IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

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 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907010 |

Secure *k*-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

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*:* 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's end, and that this lightweight approach allows a user to do the kNN query on any mobile device.

I. INTRODUCTION

Many firms are considering cloud computing as an emerging computing paradigm because of the cost-efficiency, flexibility, and offloading of administrative burden it offers. A data owner outsources his or her database T and DBMS operations to the cloud in the cloud computing model [1], [2], which has the infrastructure to host outsourced databases and provides access mechanisms for querying and controlling the hosted database. On the one hand, outsourcing benefits the data owner by lowering data administration expenses and improving service quality. Hosting and query processing of data that is not under the control of the data owner, on the other hand, poses security issues such as maintaining data confidentiality and query privacy.

One simple technique to preserve the confidentiality of outsourced data from the cloud and unauthorised users is for the data owner to encrypt the data before outsourcing [3]. The data owner can preserve the privacy of his or her own data in this way. Furthermore, authorised users must encrypt their queries before transmitting them to the cloud for review in order to protect query privacy. Furthermore, by watching the data throughout query processing, the cloud can deduce relevant and sensitive information about the actual data items.

Even if the data and query are encrypted, access patterns can be found [4], [5]. As a result of the preceding considerations, secure query processing must provide (1) the confidentiality of encrypted data, (2) the confidentiality of a user's query record, and (3) the concealment of data access patterns.

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. Various strategies for query processing over encrypted data, such as range queries [6]–[8] and other aggregate queries [9], [10], have been presented in the literature. However, for sophisticated searches like the k-nearest neighbour (kNN) query, these strategies are either not relevant or inefficient. The challenge of secure processing of k-nearest neighbour queries over encrypted data (SkNN) in

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

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907010 |

the cloud is addressed in this study. The goal of the SkNN challenge is to securely identify the k-nearest data tuples to Q utilising the encrypted database T in the cloud, without allowing the cloud to learn anything about the actual contents of the database T or the query record Q, given a user's input query Q. When encrypting data and sending it to the cloud, we discovered that an effective SkNN protocol must meet the following requirements:

- Maintaining T and Q's secrecy at all times
- Hiding data access patterns from the cloud
- Compute the k-nearest neighbours of query Q accurately
- Inflict minimal computing burden on the end-user

Researchers have proposed numerous methods [1], [11]–[13] to overcome the SkNN problem in recent years. However, we stress that present SkNN approaches lack at least one of the desirable features of a SkNN protocol described above. On the one hand, the approaches in [1], [11] are unsafe since they are vulnerable to plaintext assaults that are chosen and known. The recent method in [13], on the other hand, gives a non-accurate kNN result to the end-user. Instead of obtaining the encrypted exact k-nearest neighbours, the cloud in [13] obtains the relevant encrypted partition. Furthermore, the end-user is involved in significant computations during the query processing stage in [1], [12], and [13]. By doing so, these solutions treat the cloud as as a storage medium, requiring no additional effort.

II. LITERATURE SURVEY

In this section, we present an overview of the existing secure k-nearest neighbor techniques. Then, we discuss the security definition adopted in this paper along with the homomorphic properties of the Paillier cryptosystem as a background.

A. Existing SkNN Techniques

One of the most fundamental problems in many application domains, such as similarity search, pattern recognition, and data mining, is locating the k-nearest neighbours to a given query Q. Many solutions have been presented in the literature to address the SkNN problem, which can be divided into two categories based on whether or not the data is encrypted: centralised and distributed.

Centralized Methods: We assume that the data owner outsources his or her database and DBMS operations (e.g., kNN query) to an untrusted external service provider who handles the data on behalf of the data owner and only allows trustworthy users to query the 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 and 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 traditional technique used to protect the confidentiality of sensitive data such as medical records. Due to data encryption, the process of query evaluation over encrypted data becomes challenging. Along this direction, various techniques have been proposed for processing range [6]–[8] and aggregation queries [9], [10] over encrypted data. However, in this paper, we restrict our discussion to secure evaluation of kNN query.

Monika Rokade and Yogesh Patil [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. Yogesh Patil 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 Yogesh kumar 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

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

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907010 |

and Yogesh kumar 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.

Data Distribution Methods: Data is expected to be partitioned vertically or horizontally and dispersed among a group of independent, non-colluding parties in data distribution methods. Secure multiparty computing (SMC) techniques enable many parties to safely evaluate a function using their respective private inputs without disclosing the input of one party to the others, according to the literature. Many attempts have been made to solve the kNN query problem in a distributed context. Shaneck et al. [18] suggested a privacy-preserving k-nearest neighbour search technique. The protocol described in [18] uses secure multiparty computation to compute kNN points in a horizontally partitioned dataset privately. Qi et al. [19] suggested a single-step kNN search protocol with linear computing and communication complexity that is provably safe. Vaidya et al. [20] investigated privacy-preserving top-k queries in vertically partitioned data. Ghinita et al. [21] developed a PIR-based architecture for answering kNN queries in location-based services to answer kNN inquiries. We underline that the data on the server in [21] is in unencrypted format. However, if the data is encrypted to maintain data security, it is unclear how a user can obliviously access the output records if he or she is unaware of the indexes that match his or her input query. Even though a user can get records using PIR, the user must still do local computations in order to find the k-nearest neighbours. Our system, on the other hand, totally outsources the user's processing to the cloud..

In conclusion, the previous data distribution strategies are ineffective for performing kNN queries over encrypted data for two reasons: (1). We deal with encrypted database and query forms in our work, which is not the case with the above approaches (2). In our situation, the database is encrypted and stored in the cloud, whereas it is partitioned (in plaintext format) among different parties in the aforementioned techniques.

B. Security Definition

The amount of information released during the execution of a protocol is intimately related to privacy/security in this paper. Information disclosure can be defined in a variety of ways. We use the security definitions in the literature of secure multiparty computation (SMC) first established by Yao's Millionaires' issue for which a provably secure solution was developed [14] to maximise privacy or minimise information disclosure. In this work, we assume that parties are semi-honest (or honest-but-curious), which means that a semi-honest party follows the protocol's rules with valid input but is free to utilise what it sees during execution to compromise security afterwards. In general, secure protocols based on the semi-honest model are more efficient than malicious adversary protocols, and practically all real SMC protocols proposed in the literature are secure under the semi-honest model. We recommend the reader to [14] for thorough security definitions and models due to space constraints.

III. OUR CONTRIBUTION

We offer a unique SkNN protocol to support k-nearest neighbour search over encrypted data in the cloud while maintaining both data and query privacy in this work. Alice does not engage in any computations in our protocol once the encrypted data are outsourced to the cloud. As a result, Alice receives no information. The suggested protocol, in particular, satisfies the following conditions.

- **Data confidentiality** -Contents of T or any intermediateresultsshouldnotberevealedtothecloud.
- **Query privacy** -Bob's input query Q should not berevealedtothecloud.
- **Correctness** -The output$t'_1, \ldots, t'_k$should be revealedonlytoBob.In addition, no information other than $t'_1, \ldots, t'_k$should be revealed to Bob.
- **Low computation over head on Bob**-Aftersendinghis encrypted query record to the cloud, our protocolsincur low computation overhead on Bob compared withtheexistingworks[1],[11]–[13].
- **Hidden data access patterns** -Access patterns to thedata, such as the records corresponding to the k-nearestneighbors of Q, should not be revealed to Alice and thecloud(topreventanyinferenceattacks).

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

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907010 |

VI. RESULTS AND DISCUSSION

In this section, we go over the proposed protocols' performance in depth under various parameter settings. The suggested protocols were written in C using the Paillier cryptosystem [15]. Various tests were carried out on a Windows computer running with an Intel i3 CPU 2.07 GHz processor and 12GB RAM.

Towards Performance Improvement

The proposed techniques appear to be expensive at first glance, and they may not scale well for huge datasets. However, we underline in both protocols that the computations performed on each data record are independent of the others. As a result, we can parallelize operations on data records for increased efficiency.

To back up this assertion, we used OpenMP programming to create a parallel version of our SkNNb protocol and compared its computation costs to the serial version. Our machine, as previously said, has six cores that can be used to do parallel tasks on six threads. Figure 3 shows the comparison findings for m = 6, k = 5, and K = 512 bits. SkNNb's parallel version is around 6 times more efficient than its serial version, according to our findings. This is due to the parallel version's ability to perform operations on six data records at once (i.e., on 6 threads in parallel). For example, when n = 10000, the parallel and serial versions of SkNNb take 40 and 215.59 seconds to run, respectively.We believe that parallelizing the procedures in SkNNm will yield similar efficiency advantages. We suggest that the proposed protocols' scalability issue can be removed or lessened based on the previous arguments, especially in a cloud computing environment where high speed parallel processing is easily obtained. Furthermore, we may enhance speed even more by running parallel processes on several nodes using existing map-reduce algorithms. This analysis will be left to future research.

IV. CONCLUSION

In many data mining applications, the k-nearest neighbours query is one of the most widely used queries. Secure query processing over encrypted data becomes difficult in an outsourced database environment where encrypted data is stored in the cloud. Over encrypted data, the present SkNN algorithms are insecure. We introduced two new SkNN protocols for encrypted data in the cloud in this paper. The first protocol, which serves as a starting point, sends some data to the cloud. Our second protocol, on the other hand, is completely secure, meaning it protects the data's confidentiality, the user's input inquiry, and the data access patterns.

However, as compared to the basic protocol, the second protocol is more expensive. We also tested the performance of our protocols with a variety of parameter settings. We plan to examine and extend our research to other complicated conjunctive searches over encrypted data in the future.

REFERENCES

[1] H. Hu, J. Xu, C. Ren, and B. Choi, “Processing private queries over untrusted data cloud through privacy homomorphism,” in ICDE. IEEE, 2011, pp. 601–612.

[2]P. Mell and T. Grance, “The nist definition of cloud computing (draft),” NIST special publication, vol. 800, p. 145, 2011.

[3]M. Li, S. Yu, W. Lou, and Y. T. Hou, “Toward privacy-assured cloud data services with flexible search functionalities,” in ICDCSW. IEEE, 2012, pp. 466–470.

[4]P. Williams, R. Sion, and B. Carbunar, “Building castles out of mud: practical access pattern privacy and correctness on untrusted storage,” in CCS. ACM, 2008, pp. 139–148.

[5]M. S. Islam, M. Kuzu, and M. Kantarcioglu, “Access pattern disclosure on searchable encryption: Ramification, attack and mitigation,” in NDSS, 2012.

[6]B. Hore, S. Mehrotra, and G. Tsudik, “A privacy-preserving index for range queries,” in VLDB, 2004, pp. 720–731.

[7]E. Shi, J. Bethencourt, T.-H. Chan, D. Song, and A. Perrig, “Multi- dimensional range query over encrypted data,” in IEEE Symposium on Security and Privacy (SP’07). IEEE, 2007, pp. 350–364.

[8]B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, “Secure mul- tidimensional range queries over outsourced data,” The VLDB Journal, vol. 21, no. 3, pp. 333–358, 2012.

[9]H. Hacıgu¨mu¨s¸, B. Iyer, and S. Mehrotra, “Efficient execution of aggre- gation queries over encrypted relational databases,” in Database Systems for Advanced Applications. Springer, 2004, pp. 125–136.

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

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907010 |

[10]E. Mykletun and G. Tsudik, "Aggregation queries in the database-as-a- service model," in Data and Applications Security XX. Springer, 2006, pp. 89–103.
[11] Monika D.Rokade ,Dr.Yogesh kumar Sharma,"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.Yogesh kumar Sharma"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.Yogesh kumar 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.Yogesh kumar 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.
[17]J. Domingo-Ferrer, "A provably secure additive and multiplicative pri- vacy homomorphism," Information Security, pp. 471–483, 2002.
[18]M. Shaneck, Y. Kim, and V. Kumar, "Privacy preserving nearest neigh- bor search," Machine Learning in Cyber Trust, pp. 247–276, 2009.
[19]Y. Qi and M. J. Atallah, "Efficient privacy-preserving k-nearest neighbor search," in ICDCS. IEEE, 2008, pp. 311–319.
[20]J. Vaidya and C. Clifton, "Privacy-preserving top-k queries," in ICDE. IEEE, 2005, pp. 545–546.
[21]G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Pri- vate queries in location based services: anonymizers are not necessary," in SIGMOD. ACM, 2008, pp. 121–132.
[22]S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," SIAM Journal of Computing, vol. 18, pp. 186–208, February 1989.
[23]B. K. Samanthula and W. Jiang, "An efficient and probabilistic secure bit-decomposition," in ACM ASIACCS, 2013, pp. 541–546.
[24]S. Bugiel, S. Nu¨rnberger, A.-R. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing (extended abstract)," in Workshop on Cryptography and Security in Clouds, March 2011.
[25]Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k-nearest neighbor query over encrypted data in outsourced environments," eprint arXiv:1307.4824, 2013.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details