e-ISSN:2582 - 7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 4, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 5.928

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 5.928|

| Volume 4, Issue 7, July 2021 |

|DOI:10.15680/IJMRSET.2021.0407021|

Efficient Secure k-Nearest Neighbours over Encrypted Data

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

PG Student, Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India[1]

Assistant Professor (ME Co-coordinator), Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India[2]

ABSTRACT:Due to inherent security and privacy concerns, enterprise cloud clients are apprehensive about outsourcing sensitive user and corporate data. In this context, directly storing and computing on encrypted data is an appealing alternative, particularly in the face of insider threats. Unfortunately, the fundamental enabling technology, homomorphic encryption, is extremely expensive. We focus on identifying k-Nearest Neighbours (k-NN) directly on encrypted data in this work, which is a basic data-mining and machine learning approach.
Without the cloud knowing anything about the data, question, results, or access and search habits, the goal is to compute the nearest neighbours to a given query and return accurate results to the clients. In a two-party cloud setting, we describe a novel protocol that uses an underlying homomorphic encryption technique. We provide asymptotically faster performance in this context than the current state-of-the-art protocol, without sacrificing any security guarantees. On simulated data, we constructed our protocol to show that it is efficient and feasible on big and important real-world datasets, as well as to investigate how it scales well across different parameters.

I. INTRODUCTION

Finding k-Nearest Neighbours (k-NN) is considered one of the most basic data mining strategies for pattern discovery. It's non-parametric, which means it doesn't make any assumptions about the underlying data distributions. It's also a lazy learning strategy, as there's no attempt to generalise the data until a query is supplied.The purpose of a k-NN method for a given query is to locate the query's k "nearest" neighbours using an acceptable measure of proximity or distance. It has uses in picture segmentation candidate patterns, location-based search, and the classification of symptoms and diagnoses in medical data, to mention a few. These patterns could be utilised as inputs for more advanced learning algorithms in the future.

Client data records held on clouds, such as customer transactions, order histories, credit card details, and other personally identifiable information, are becoming increasingly secret in the context of cloud computing services (PII). Algorithms, and especially algorithmic parameters, are also proprietary and sensitive, as shown in recommendation systems. Inadvertent or unauthorised data or computation disclosure might have major legal and financial ramifications. Clients can store encrypted data in data centres to preserve the secrecy of important information. Importing all of the data back to the client, decrypting it, and doing computations is not cost effective, since it defeats the benefits of migrating to the cloud platform in the first place.

II. BACKGROUND

We detail our adversary model and trust assumptions in this section, as well as our two-party cloud architecture. With the purpose of making this study self-contained, we also explain the Secure k-NN problem in terms of what functions must be computed on encrypted data and briefly present homomorphic encryption.

Cloud Architecture

Outsourcing data or computation to cloud servers has security issues that must be thoroughly investigated. Private clouds, public clouds, and hybrid clouds are among the deployment options. Outsourcing data and computation may not be a justifiable risk for businesses that place a high value on data security and computational integrity. Private cloud solutions fill this gap, with services available exclusively within corporate intranets and restricted cloud computing benefits like on-demand scalability and load balancing, as well as higher initial infrastructure and set-up expenses for new services. The public clouds, on the other hand, are massive third-party owned server farms and data centres that

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 5.928|

| Volume 4, Issue 7, July 2021 |

|DOI:10.15680/IJMRSET.2021.0407021|

house client data and computing services for hire. Insider attacks, in which personnel within a public cloud organisation might watch and infer trade secrets, and side channels, such as unintended or indirect information leaks when data from competing companies is shared on the same bare metal, are two major problems. The scope of this paper does not include side channel analysis. A third option, the hybrid cloud model, combines the best of both worlds by allowing businesses to offload only a portion of their services to public clouds, depending on the sensitivity of the data and processing.

Trust Assumptions

As with public clouds, the trust model for federated clouds is that of an honest-but-curious or semi-honest opponent who does not meddle directly with the computation or data. On the cloud networks and servers, however, the adversary is free to see inputs and outputs, as well as side-effects of computation and other behavioural traits. This opponent differs from the passive observer or malevolent adversarial model previously considered, in that the adversary is trusted to do the computation successfully, but also has access to the service's internal state, which includes client data. An adversary of this type can monitor memory and network traffic, as well as examine how operating systems respond to client queries. Because data owners must give control to cloud service providers, the choosing of this specific sort of adversary is warranted. While legal contracts and liabilities can limit and regulate these risks, the threat of a curious insider cannot be ruled out. The gap is filled by computing directly over encrypted data. The idea is that a hostile insider cannot get relevant information from the data or computation by observation, even when we operate in the honest-but-curious adversary model. We also want to keep the adversary from learning database access patterns, such as the set of (encrypted) result tuples provided corresponding to a specific input query, and query search patterns, which could reveal information like how many times the same query was performed.

III. PROPOSED METHODOLOGY

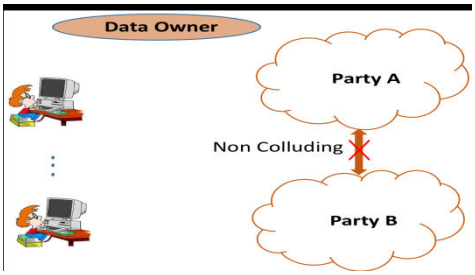

Figure 1: Entities and Relationships

The main entities in our protocol are shown in Figure: 1.

• Data owner: The data owner is a trustworthy entity, the legal owner of the plaintext database, who outsources the encrypted data storage to Party A for k-NN computation, i.e. the plaintext database is not revealed to Party A. The data owner can be offline once the data is outsourced.

• Party A: Party A uses an encrypted database for storage and calculation. It takes encrypted query inputs and returns k encrypted database points representing the query point's k-NN. It is not in possession of any secret keys and only operates with encrypted data. Party A is believed to be a trustworthy but inquisitive opponent who will not tamper with the protocol's normal execution but has access to its internal state and could try to infer extra information about the plaintext data, query, results, or access and search habits.

• Party B: Party B has access to the secret keys that are used to encrypt a database, but not to the encrypted database or query itself. Instead, it has only (partial) access to the results of computations performed on the encrypted data. Party B, like Party A, is truthful but curious, and does not meddle with the calculations. It can, however, infer properties about the plaintext database and query using any information provided.A Secure k-NN solution will show that this informationcannot be learned by Party B even if it has the secret key.

• Clients : These are users who have been given permission by the data owner to communicate with Party A and submit k-NN queries to the outsourced database. Clients have access to keys that allow them to send and decrypt encrypted queries and responses.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 5.928|

| Volume 4, Issue 7, July 2021 |

|DOI:10.15680/IJMRSET.2021.0407021|

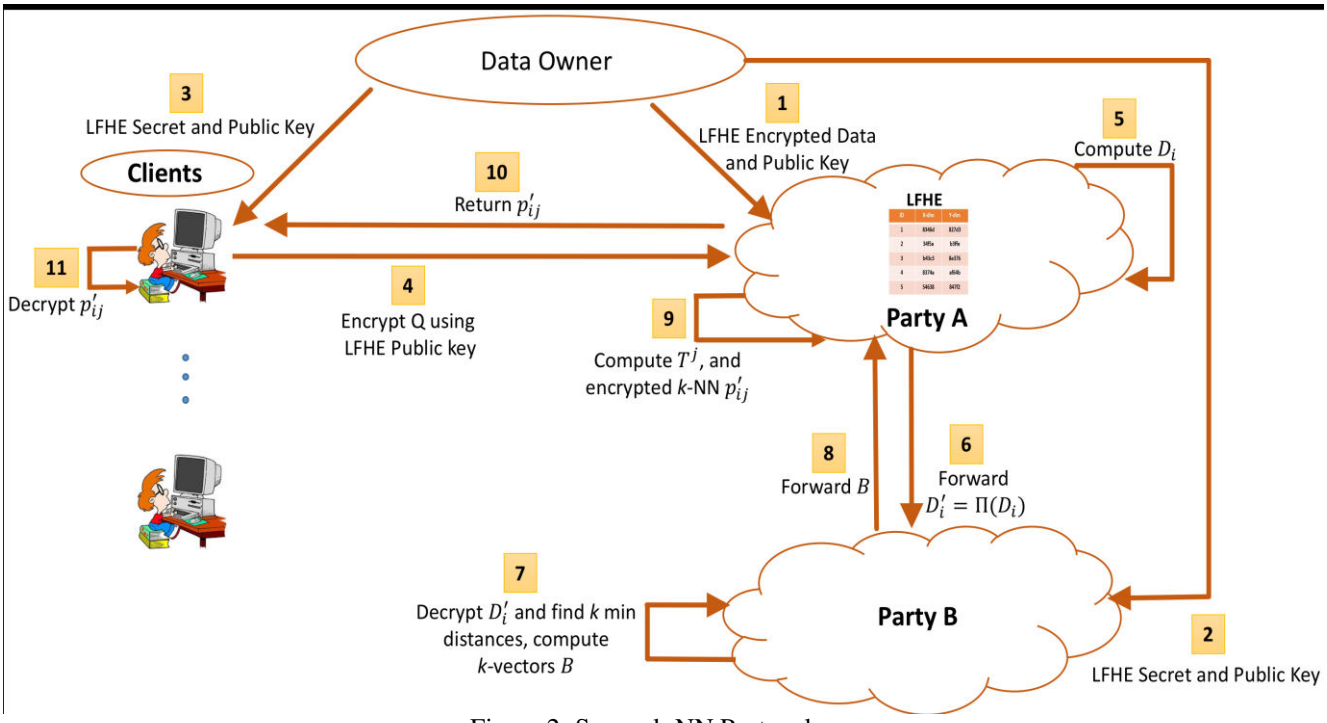

Figure 2: Secure k-NN Protocol

- Figure 2 shows the entities, the communication and the computation phases in our main protocol. The Setup phase is executed once at the time of transferring the encrypted data to Party A. Let (sk,pk) ← KeyGen($) be the secret-key and public-key respectively for the the chosen (S)HE
- Party A receives the public key pkfrom the data owner, as shown in label 1 in Figure 2.
- Party A also receives the encrypted database P ′ from the data owner, where the magnitude of each dimension d of each point pi is encrypted under (S)HE using Encpk . These d encrypted values together constitute the d dimensional encrypted data points p′i .
- Party B receives both skand pk, as shown in label 2 in Figure 2.
- Clients receive both skand pk, as shown in label 3 in Figure 2.
- The inputs to our Securek-NN protocol are the n encrypted points in P ′ and an encrypted d-dimensional query point Q′ ←Encpk(Q), of the plaintext query point Q computed by the client as shown in label 4 in Figure 2. The output of the protocol, returned by Party A to the client,is the set of k encrypted points ⟨p′i1,p′i2, ・・・ ,p′ik⟩, which are the encrypted k nearest neighbours to query Q in database P.

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.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 5.928|

| Volume 4, Issue 7, July 2021 |

|DOI:10.15680/IJMRSET.2021.0407021|

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.

IV. CONCLUSION

This work provides a new methodology for k-Nearest Neighbors on encrypted data that is both efficient and secure. In the two-party federated cloud concept, we adopt LFHE, a homomorphic encryption technique, as our basic building block. Under the honest-but-suspicious premise, our adversary model captures insider attacks.

In comparison to the current state of the art, our protocol is fast without sacrificing security assurances. Our implementations are quick and scalable, and our real-world data trials indicate that basic data mining on encrypted data is feasible. We intend to expand our research into other data mining methods in the future, such as k-Means and Apriori.

REFERENCES

[1] R. Agrawal, D. Asonov, M. Kantarcioglu, and Yaping Li. 2006.SovereignJoins. In 22nd International Conference on Data Engineering (ICDE'06). 26–26.https://doi.org/10.1109/ICDE.2006.144
[2] RakeshAgrawal, Jerry Kiernan, RamakrishnanSrikant, and YirongXu. 2004.Order preserving encryption for numeric data. In Proceedings of the 2004 ACMSIGMOD international conference on Management of data. ACM, 563–574.
[3] RakeshAgrawal and RamakrishnanSrikant. 2000. Privacy-preserving DataMining. SIGMOD Rec. 29, 2 (May 2000), 439–450. https://doi.org/10.1145/335191.335438
[4] ArvindArasu, Spyros Blanas, Ken Eguro, RaghavKaushik, Donald Kossmann,Ravi Ramamurthy, and RamaratnamVenkatesan. 2013. Orthogonal securitywith cipherbase. In Proc. of the 6th CIDR, Asilomar, CA.
[5] S. Bajaj and R. Sion. 2014. TrustedDB: A Trusted Hardware-Based Databasewith Privacy and Data Confidentiality. IEEE Transactions on Knowledge andData Engineering 26, 3 (March 2014), 752–765. https://doi.org/10.1109/TKDE.2013.38
[6] Alexandra Boldyreva, Nathan Chenette, Younho Lee, and Adam O'neill. 2009.Order-preserving symmetric encryption. In Annual International Conferenceon the Theory and Applications of Cryptographic Techniques. Springer, 224–241.
[7] Alexandra Boldyreva, Nathan Chenette, and Adam O'Neill. 2011. Orderpreservingencryption revisited: Improved security analysis and alternativesolutions. In Annual Cryptology Conference. Springer, 578–595.
[8] Dan Boneh, Eu-Jin Goh, and KobbiNissim. 2005. Evaluating 2-DNF Formulason Ciphertexts. In Proceedings of the Second International Conference on Theoryof Cryptography (TCC'05). Springer-Verlag, Berlin, Heidelberg, 325–341. https://doi.org/10.1007/978-3-540-30576-7_18
[9] ZvikaBrakerski, Craig Gentry, and VinodVaikuntanathan. 2012. (Leveled)Fully Homomorphic Encryption Without Bootstrapping. In Proceedings of the3rd Innovations in Theoretical Computer Science Conference (ITCS '12). ACM,New York, NY, USA, 309–325. https://doi.org/10.1145/2090236.2090262
[10] ZvikaBrakerski, Craig Gentry, and VinodVaikuntanathan. 2014. (Leveled)Fully Homomorphic Encryption Without Bootstrapping. ACM Trans. Comput.Theory 6, 3, Article 13 (July 2014), 36 pages. https://doi.org/10.1145/2633600
[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

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 5.928|

| Volume 4, Issue 7, July 2021 |

|DOI:10.15680/IJMRSET.2021.0407021|

[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.KhatalDr.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] Craig Gentry and ShaiHalevi. 2011. Implementing Gentry's FullyhomomorphicEncryption Scheme. In Proceedings of the 30th Annual InternationalConference on Theory and Applications of Cryptographic Techniques:Advances in Cryptology (EUROCRYPT'11). Springer-Verlag, Berlin, Heidelberg,129–148. http://dl.acm.org/citation.cfm?id=2008684.2008697

[18] Craig Gentry, ShaiHalevi, and Nigel P. Smart. 2012. Better Bootstrapping inFully Homomorphic Encryption. In Proceedings of the 15th International Conferenceon Practice and Theory in Public Key Cryptography (PKC'12). Springer-Verlag, Berlin, Heidelberg, 1–16. https://doi.org/10.1007/978-3-642-0057-8_1

[19] Gabriel Ghinita, PanosKalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-LeeTan. 2008. Private Queries in Location Based Services: Anonymizers Are NotNecessary. In Proceedings of the 2008 ACM SIGMOD International Conferenceon Management of Data (SIGMOD '08). ACM, New York, NY, USA, 121–132.https://doi.org/10.1145/1376616.1376631

[20] NikolayGrozev and RajkumarBuyya. 2014. Inter-Cloud architectures andapplication brokering: taxonomy and survey. Software: Practice and Experience44, 3 (2014).

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 5.928

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

9710 583 466 9710 583 466 ijmrset@gmail.com

www.ijmrset.com