behalf and allows only trusted users to query the service provider's hosted data. Many security risks exist when data is outsourced to an untrustworthy server, such as data privacy (protecting the confidentiality of the data from the server as well as from query issuer). Before outsourcing his or her data to the server, the data owner must employ data anonymization models (e.g., k-anonymity) or cryptography (e.g., encryption and data perturbation) techniques to ensure data privacy. Encryption is a well-known method for safeguarding the privacy of sensitive data, such as medical records. The process of query evaluation over encrypted data gets difficult due to data encryption. Various strategies for processing range and aggregate queries over encrypted data have been developed in this direction. Encryption as a means of ensuring data secrecy may cause a problem during the cloud query processing stage. In general, processing encrypted data without needing to decrypt it is quite challenging. The question is how the cloud can conduct queries over encrypted data while the data is encrypted at all times in the cloud.

II.GOALS AND OBJECTIVE

- Improving SMINn's efficiency is a critical first step toward bettering the performance of our PPkNN technique.
- Our protocol safeguards the data's confidentiality, as well as the user's input query and data access behaviors.
- We also tested our protocol's performance with various parameter values.

| DOI: 10.15680/IJIRCCE.2021.0906296 |

III.MOTIVATION OF THE PROJECT

We used encrypted data to inspire the PPKNN in order to obtain Cloud Computing economies of scale. After that, we developed two new security primitives: secure minimum (SMIN) and secure frequency (SF), as well as new solutions for them. Second, there was no systematic security study of the underlying sub-protocols in the work. On the other hand, under the semi-honest paradigm, this study gives formal security proofs of the underlying sub-protocols as well as the PPkNN protocol. We demonstrate that our proposed technique is secure and privacy-preserving while achieving the PPKNN goal appropriately.

IV.EXISTING SYSTEM PROBLEM

Suppose Alice owns a database D of n records t1. . . tnand m + 1 attributes. Let ti,jdenote the jthattribute value of record ti. Initially, Alice encrypts her database attribute-wise, that is, she computes Epk(ti,j), for $1 \leq i \leq n$ and $1 \leq j \leq m+1$, where column (m+1) contains the class labels. We assume that the underlying encryption scheme is semantically secure . Let the encrypted database be denoted by D′. We assume that Alice outsources D′ as well as the future classification process to the cloud.Let Bob be an authorized user who wants to classify his input record q = hq1, . . . ,qmiby applying the k-Classification method based on D′. We refer to such a process as privacy-preserving k-NN (PPkNN) classification over encrypted data in the cloud. Formally, we define the PPkNN protocol as:

$$\text{PPkNN}(D', q) \rightarrow cq$$

Wherecqdenotes the class label for q after applying k-NN classification method on D′ and q.

V.LITERATURE SURVEY

1. Project Title: Survey on Privacy Preserving Data Mining From this paper We Referred

The extraction of interesting patterns or knowledge from large amounts of data is known as data mining. With the rapid advancement of Internet, data storage, and data processing technology in recent years, privacy preservation has become one of the most pressing challenges in data mining. For privacy-preserving data mining, a number of methodologies and strategies have been developed. This paper presents a comprehensive overview of various privacy-preserving data mining algorithms, as well as an analysis of sample strategies for privacy-preserving data mining and a discussion of their benefits and drawbacks. Finally, current issues and prospective research directions are reviewed.

2. Project Title: Proving in Zero-Knowledge that a Number Is the Product of Two Safe Primes. From this paper We Referred

We present the reticent statistical zero-knowledge protocols to prove statements such as:

- A committed number is a prime.
- A committed (or revealed) number is the product of two safe primes, i.e., primes p and q such that (p − 1)=2and (q − 1) =2areprime.
- A given integer has large multiplicative order modulo a composite number that consists of two safe prime factors.

3. Project Title: Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments. From this paper We Referred

Query processing on relational data has been intensively investigated over the last decade, with several theoretical and practical solutions to query processing suggested in a variety of circumstances. Users may now outsource their data as well as data management chores to the cloud, thanks to the rising popularity of cloud computing. Sensitive data (e.g., medical records) must be encrypted before being sent to the cloud, due to the rise of numerous privacy concerns. Furthermore, the cloud should perform query processing chores; otherwise, there would be no reason to outsource the data in the first place. It's a difficult task to perform queries over encrypted data without the cloud ever decrypting the data. The goal of this work is to solve the k-nearest neighbour (kNN) query problem using an encrypted database that has been outsourced to the cloud: a user sends an encrypted query record to the cloud, and the cloud returns the k closest records to the user. We begin by presenting a rudimentary system and demonstrating that such a simplistic approach is insecure. To improve security, we offer a secure kNN protocol that safeguards data confidentiality, user input query confidentiality, and data access patterns. In addition, we use a variety of tests to test the efficacy of our processes. These findings show that our secure protocol is very efficient on the user end, and that this lightweight method allows a user to do the kNN query on any mobile device.

4. Project Title: Managing and Accessing Data in the Cloud Privacy Risks and Approaches From this paper We Referred

One of our modern society's big concerns is ensuring sufficient privacy and protection of the information stored, communicated, processed, and distributed in the cloud, as well as the users who access such information. In fact, while advances in information technology and the spread of novel paradigms such as data outsourcing and cloud computing enable users and businesses to easily access high-quality applications and services, they also introduce new privacy risks associated with improper information disclosure and dissemination. We shall characterise many facets of the privacy challenge in developing scenarios in this study. We'll go over the dangers, remedies, and unsolved issues associated with maintaining the privacy of users accessing cloud services or resources, sensitive data maintained by third parties, and access to such data.

5. Project Title: Privacy-preserving data mining in the malicious model From this paper We Referred

The majority of cryptographic work in privacy-preserving distributed data mining is focused on semi-honest adversaries that are expected to follow the prescribed protocol but attempt to deduce private information from the messages they receive. Although the semi-honest paradigm makes sense in some situations, it is impractical to expect opponents to constantly follow the procedures to the letter. Malicious enemies, in instance, could diverge from their authorised protocols at will. Complex strategies are required for developing secure protocols against malevolent adversaries. Protocols that can survive malevolent adversaries are clearly more secure. However, there is a clear trade-off: protocols that are secure against malevolent adversaries are typically more expensive than those that are just secure against semi-honest opponents. Our purpose in this research is to compare the performance and security trade-offs in privacy-preserving distributed data mining techniques in the two scenarios. To make a meaningful comparison, we employ zero knowledge proofs to modify commonly used sub protocols that are secure in the semi-honest model to make them secure in the malicious model. In both models, we compare the performance of various procedures.

Monika Rokade and YogeshPatil [11] proposed a system deep learning classification using nomaly detection from network dataset. The Recurrent Neural Network (RNN) has classification algorithm has used for detection and classifying the abnormal activities. The major benefit of system it can works on structured as well as unstructured imbalance dataset.

The MLIDS A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset has proposed by Monika Rokade and Dr. YogeshPatil in [12]. The numerous soft computing and machine learning classification algorithms have been used for detection the malicious activity from network dataset. The system depicts around 95% accuracy ok KDDCUP and NSLKDD dataset.

Monika D. Rokade and Yogesh Kumar Sharma [13] proposed a system to identification of Malicious Activity for Network Packet using Deep Learning. 6 standard dataset has sued for detection of malicious attacks with minimum three machine learning algorithms.

Sunil S. Khatal and Yogeshkumar Sharma [14] proposed a system Health Care Patient Monitoring using IoT and Machine Learning for detection of heart and chronic diseases of human body. The IoT environment has used for collection of real data while machine learning technique has used for classification those data, as it normal or abnormal.

Data Hiding In Audio-Video Using Anti Forensics Technique For Authentication has proposed by Sunil S.Khatal and Yogeshkumar Sharma [15]. This is a secure data hiding approach for hide the text data into video as well as image. Once sender hide data into specific objects while receivers does same operation for authentication. The major benefit of this system can eliminate zero day attacks in untrusted environments.

Sunil S.Khatal and Yogesh Kumar Sharma [16] proposed a system to analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. This is the analytical based system to detection and prediction of heart disease from IoT dataset. This system can able to detect the disease and predict accordingly.

VI.METHODOLOGY USED

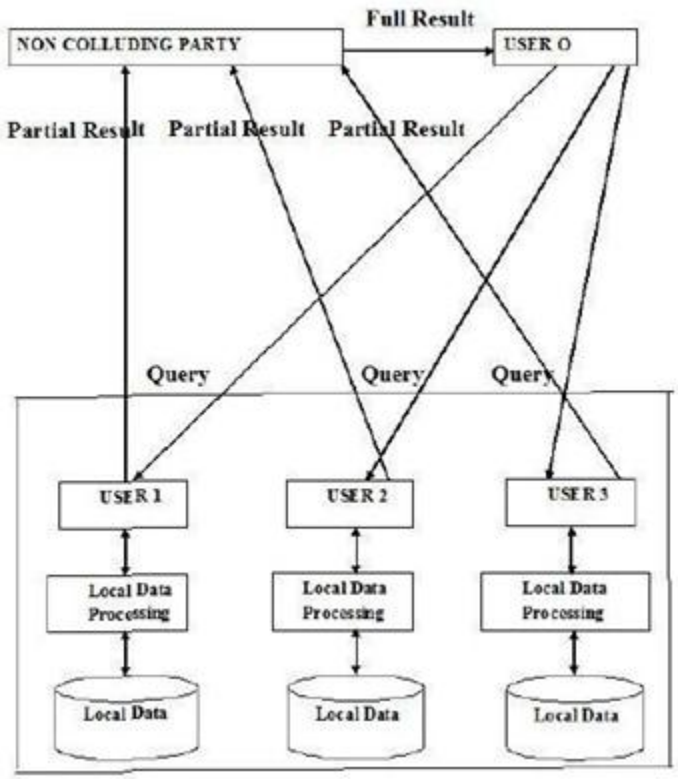


Fig. 1 System Architecture

1. Data confidentiality - Contents of T or any intermediate results should not be revealed to the cloud.
2. Query privacy - Bob’s input query Q should not be revealed to the cloud.
3. Correctness - The output (t’1……t’k) should be revealed only to Bob. In addition, no information other thant’1…..t’kshould be revealed to Bob.
4. Low computation overhead on Bob - After sending his encrypted query record to the cloud, Bob involves only in a little computation compared with the existing works. More details are given in Section.
5. Hidden data access patterns - Access patterns to the data, such as the records corresponding to the k-nearest neighbors of Q, should not be revealed to Alice and the cloud (to prevent any inference attacks).

ALGORITHM AND TECHNIQUE USED

- Search algorithm.

Our DMRS scheme's search method begins at the root node with a recursive operation that traverses the tree in a specific depth-first manner known as "Greedy Depth First Traverse Strategy." If the similarity score of the node is less than or equal to the minimum similarity score of the currently selected top-k documents, the search process returns to the parent node; otherwise, it proceeds to the child node. Formula (1), i.e. the inner product of query vector Q and data vector Du, is used to determine the similarity score of each node u. This approach is repeated until the objects with the highest k scores are chosen. Because of the rather accurate maximum score prediction, the search may be done quickly because just a portion of the index tree is visited. The procedure of our proposed search strategy is depicted in Algorithm 1.

- Secure algorithm.

To prevent information leaking, a safe encryption scheme must be implemented as soon as the plaintext index tree is generated. To secure our index tree, we use the encryption approach provided in [4], with the following steps:

- Setup
- GenIndex

- GenTrapdoor
- SimEvaluation

APPLICATIONS

It's employed in the cloud when we want to store data in an encrypted format. And then decrypt the data using the secret key. The k-NN classifier is well-known, and we've created a privacy-preserving mechanism for it over encrypted data.

- Increased website ROI

Users are more likely to perform the intended action on a website if they can readily locate what they're searching for, whether it's a product purchase, an information request, or simply learning what they wanted to know.

- Reduced customer service costs

The amount of calls or emails to customer care can be reduced by providing a self-service mechanism to access common information on a website. In addition, when addressing queries, customer service might use the same search.

- Increased productivity

If your company is like many others, you have file shares full of documents but aren't sure what's in them or where they are. By conducting a thorough search, you may quickly identify both the documents you need and any associated documents that may already exist, saving time and effort.

VII.EXPERIMENTAL SETUP AND RESULT

Here are several studies that show how the Privacy Preserving k-Nearest Neighbor (PPKNN) classification approach performs with different parameter choices. The suggested PPkNN protocol is implemented in JAVA using the Partial Homomorphic encryption scheme as the underlying additive homomorphic encryption scheme.

1 Dataset Details and Experimental Setup

The Car Evaluation dataset from the UCI KDD archive[12] was used in this technique. There are 1728 records in this dataset (n = 1728) and 6 attributes (m = 6). There is a distinct class attribute, and the dataset is divided into four categories (w = 4). This dataset was encrypted attribute-by-attribute using Homomorphic encryption, with the key size adjusted in tests, and the encrypted data saved on a server system. Executed a random query on this encrypted data using the PPkNN protocol.

2 Performance of k-Nearest Neighbor Classification with Partial Homomorphic Encryption for Privacy Preservation

When the encryption key size K is set to 512 or 1024 bits, the computation time ranges from 9.98 to 46.16 minutes, depending on whether K is set to 5 or 25 bits. When K=1024 bits, the computation time ranges from 66.97 to 309.98 minutes, depending on whether k is set to 5 or 25. When k is varied from 5 to 25, the computation time for Stage 2 to generate the final class label corresponding to the input query ranges from 0.118 to 0.285 seconds for K=512 bits. Stage 2 took 0.789 and 1.89 seconds for K=1024 bits when k = 5 and 25, respectively. The computation time of Stage 1 accounts for at least 99 percent of the entire time in PPkNN, as shown above. Stage 1 and 2 processing costs, for example, are 19.06 minutes and 0.175 seconds, respectively, when k = 10 and K=512bits. Stage 1 costs 99.98 percent of the total cost of PPkNN in this scenario. With n and k, the overall computation time of PPkNN climbs practically linearly.

RESULTS

We need to show that the simulated image of SMIN is computationally indistinguishable from the actual execution image of SMIN, as indicated in Section 2.3, to formally establish that SMIN is secure under the semi-honest model. The messages sent and the information generated from these messages are usually included in an execution image.

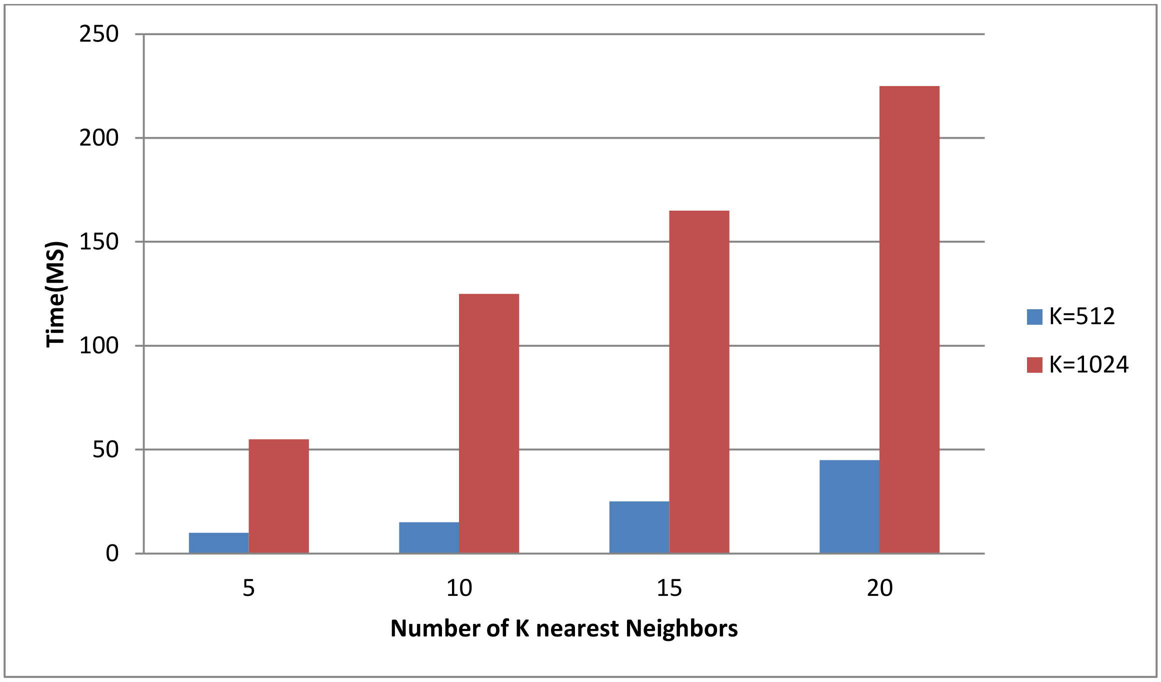


Fig No 2 Computation costs of PPkNN for varying number of k nearest neighbors and encryption key size K

VIII.CONCLUSION

Various privacy-preserving classification algorithms have been presented during the last decade to protect user privacy. Existing solutions are ineffective in outsourced database environments when data is stored on a third-party server in encrypted form. Over encrypted data in the cloud, this research developed a novel privacy-preserving k-NN classification system. Our protocol safeguards the data's confidentiality, as well as the user's input query and data access patterns. We have tested our protocol's performance using a variety of parameter settings. We plan to study alternate and more efficient solute-ions to the SMIN n problem in our future work because enhancing the efficiency of SMIN n is an important first step toward improving the performance of our PPkNN protocol. We'll also look into and expand our study into other classification algorithms.

REFERENCES

[1] Mell and T. Grance, “The nist definition of cloud computing(draft),” NIST special publication , vol. 800, p. 145, 2011.
[2] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, “Managing and accessing data in the cloud: Privacy risks and approaches,” in CRiSIS, pp. 1 –9, 2012.
[3] P. Williams, R. Sion, and B. Carbunar, “Building castles out of mud: practical access pattern privacy and correctness on untrusted storage,” in ACM CCS , pp. 139–148, 2008.
[4] P. Paillier, “Public key cryptosystems based on composite degree residuosity classes,” in Eurocrypt , pp. 223–238, 1999.
[5] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, “k-nearest neighbor classification over semantically secure encrypted re-lational data.” eprint arXiv:1403.5001, 2014.
[6] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in ACM STOC , pp. 169–178, 2009.
[7] C. Gentry and S. Halevi, “Implementing gentry’s fully- homomorphic encryption scheme,” in EUROCRYPT , pp. 129– 148, Springer, 2011.
[8] A. Shamir, “How to share a secret,”Commun.ACM, vol. 22, pp. 612–613, 1979.

[9] D. Bogdanov, S. Laur, and J. Willemson, “Sharemind: A framework for fast privacy-preserving computations,”in Proc. 13th Eur.Symp. Res. Comput. Security: Comput.Security, 2008, pp. 192–206.

[10] R. Agrawal and R. Srikant, “Privacy-preserving data mining,”ACMSigmod Rec., vol. 29, pp. 439–450, 2000.

[11] Monika D.Rokade ,Dr.YogeshkumarSharma,”Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic.”IOSR Journal of Engineering (IOSR JEN),ISSN (e): 2250-3021, ISSN (p): 2278-8719

[12] Monika D.Rokade ,Dr.YogeshkumarSharma”MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset”, 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE

[13]Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2324 - 2331.

[14] Sunil S.Khatal ,Dr.Yogeshkumar Sharma, “Health Care Patient Monitoring using IoT and Machine Learning.”, **IOSR Journal of Engineering (IOSR JEN),** ISSN (e): 2250-3021, ISSN (p): 2278-8719

[15]Sunil S.Khatal ,Dr.Yogeshkumar Sharma, “Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ", IJSRDV4I50349, Volume : 4, Issue : 5

[16]Sunil S.Khatal Dr. Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2340 - 2346.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com