National Conference on Sustainable Engineering And Management (NCSEM24)
ISSN 2349-6002

Searching Encrypted Data Using KNN Algorithm

Remya R S
Assistant Professor, Computer science and Engineering, Sivaji College of Engineering and Technology, Manivilla.

Abstract-In medical cloud computing, a patient can remotely export her medical records to a cloud server. In this case, only authorized physicians are allowed access due to the extremely sensitive nature of the medical data. One common practice is to encrypt the data prior to outsourcing; in this scenario, the patient just needs to give the encryption key to the approved medical professionals. Nevertheless, this significantly limits the value of outsourced medical data because it is challenging to search through the encrypted data. This research proposes two Secure and Efficient Dynamic Searchable Symmetric Encryption (SEDESE) algorithms for medical cloud data. First, we suggest a secure k-Nearest Neighbor (kNN) and attribute-based encryption (ABE) approach that is dynamically searchable and symmetric. Two very difficult-to-obtain security properties in the dynamic encryption domain are forward privacy and backward privacy, which this approach can provide.

Index Terms— cloud**, **SEDESE, Encryption, ABE**, **Dynamic searching algorithm**, K**-Nearest neighbor**, **medical, secure

1. INTRODUCTION

In order to lower medical costs and increase medical quality, health care services have been the subject of much research. A health care system must expand its scope in order to provide effective and secure services given the volume of medical data.

Medical cloud computing licenses out the processing and storage capacities to the general public patients and physicians, treating computing as a utility. This cutting-edge computing paradigm allows for resource transparency, self-demand services, dynamic resource allocation and service measurement, among Other things. In this way, the patient can remotely save her data on the cloud server (data outsourcing) and thereafter grant the doctors access to her cloud data. Information that is private and sensitive may be included in the outsourced medical data (ex. medical Case and diagnostic report).Prior to being uploaded to the cloud, medical data must frequently be encrypted. However, because it is difficult to search via encrypted data, the encrypted data cannot offer high usability. Fuzzy keyboard search, ranked keyword search, multi-keyword search, and other features can be achieved using the encryption strategy using searchable symmetric encryption (SSE) technology, which has been presented as a solution to these problems. Many SSE techniques based on k-Nearest Neighbor (KNN) have been presented recently for searching over encrypted material. However, with these techniques, each user's search is conducted using the same secret key, potentially exposing personal information. However, creating a dynamic version of SSE (DSSE) that supports encrypted keyword searches even in the event that data is unilaterally added to or removed from collections (forward privacy) or deleted (backward privacy) is a difficult problem, particularly in the context of the health care system. An effective DSSE strategy that can achieve forward privacy but not guarantee backward privacy was proposed by Stefano et al.

Same researchers accomplish both forward and backward privacy in DSSE by using the oblivious Random-Access Memory (ORAM) approach. However, the complexity of the storing, search, and updating procedures is greatly increased by these approaches.
To address the above issue, in this paper, we propose a secure Efficient Dynamic Searchable Symmetric Encryption (SEDSSE) scheme over the medical cloud data. This work extends and improves our previous research. Specifically, this paper accesses two new issues: the collusion between the cloud server and search user.in addition, we apply the new design to the health care system. Furthermore, the security and performance are analyzed. The original contributions of the paper are firstly, we combine the k-Nearest Neighbor (KNN)

And Attribute-Based Encryption (ABE) technique to propose a Secure and Efficient Dynamic Searchable Symmetric Encryption scheme, named SEPSSE I. The propose scheme can achieve forward privacy, backward privacy, and collusion resistance between the cloud server and search users. Secondly, based on the scheme, we further propose an enhanced scheme, name SEPSSE II to solve the key sharing problem which widely exist in the KNN based searchable encryption schemes. Compared with the existing DSSE schemes, our propose schemes are have less storage costs, search and updating complexity. Extensive experiments demonstrate the efficiency of our scheme in term of storage overhead, index building, and trapdoor generating query.

II.LITERATURE SURVEY

Searching encrypted data while preserving privacy is a crucial challenge in today's world, and the K-Nearest Neighbors (KNN) algorithm offers a promising approach. This survey explores the various techniques and challenges associated with using KNN for encrypted data search B.

M.Li, S. Yu, K. Ren, and W. Lou (2010) the encrypted data using the KNN algorithm encompasses several significant contributions. Proposed an improved KNN algorithm for fine-grained classification of encrypted network flow, achieving high accuracies in identifying the encryption status, application type, and content type of encrypted network flows (Ma et al., 2020). Focused on privacy-preserving kNN query processing algorithms via secure two-party computation over encrypted databases in cloud computing, addressing the challenges of secure computation over encrypted data (Kim et al., 2022). Additionally, presented a privacy-preserving KNN classification algorithm for smart grid applications, demonstrating enhanced efficiency and accuracy compared to previous algorithms, particularly.

A.M.H Kuo (2011) cloud computing is a new way of delivering computing resources and services. Many managers and experts believe that it can improve health care services. Benefit health care research and changes the face of health information technology. However, as with any innovation cloud computing should be rigorously evaluated before its widespread adoption. This paper discusses the concept and its current place in health care uses 4 aspects to evaluate the opportunities and challenges of this model. Strategic planning that could be used by a health organization to determine, strategy, and resources allocation when it has decided to migrate from traditional to cloud-based health service is also discussed.

L.M Vaquero cloud computing to achieve a complete definition of what a cloud computing to achieve a complete definition of what a cloud, using the main characteristics typically associated with these paradigms in this literature. More than 20 definitions have been studied allowing for the extraction of a consensus definition as well as a minimum definition containing the essential characteristics. This paper pays much attention to the grid paradigm.as it is often confused with cloud technologies. We also describe the relationships and distinction between the Grid and cloud approaches.

H. Liang, L.X. Cai based on the provided references, a comprehensive literature survey on cloud computing can be synthesized. Cloud computing has emerged as a transformative technology with significant implications for various domains. The literature review encompasses various dimensions of cloud computing, including technological advancements, security challenges, adoption frameworks, and regulatory impacts.

Senyo et al. (2018) highlighted the skewness of extant cloud computing literature towards technological dimensions, neglecting under-researched areas such as business and application domains. Table& Mohamed (2020) provided a comprehensive review of cloud computing trends, emphasizing its rapid growth and benefits for organizations. Yigitbasioglu (2015) contributed to the empirical literature on cloud computing adoption, offering an institutional framework for interpreting cloud-based information technology outsourcing. Asiaei & Rahim (2019) developed an integrative model to identify contextual factors influencing the adoption and usage of cloud computing in SMEs, particularly in developing countries.

Wang, N. Cao, k. Ren Enabling secure and efficient rank keyword search over outsourced cloud data owners are outsourcing their complex data management system from local sites to the commercial public cloud for greater flexibility and economic savings. But sensitive data has to be encrypted before outsourcing, for protecting data privacy. However, data has to be encrypted makes effective data utilization a challenging task.

III. SYSTEM OVERVIEW

The process of defining the architecture, components, modules, interfaces and data for a system to satisfy the specified requirement.to realize secure and efficient dynamic searchable symmetric encryption scheme over medical cloud data.

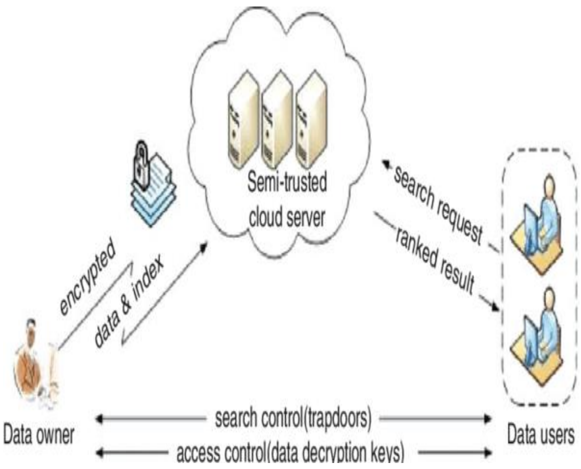


Fig 1.1: cloud accessing

A secure and efficient dynamic searchable symmetric encryption (SEDESE) scheme over medical cloud data. This work extend improves our previous research. Specifically, this paper addresses two new issues. Collision between the cloud server and search users as well as different secret key distribution among search user.in addition, we apply the new design to the health care system. Furthermore, the security and the performance are analyzed. the original contribution of the paper is:

1.firstly, we combine the k-Nearest Neighbor (KNN) an attribute-based encryption (ABE) Techniques of propose efficient and dynamic searchable symmetric encryption scheme name SEPSSE I. The proposed scheme can achieve forward privacy and backward privacy and collision resistance between the cloud server and the search users.

2.Secondly, based on the scheme the further propose the enhanced scheme, name SEPSSE II to solve the key sharing problem widely exist in the kNN based on searchable encryption scheme. Compared with existing DSSE schemes, our proposed schemes have less storage costs, search and updating complexity. Extensive experiment demonstrated the efficiency for our schemes in terms of storage overhead, index building trapdoor generating a query.

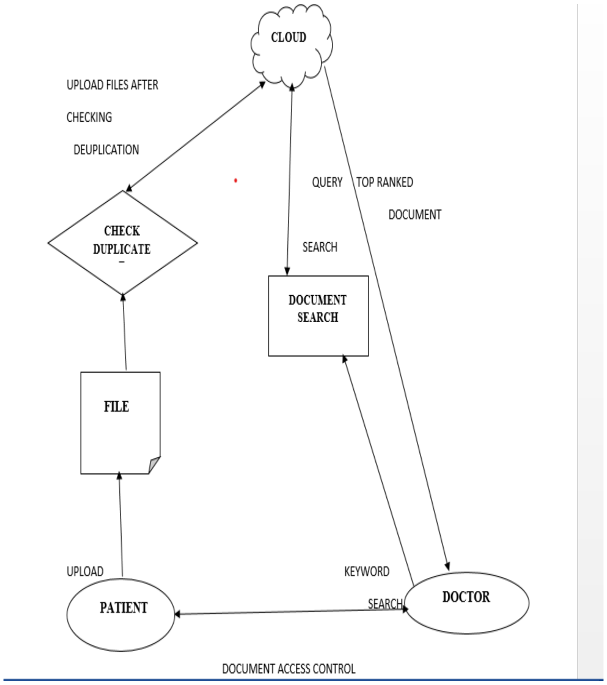


Fig 2.1: Architectural Diagram

1) TRUSTED AUTHORITY

A trusted authority (TA) is a trusted third party. We use it to generate attribute-based encryption (ABE) key to encrypt medical documents. Patient's documents will be encrypted and only some doctors that satisfy the corresponding access policy can decrypt them. The trusted authority in the cloud can deduplicate the files that are duplicate send by the patient that will be checked and send to the corresponding doctors who need the file.

2) PATIENT

A patient outsources the documents to the cloud server to provide convenient and reliable data access to the corresponding search doctors. To protect data privacy, the patient encrypts the original documents under an access policy using attribute-based encryption. To improve the search efficiency, she also generates according to the keywords using the secret key of secure KNN scheme after that the patient send the encrypted document and the corresponding index to the cloud server, and submit the secret key to the search doctors. The patient can send the same file content with different name to the cloud. That will de duplicate by the trusted authority.

3) CLOUD SERVER

A cloud server is intermediary entity which store the encrypted document and corresponding indexes received from the received patient, and then provides data access and search services to authorized search doctors. When a search doctor sends a trapdoor to the cloud server it could return a collection of matching documents based on certain operations. Performs the avoidance of duplication by Merkle hash function to avoid duplication.

4) DOCTOR

An authorized doctor can obtain the secret key from the patient, where this key can be used to generate trapdoors. When she needs to search the outsourced documents stored in the cloud server, she will generate a search keyword set then according to the keyword set the doctor uses the secret key to generate a trapdoor and sends it to the cloud server. Finally, she receives the matching document collection from the cloud server and decrypt them with the ABE key received from the trusted authority. After getting the health information of the patient the doctor can also outsource medical report to the cloud server by the same way. For simplicity we just consider one-way communication in our schemes. The doctor can view only one file with the same content by Merkle hash function method.

TOOLS

1) C#

C# is a general-purpose, object-oriented programming language designed by Microsoft as part of the .NET Framework. It is a statically typed, compiled language that is syntactically similar to C++ and Java. C# is used for a wide variety of applications, including web development, desktop applications, games, and more c# is a relatively easy language to learn, even for those with no prior programming experience. This is due to its clean syntax and well-defined semantics. However, C# is also a powerful language that can be used to create complex and sophisticated applications.

2) ASP.NET

ASP.NET offers three frameworks for creating web applications: Web Forms, ASP.NET MVC, and ASP.NET Web Pages. All three frameworks are stable and mature, and you can create great web applications with any of them. No matter what framework you choose, you will get all the benefits and features of ASP.NET everywhere. Each framework targets a different development style. The one you choose depends on a combination of your programming assets (knowledge, skills, and development experience), the type of application you're creating, and the development approach you're comfortable with.

3) SQL SERVER

Microsoft SQL Server is a relational database management system (RDBMS) developed by Microsoft. It is a software application that helps you organize, store, manage, and retrieve data in a relational database. A relational database is a collect*ion of tables that are linked together by relationships. This allows you to efficiently store and access data, and to create complex queries to retrieve specific information. SQL Server can store and retrieve large amounts of structured data in a relational format. This means that the data is organized into tables, which are made up of rows and columns. Each row represents a single record, and each column represents a specific field of data. SQL Server provides a powerful query language called Transact-SQL (T-SQL) that allows you to retrieve and manipulate data in the database.

4) DATABASE

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database. Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

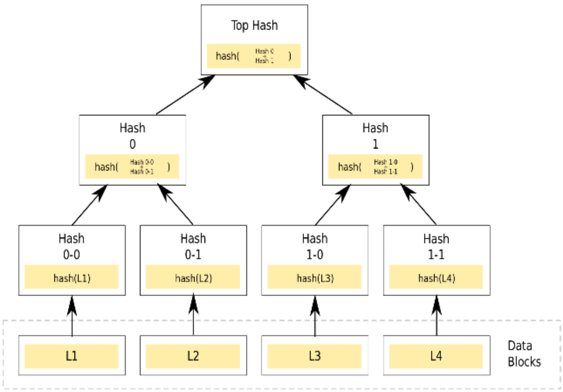
Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized.

5) VISUAL STUDIO

Integrated Development Environment (IDE) developed by Microsoft to develop Desktop applications, GUI (Graphical User Interface), console, web applications, mobile applications, cloud, and web services, etc. With the help of this IDE, you can create managed code as well as native code. It uses the various platforms of Microsoft software development software like Windows store, Microsoft Silverlight, and Windows API, etc. It is not a language-) specific IDE as you can use this to write code in C#, C++, VB (Visual Basic), Python, JavaScript, and many more languages. It provides support for 36 different programming languages. It is available for Windows as well as for macOS.

MERKLE HASH FUNCTION

A Merkle hash function isn't actually a single function, but rather a cryptographic technique used to ensure data integrity and authenticity. It's particularly useful in scenarios where data is distributed across multiple locations or needs to be verified without having access to the entire dataset. A Merkle tree is a hash-based data structure that is a generalization of the hash list. It is a tree structure in which each leaf node is a hash of a block of data, and each non-leaf node is a hash of its children. Typically, Merkle trees have a branching factor of 2, meaning that each node has up to 2 children. Merkle trees are used in distributed systems for efficient data verification. They are efficient because they use hashes instead of full files. Hashes are ways of encoding files that are much smaller than the actual file itself. Currently, their main uses are in peer-to-peer networks.



and input of data broken up into blocks labelled L1 though L4. Each of these blocks are hashed using some hash function. Then each pair of nodes are recursively hashed until we reach the root node, which is a hash of all nodes below it.

SIMPLIFIE PAYMENT VERIFICATION (SPV)

SPV is to afford users a trust-minimized way of examining the block chain without the inconvenience of running a node. SPV clients do trust the nodes through which they query the block chain, but the ability to abuse this trust is extremely limited thanks to the cryptographic security of Merkle trees and the presence of other nodes. A Merkle tree is a data structure with unique properties which make it useful for Bitcoin. Merkle trees are used to store all transactions in a given block. The advantage of this system is that one node can easily prove to another that a given transaction was contained in a specific block. This is useful for SPV nodes and light clients, who do not store the entire block chain and are only interested in certain transactions or blocks.

HARDWARE SPECIFICATION

System: i3 processor
Hard Disk: 500GB
Monitor: 14’Colour Monitor
Mouse: optical Mouse
Ram: 4 GB

SOFTWARE SPECIFICATION
Operating System: Windows 7 Ultimate
Front-End: ASP.Net with C#
Database: SQL Server 2008

IV. CONCLUSION

We propose two high-security dynamic searchable encryption systems. The first bone can accomplish both forward and backward privacy in addition to preventing cooperation between the cloud server and search users. The second one addresses the issue of key sharing that is prevalent in searchable encryption schemes based on Nonperformance analysis shows that, in terms of storage, search, and updating complexity, the suggested schemes can outperform the current works in terms of efficiency. Our schemes have been shown effective in reducing storage overhead, index building, trapdoor producing, and query through numerous testing. We have also received funding from the State Key Laboratory of Information Security Foundation Open Foundation under grant 2015-ms.

V.REFERENCES

[1] X. Xie, X. Yang, X. Wang, H. Jin, D. Wang, and X. Ke, "BFSI-B: an improved K-hop graph reachability queries for cyber-physical systems," *Information Fusion*, vol. 38, pp. 35–42, 2017.

[2] L. Qi, X. Wang, X. Xu, W. Dou, and S. Li, "Privacy-aware cross-platform service recommendation based on enhanced locality-sensitive hashing," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1145–1153, 2021.

[3] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, 2018.

[4] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on Systems, Man, And Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[5] S. Zhang, "Nearest neighbor selection for iteratively KNN imputation," *Journal of Systems and Software*, vol. 85, no. 11, pp. 2541–2552, 2012.

[6] X. Wu, V. Kumar, and J. R. Quinlan, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.

[7] T. Wang, Z. Qin, S. Zhang, and C. Zhang, "Cost-sensitive classification with inadequate labelled data," *Information Systems*, vol. 37, no. 5, pp. 508–516, 2012.

[8] H. Zhou, N. Li, X. Che, and X. Yang, "Multi-key fully homomorphic encryption scheme over prime power cyclotomic rings," *Net info Security*, vol. 20, no. 5, pp. 83–87, 2020.

[9] N. Li, H. Zhou, X. Che, and X. Yang, "Design of directional decryption protocol based on multi-key fully homomorphic encryption in cloud environment," *NetinfoSecurity*, vol. 20, no. 6, pp. 10–16, 2020.

[10] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," pp. 409–437, Hong Kong, China, December 2017.

[11] G. Góra and A. Wojna, "RIONA: a classifier combining rule induction and k-NN method with automated selection of optimal neighborhood," in *Proceedings of the European Conference on Machine Learning*, pp. 111–123, Helsinki, Finland, August 2002.

[12] B. Li, Y. W. Chen, and Y. Q. Chen, "The nearest neighbor algorithm of local probability centers," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 1, pp. 141–154, Feb. 2008.

[13] J. Wang, P. Deskovic, and L. N. Cooper, "Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence," *Pattern Recognition*, vol. 39, no. 3, pp. 417–423, 2006.

[14] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient k NN classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, 2016.

[15] S. Zhang, X. Li, M. Zong M, X. Zhu, R. Wang, and X. Zhu, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2017.