Adaptive Query Processing Over Encrypted Data Using Blow-fish

G.Ram Sankar[1], Jeeva S[2], Pavan Kumar N[3], Thirumugil T[4], Ugendhar U[5]
[1,2,3,4,5]*Information Technology, Adhiyamaan College of Engineering*

Abstract- Processing nearest neighbor queries is a fundamental problem that occurs in many sectors, including machine learning and geographic databases. This research focuses on the Secure Nearest Neighbor (SNN) issue in cloud computing. Prior SNN systems have been ineffective and unsafe. In the present article, we officially establish and empirically demonstrate that the SNN scheme ASPE is truly vulnerable to ciphertext-only attacks. Although previous study showed that building an SNN method is difficult even in significantly permissive standard security models, we highlight the shortcomings of the hardness proof. We present an SNN architecture and show how it can withstand adaptive chosen keyword assaults. Because the complexity of processing queries is exponential, our method is efficient. We created our SNN scheme in C++ and compared its performance with a plain text scheme, a binary scheme, and a PIR scheme on a massive collection of over 10 million real-world data points to determine its efficiency. Experiment findings demonstrate that our scheme is both fast (0.124 millisecond per query when the data set size is 10 million) and scalable in terms of data points.

Keywords: Secure Nearest Neighbor Queries, Adaptive IND-CKA security, and Cloud Computing

I. INTRODUCTION

Because of their lower costs, higher reliability, better performance, and faster deployment, enterprises and users are attracted to public clouds such as Amazon EC2 and S3, Microsoft Azure, and Google App Engine. However, the primary barrier is privacy, because data owners may not completely trust public clouds. First, clouds may have dishonest personnel. In 2010, for example, a Google employee gained access to many children's Gmail and Google Voice accounts. Second, clouds may be jeopardized. As an example, in 2014, Apple iCloud was breached, resulting in the online publication of hundreds of celebrity private photographs. Third, some cloud computing facilities may be located in countries where privacy regulations may be difficult to enforce. Amazon, Microsoft, and Google, for example, have data centers located all over the globe. The cloud computing model used in this study includes a data owner storing data on the cloud and several data users querying the data. The computational problem in this research is closest neighbor searching. The aim of nearest neighbour searching is to find the point in D that is closest to a query q given a collection of multidimensional Euclidean points D = d1, d2, dn and a query q described as a multidimensional point. This fundamental problem affects numerous disciplines, including spatial databases, machine learning, and computer vision. The Secure Nearest Neighbor (SNN) problem is the focus of this study. We concentrate on Searchable Symmetric Encryption (SSE) methods in particular because they are more effective than asymmetric searchable encryption schemes. The data owner first encrypts each data point di D into (di)K using a key K shared by the data owner and data users, then creates a secure index for the data points in D and sends both the encrypted data points (d1) K, (d2) K,, (dn)K and the secure index to the cloud to securely store and search a data set D = d1, d2,, dn. A backdoor tq is created by the data consumer from a query q and sent to the cloud. The cloud scans the secure index for data points that fit the query using the trapdoor. Given the encrypted data, trapdoors, and query results, this process should prevent the cloud from deducing useful information about the data and queries, such as the values of the data points, the substance of the searches, and the statistical features of the data points.

This project employs the adaptive IND-CKA (indistinguishable (IND) under chosen keyword attack (CKA)) security paradigm proposed in. The term "IND-CKA" refers to the index's ability to be undetectable to chosen keyword attacks, whereas "adaptive" refers to the adversary's query selection technique. In the adaptive INDCKA paradigm, the adversary makes query selections based on previously chosen questions as well as the trapdoors and

outcomes of those queries. In the non-adaptive IND-CKA model, the attacker selects all questions at once and gets the relevant trapdoors and query results. The levels of security for these two kinds differ. A method is only secure under the non-adaptive IND-CKA model if the queries are not reliant on the secure index or previous search results. A secure system, on the other hand, guarantees security even when queries are reliant on the secure index and the results of previous searches.

II. LITERATURE REVIEW

Fernando Krell et.al While rarely explicitly taken into consideration outside of the theoretical community, query privacy in secure DBMS is a crucial characteristic. Nearly all prior works addressing practical applications either handle limited queries (e.g., merely keyword search) or provide a weak guarantee of privacy due to the enormous overheads of ensuring privacy in complicated searches. In this work, we tackle an important outstanding issue in private databases: effective sub-linear search for any type of Boolean query. We take into account scalable DBMS with demonstrable security for all parties, which includes safeguarding the data from both the server (which stores encrypted data) and client (which searches it), as well as safeguarding the query and access control for the inquiry. We create a robust DBMS system that is appropriate for deployment in the real world, and then we test its performance(1). Rui Li et.al The first conjunctive query processing system that respects privacy and adheres to the aforementioned criteria is what we provide in this paper. We provide an Indistinguishable Bloom Filter (IBF) data structure for indexing to achieve adaptive security. We propose an extremely balanced binary tree data structure called the Indistinguishable Binary Tree to provide effective query processing and structure indistinguishably (IBtree). We suggest a traversal depth minimising algorithm and a traversal width minimization algorithm to increase search efficiency. We provide an IBtree space compression approach to eliminate extraneous information in IBFs in order to obtain salable and compact index size. Using a random oracle model, we explicitly demonstrate the adaptive security of our method(2).Minxin Du et.al In this research, we examine novel privacy-preserving indexing and query processing protocols that satisfy various desirable aspects, such as multi-keyword query processing with conjunction and disjunction logic queries, substantially strong privacy guarantees with adaptive chosen keyword attack (CKA2) security and forward privacy, the support of dynamic data operations, etc. Our solutions are much more compact, useful, and versatile than earlier plans. Rigorous analysis is used to precisely characterize their security and performance. Our solutions can provide small search time efficiency, as shown by experimental evaluations carried out across a sizable representative data set, and they are applicable for usage in sizable encrypted database systems(3).Lili Zhang et.al The conjunctive keyword search method we offer in this work protects user privacy while supporting dynamic update operations on encrypted cloud data. We specifically build an index structure based on the multi-attribute tree (MAT) and provide the searchMAT algorithm, an effective search technique over the index tree. The MCKS-MAT strategy, which we suggest, is a multi-attribute conjunctive keyword search method based on MAT that can achieve equality conjunction, subset conjunction, and range conjunction as well as meet privacy requirements in the context of the well-known background attack model.A sufficient number of trials are included with this paper to allow readers to assess the effectiveness of the suggested plan. The suggested technique delivers decreased computing overhead in initialization, trapdoor generation, and queries, but requires a somewhat greater prepossessing cost relative to the linear search due to the construction of the tree-based index, according to experiments(4).Hyeong-Il Kim et.al Database outsourcing has become a new platform as cloud computing has gained traction. Databases must be encrypted before being transferred to the cloud due to the major privacy concerns there. As a result, numerous kNN query processing methods for the encrypted database have been presented. Nevertheless, the current plans are either ineffective or insecure. Thus, we suggest a novel secure kNN query processing technique in this study.(5)Cheng Hong et.al A particular security flaw with symmetric searchable encryption is that, when performing CKS (Conjunctive Keywords Search), the trapdoors and search results may betray the connections between the keywords being searched. For instance, it suggests that A is likely a subset of B if the search result for keyword set A is the superset of keyword set B's.

These inclusion-relation (IR) attacks affect the majority of the search techniques now in use that support CKS. We outline metrics for IR security and provide CKS-SE, a safe CKS implementation based on a bloom filter that makes trapdoor expressions IR-secure through randomization and integration. The performance of CKS-SE is among the finest, according to experiments, and the average false positive rate is within an acceptable range(6).Stanislaw Jarecki et.al This work presents the design and analysis of the first searchable symmetric encryption (SSE) protocol that supports conjunctive search and general Boolean queries on outsourced symmetrically- encrypted data and that scales to very large databases and arbitrarily-structured data including free text search. Work in this field up to this point has mostly centred on single-keyword searches. For the case of conjunctive search, past SSE constructs needed labour linear in the total number of documents in the database and guaranteed acceptable privacy only for organised attribute-value data, rendering these methods too slow and inflexible for large practical databases(7).Hugo Krawczyk et.alThis article discusses the development and analysis of the first searchable symmetric encryption (SSE) protocol, which scales to very large databases and arbitrarily-structured data, including free text search, and supports conjunctive search and general Boolean queries on outsourced symmetrically-encrypted data. Work in this field up to this point has mostly centred on single-keyword searches. Prior SSE constructs for conjunctive search only gave adequate privacy for structured attribute-value data and needed labour linear in the number of documents in the database, making these solutions prohibitively slow and rigid for large practical databases(8).Seny Kamara et.al

A party can transfer the private storage of its data to a third party (a server) while still having the option to conduct limited searches on it thanks to searchable symmetric encryption (SSE). In recent years, this issue has been the subject of considerable research. In this study, we provide two SSE solutions that share the qualities listed below.

Each of the two options is more effective than the constant-round alternatives. In particular, the server's labour per document returned is constant rather than linearly increasing with data size(9).Antonios Deligiannakis et.al We take into account a data owner who offloads a dataset to a server that is not reliable. Without compromising the privacy of the data and the queries, the owner wants to make it possible for the server to respond to range queries on a single property. There are various "practical" private range search algorithms (mostly in databases venues) that try to balance security and efficiency. Yet, these techniques either don't provide verifiable security guarantees or allow for serious privacy leaks. The rigour of Security formulations and proofs is combined in this work with effective Data Management strategies using an interdisciplinary approach. Using the idea of Searchable Symmetric Encryption (SSE), which was mainly developed for keyword search, we construct a broad range of innovative schemes with realistic security/performance trade-offs. (10).

III.METHODOLOGY

3.1 PROBLEM STATEMENT

The processing overhead of the current method is, however, prohibitively expensive, rendering it impractical in a cloud setting.

We focus on Searchable Symmetric Encryption (SSE) methods in particular because searchable encryption based on symmetric encryption is more effective than searchable encryption based on asymmetric encryption.The issue of safely calculating the distance between an encrypted query and encrypted data points in order to identify the nearest neighbor(s) of the query among the encrypted data points is known as the problem of Adaptive Secure Nearest Neighbor Query Processing over Encrypted Data (ASNNQP).

The query in this issue is encrypted using the same public-key encryption technique as the data points in the problem. The objective is to create a secure computation protocol that enables a client to transmit an encrypted query to a server, which then determines the distance between each encrypted data point and the query without disclosing any information about the data or the query to the client or any other party.

DISADVANTAGE:

1. Encryption and decryption operations significantly increase computing overhead while processing nearest neighbour searches, which can have a negative impact on query performance and response time.

2. Limitations on query capabilities: Using encrypted data may restrict the kinds of queries that may be made because doing some operations on it may be challenging or impossible.
3. Vulnerabilities in the algorithms and protocols used for query processing could be exploited by attackers, despite encryption's ability to secure data.
4. Restricted scalability: As the dataset size grows, the computational difficulty of handling k nearest queries likewise grows, which may restrict the system's ability to scale.

3.2 PROPOSED APPROACH

In this study, we present Sean, a highly efficient and secure SNN algorithm for one- and two-dimensional data points.SecNN is protected from security threats according to the adaptive IND-CKA security paradigm suggested in. SecNN's query processing complexity is O(log n) for efficiency, where n is the number of data points.Processing SNN queries in high-dimensional space is still unknown since it is difficult to handle textual closest neighbour requests in this setting. We want the scholarly community to pay more attention to this crucial and difficult issue.To encrypt the data and enable secure calculations on the encrypted data, the system employs homomorphic encryption. To enhance query performance, the system also makes use of an adaptive indexing strategy. By dynamically altering the index structure in response to the queries being run, the adaptive indexing technique can help to shrink the search space and enhance query performance.

ADVANTAGES:

1. Top k multi-keyword retrieval over encrypted cloud data with excellent security and practical effectiveness.

2. Confidentiality: Even when a query is made, the encrypted data is still safe and confidential. Using encrypted data, the query is performed without disclosing any private data to the server or other unauthorized parties.

3. Flexibility: Adaptive secure nearest neighbour query processing is flexible and may be utilized with various encryption systems, making it suitable for a range of use cases.

4. Efficiency: By reducing the requirement for data decryption and re-encryption, which can be time-consuming and computationally expensive, this approach can increase the efficiency of searching huge databases.

3.3 ALGORITHMS

Blow fish Algorithm:

A variable-length key, used by blowfish, can be between 32 and 448 bits long. The plaintext is converted into ciphertext through a series of rounds by first expanding the key into an array of subkeys. Each 64-bit block of the input data is subjected to a total of 16 rounds of encryption as part of the algorithm's operation. A Feistel network along with straightforward substitution and permutation operations make up each round.

The Advanced Encryption Standard (AES) has entirely replaced Blow fish in many applications, although despite its antiquity, it is still commonly used today.

3.4 SYSTEM ARCHITECURE

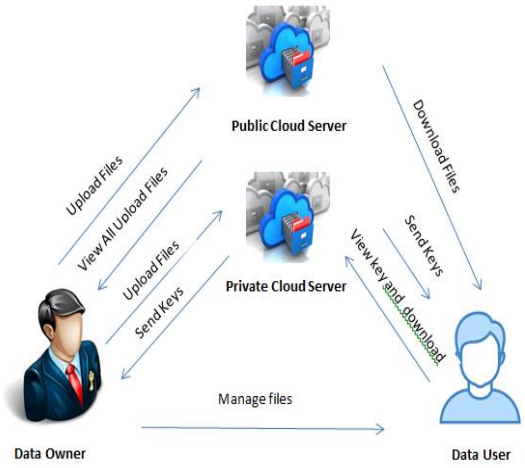


Fig. 1: Design for Cloud Technology

3.5 DESCRIPTION OF A MODULE

The project consists of three modules:

1. Data Owner
2. Data User
3. Cloud Server

DATA OWNER

Register the account with the basic information .After authorized by cloud owner may login the account.The data owner is uploading the encrypted file means if give request for cloud .The cloud is accept for file upload request next owner Upload the file with the encrypted file file .

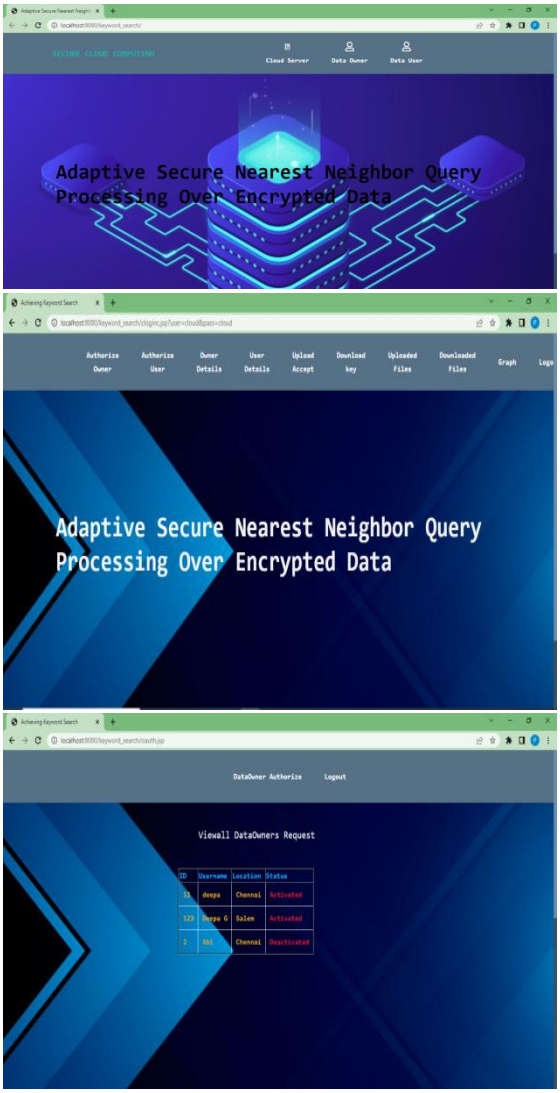
DATA USER

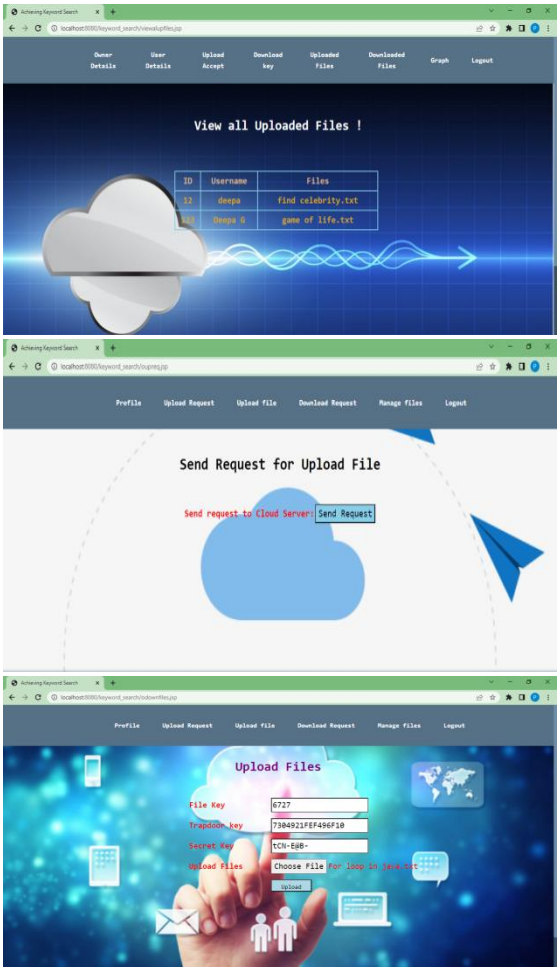
Register the account with the basic details. After authorized by cloud user can login the account.Search the keyword and make a request to particular file.Cloud transmits the File key. If we enter the proper key implies, download the file with the Decrypted Format.

CLOUD SERVER

Login the account using the right username and password. View owner and user authorize them View every uploaded document. Approve the request for the data owner's cloud file upload. See the request for download and send the key. View graph

IV.RESULT





V.CONCLUSION

In this paper, we make the following important advances. First, we formalize and experimentally show that the widely used SNN scheme ASPE is vulnerable to even ciphertext-only attacks. Second, we bring out the flaws in SecNN's hardness analysis in. Third, we suggest SecNN, the first SNN scheme for one and two-dimensional data points that meets the requirements of security against chosen plain-text attacks, query processing by clouds, and query processing time efficiency. Our scheme incorporates several novel concepts, such as multi-homing Bloom filters and single-homing Bloom filters, that can be applied in other contexts. Fourth, we used C++ to build SecNN and tested its performance on a large real-world data set. The experimental results indicsate that SecNN is fast in terms of query processing. The experimental findings show that SecNN is both fast

and scalable in terms of query processing time and index size.

VI.FUTURUE SCOPE

It's a difficult research problem, but one that hopes to make it possible to process encrypted data securely and effectively. A popular symmetric encryption algorithm called blowfish can be used to give encrypted data secrecy and integrity. There are several potential future lines of inquiry in this field: Machine learning with privacy protection: Investigating how to use Blow fish to execute privacy-preserving machine learning over encrypted data is an exciting area for future research. This can allow for secure data analytics without disclosing private data. Access control is yet another crucial aspect of secure query processing. Future research can look into how to create effective access control systems that can be employed with Blow fish-based encryption techniques. Finally, research might concentrate on accelerating Blow fish-based encryption algorithms using hardware solutions like FPGA or ASIC-based implementations. By doing so, the efficiency of safe query processing on encrypted data can be greatly increased.

REFERENCE

[1] K. Bache and M. Lichman. UCI machine learning repository, 2013. http://archive.ics.uci.edu/ml.

[2] Blind seer: A scalable private DBMS
V Pappas, F Krell, B Vo, V Kolesnikov…… IEEE Symposium on …, 2014 ieeexplore.ieee.org

[3] Adaptively secure conjunctive query processing over encrypted data for cloud computing
R Li, AX Liu - 2017 IEEE 33rd International Conference on Data …, 2017 - ieeexplore.ieee.org

[4] Privacy-preserving indexing and query processing for secure dynamic cloud storage
M Du, Q Wang, M He, J Weng - IEEE Transactions on …, 2018 - ieeexplore.ieee.org.

[5] Privacy-preserving and dynamic multi-attribute conjunctive keyword search over encrypted cloud data L Zhang, Y Zhang, H Ma - IEEE Access, 2018 - ieeexplore.ieee.org

[6] A kNN query processing algorithm using a tree index structure on the encrypted database HI Kim, HJ Kim, JW Chang -Conference on Big Data and Smart, 2016 - ieeexplore.ieee.org

[7] A secure conjunctive keywords search over encrypted cloud data against inclusion-relation attack K Cai, C Hong, M Zhang, D Feng… - 2013 IEEE 5th …, 2013 - ieeexplore.ieee.org

[8] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Steiner. Highly-scalable searchable symmetric encryption with support for boolean queries. In Proc. Inte. Cryptology Conf. (CRYPTO), pages 353– 373, California, 2013. Springer.

[9] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In Proc. of the ACM Conf. on Computer and Communications Security (CCS), pages 79–88, Virginia, 2006. ACM.

[10] I. Demertzis, S. Papadopoulos, and O. Papapetrou. Practical private range search revisited. In Proc. of the International Conference on Management of Data (SIGMOD), pages 185–198, San Francisco, 2016. ACM.

[11] Y. Elmehdwi, B. K. Samanthula, and W. Jiang. Secure k-nearest neighbor query over encrypted data in outsourced environments. In Proc. of the IEEE International Conference on Data Engineering (ICDE), pages 664– 675, Chicago, 2014. IEEE Computer Society.

[12] S. Faber, S. Jarecki, H. Krawczyk, Q. Nguyen, M. Rosu, and M. Steiner. Rich queries on encrypted data: Beyond exact matches. In Proc. of the 20th European Symposium on Research in Computer Security (ESORICS), pages 123–145, Vienna, 2015. Springer.

[13] O. Fcontributors. Openstreetmap data for china, 2016. http://download. geofabrik.de/asia/china.html

[14] J. Furukawa. Short comparable encryption. In Proc. of the 13th Cryptology and Network Security International Conference (CANS), pages 337– 352, Heraklion, 2014. Springer.

[15] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. Tan. Private queries in location-based services: Anonymizers are not necessary. In Proc. of the International Conference on Management of Data (SIGMOD), pages 121–132, Vancouver, 2008. ACM.

[16] P. Grubbs, K. Sekniqi, V. Bindschaedler, M. Naveed, and T. Ristepart. Leakage-abuse attacks against order-revealing encryption. In Proc. of the IEEE Symposium on Security and Privacy (S&P), pages 655–672, San Joes, CA, USA, 2017. IEEE.

[17] P. Gupta and N. McKeown. Algorithms for packet classification. IEEE Network, 15(2):24–32, 2000.
[18] H. Hu, J. XU, C. Ren, and B. Choi. Processing private queries over untrusted data cloud through privacy homomorphism. In Proc. of the IEEE International Conference on Data Engineering (ICDE), pages 601– 612, Hannover, 2011. IEEE Computer Society.
[19] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. IEEE Transaction on Neural Networks, 10(3):626– 634, 1999.
[20] S. Kamara and T. Moataz. Boolean searchable symmetric encryption with worst-case sub-linear complesity. In Annual International Conference on the Theory and Application of Cryptographic Techniques (EUROCRYPT), pages 94–124. Springer, 2017.
[21] S. Kamara and T. Moataz. Boolean searchable symmetric encryption with worst-case sub-linear complexity. In EUROCRYPT, 2017.
[22] S. Kamara and C. Papamanthou. Parallel and dynamic searchable symmetric encryption. In Proc. of the Financial Cryptography and Data Security (FC), pages 258–274, Okinawa, 2013. Springer.
[23] S. Kamara, C. Papamanthou, and T. Roeder. Dynamic searchable symmetric encryption. In Proc. of the ACM Conf. on Computer and Communications Security (CCS), pages 965–976, Raleigh, 2012. ACM.
[24] P. Karras, A. Nikitin, M. Saad, R. Bhatt, D. Antyukhov, and S. Idreos. Adaptive indexing over encrypted numeric data. In Proc. of the International Conference on Management of Data (SIGMOD), pages 171–183, San Francisco, 2016. ACM.
[25] J. Katz and Y. Lindell. Introduction to Modern Cryptography. CRC press, Washington,D.C., 2007.
[26] F. Kerschbaum and A. Schropfer. Optimal average-complexity idealsecurity order-preserving encryption. In Proc. of the ACM Conf. on Computer and Communications Security (CCS), pages 275–286, Scottsdale, 2014. ACM.
[27] H.-I. Kim, H.-J. Kim, and J.-W. Chang. A knn query processing algorithm using a tree index structure on the encrypted database. In Proc. of the International Big Data and Smart Computing (BigComp), pages 93–100, Hong Kong, China, 2016. IEEE.
[28] K. Kurosawa and Y. Ohtaki. Uc-secure searchable symmetric encryption. In Proc. of the Financial Cryptography (FC), pages 285–298, Kralendijk, 2012. Springer.
[29] X. Lei, A. X. Liu, R. Li, and G.-H. Tu. Seceqp: A secure and efficient scheme for sknn query problem over encrypted geodata on cloud. In Proc. of the IEEE International Conference on Data Engineering (ICDE), pages 662–673. IEEE, 2019.
[30] R. Li and A. X.Liu. Adaptive secure conjunctive query processing over encrypted data for cloud computing. In Proc. of the IEEE International Conference on Data Engineering (ICDE), San Diego, 2017. IEEE Computer Society.