IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Ensemble Multiple-Keyword Search With Data Security In Cloud Environment

Afsha Jabeen[1],

Mallareddy Engineering College for Women's

Dr. C.V.P.R.Prasad[2],

Head, Dept of CSE, Mallareddy Engineering College for Women's

ABSTRACT: With the increasing adoption of cloud computing, data security has become a significant concern for organizations and individuals. In this context, efficient and secure keyword search techniques are essential to retrieve relevant information from encrypted data stored in the cloud. This paper proposes an ensemble multiple-keyword search approach with enhanced data security in a cloud environment. The proposed system combines the strengths of multiple keyword search techniques to achieve better search accuracy and efficiency. The ensemble method leverages various algorithms, such as inverted index, bloom filters, and homomorphic encryption, to overcome the limitations of individual techniques and provide robust search capabilities. The proposed approach employs encryption mechanisms to protect sensitive information to ensure data security. The data is encrypted before being uploaded to the cloud, ensuring only authorized users can access and retrieve the information. Homomorphic encryption enables the cloud server to perform search operations on the encrypted data without decrypting it, preserving the privacy of the stored information.

Furthermore, the proposed approach addresses the issue of keyword privacy by introducing a bloom filter-based technique. The bloom filter masks the keywords used in the search query, preventing unauthorized parties from gaining insights into the user's search intentions. This technique adds a layer of privacy protection to the search process. Experiments were conducted using real-world datasets to evaluate the effectiveness of the proposed ensemble multiple-keyword search approach. The results demonstrate that the system achieves high search accuracy and efficiency while maintaining data security.

Keywords: Cloud computing, Symmetric-key, homomorphic, and Attribute-based Encryption (ABE).

I. INTRODUCTION

Cloud computing has revolutionized how businesses and individuals store, access, and process data [1]. It refers to delivering computing resources, such as storage, processing power, and software applications, over the Internet on a pay-as-you-go basis. This technology has gained immense popularity due to its flexibility, scalability, and cost-effectiveness [2]. One of the critical advantages of cloud computing is its ability to handle vast amounts of data efficiently. With the ever-increasing volume of information generated in today's digital age, traditional data storage and retrieval methods still need to be improved. Cloud computing offers a solution by providing virtually unlimited storage capacity and powerful processing capabilities. Another significant feature of cloud computing is its support for multiple keyword searches [3]. Keyword search is a fundamental method for retrieving information from large datasets. However, traditional keyword search techniques often have limitations, such as low precision and recall rates, especially when dealing with complex or unstructured data [4].

Cloud computing platforms leverage advanced algorithms and distributed computing architectures to improve the efficiency and accuracy of keyword searches. Cloud-based systems can process queries in parallel by distributing the search workload across multiple nodes or servers, significantly reducing search times. This distributed approach also enables scalability, allowing the system to handle large search requests simultaneously. Moreover, cloud-based keyword search systems often employ techniques like indexing and caching to enhance search performance [5]. By creating indexes of the data and storing frequently accessed information in cache memory, search operations can be executed more quickly and efficiently.

Combining cloud computing and multiple keyword search opens up new possibilities for data-driven applications and services. Businesses can leverage these technologies to perform complex data analytics, conduct market research, and gain valuable insights from their data repositories. Additionally, individuals can benefit from faster and more accurate information retrieval, enabling them to find relevant content efficiently [6].

The main aim of this paper is to provide the ensemble security for the cloud data based on the multiple-keyword searching using advanced searching technique.

II. LITERATURE SURVEY

N. Cao et al. [7] proposed a new security model that searches the encrypted data by using a multi-keyword to protect the data in a cloud server by utilizing the advanced similarity measures combined with "coordinate matching," which helps to match the maximum number of documents based on the given keyword. The proposed model also focused on reducing overloading and memory management. Thus, the proposed approach achieved better search results compared with existing models. Y. Zhang et al. [8] proposed the public-key-based encryption (PKBE) model that solves various issues securing data. The PKBE mainly focused on detecting the keyword analyzing attacks that finds the low-entropy. The performance of the proposed model is integrated with the blockchain (e.g., Ethereum), which explores the total number of requests received from several types of users. The proposed system provides high security to the data stored in a cloud server. Y. Zhang et al. [9] proposed verifying data present in the blockchain. Blockchain transactions are very secure and time-sensitive because of the combination of blockchain that enables data auditing. The proposed approach finally built the integrated security that secures the data efficiently. Y. Yang et al. [10] proposed a new security model that solves the escrow problem in cloud security data in cloud servers. The proposed model has used better encryption and decryption techniques that compute the energy management among the cloud servers. The proposed system, combined with the revocation model, detects malicious users among the cloud servers. An efficient searching process is introduced to search the data and provide security for the cloud data. S. K. Ocansey et al. [11] developed a framework to secure the data based on privacy and searching for outsourced data. The proposed framework follows the two steps, such as exploring the data using Searchable symmetric encryption (SSE) that supports the updated functions in cloud computing. The updated operations mainly focused on multi-keyword searching using the KNN and Bloom Filter (BF) to get a better ranking search technique to retrieve the files. F. Wang et al. [12] proposed advanced keyword-based encryption (AKBE) that can detect attackers from cloud users. The proposed AKBE mainly checks the authentication and authorization of the users and provides fine-grained access to data by updating the access key generation. The proposed solves the Diffie-Hellman issue in securing the data and performing high computation efficiency. R. Chen et al. [13] proposed a new encryption model that secures the data based on keyword guessing attacks (KGA). The proposed model creates the keyword-based ciphertext that provides the authenticated model combined with the keyword server. The security is improved by using the RSA algorithm. C. Guo et al. [14] proposed an advanced keyword search that protects by using the encrypted key to provide the cloud data by using the Bloom filter. Thus, the proposed model offered high security for the cloud data.

III. HOMOMORPHIC ENCRYPTION

Homomorphic encryption is a cryptographic technique that allows computations to be performed on encrypted data without decrypting it first. It enables secure computation on sensitive information while preserving privacy. The term "homomorphic" refers to the property that operations performed on encrypted data produce results that, when decrypted, correspond to the same operations performed on the plaintext data. Traditional encryption methods require data to be decrypted before any computation can be performed on it. This poses a significant security risk, especially when dealing with sensitive data. Homomorphic encryption addresses this challenge by providing a way to perform computations on encrypted data directly, without revealing the underlying information. Homomorphic encryption schemes are designed to support various mathematical operations, such as addition and multiplication, while maintaining the confidentiality of the encrypted data. These schemes involve complex mathematical concepts, including lattice-based cryptography or the use of mathematical structures like rings and groups. Homomorphic encryption has numerous applications, particularly in scenarios where privacy is crucial, such as secure cloud computing, confidential data processing, and secure outsourcing of computations. By enabling computations on encrypted data, homomorphic encryption ensures data privacy while still allowing valuable insights to be derived from the encrypted information. Here are the general steps involved in using homomorphic encryption:

Key Generation: Generate the necessary cryptographic keys for the homomorphic encryption scheme. This typically involves generating a public key and a private key.

Encryption: Encrypt the data you want to perform computations on using the public key. The encrypted data will be transformed into a ciphertext that cannot be easily understood or manipulated without the corresponding private key.

Computation: Perform the desired computations on the encrypted data without decrypting it. This is the key feature of homomorphic encryption. The computations are performed directly on the encrypted ciphertext.

Decryption: After the desired computations have been performed, decrypt the result using the private key. This step is necessary to obtain the final result in a readable format.

IV. METHODOLOGY

Ensemble security refers to the concept of combining multiple security mechanisms or techniques to enhance the overall security of a system or application. One approach to achieving ensemble security is by integrating homomorphic encryption and multiple keyword searching methodologies. Homomorphic encryption is a cryptographic technique that allows computations to be performed on encrypted data without decrypting it first. This means that sensitive data can be stored and processed in an encrypted form, protecting it from unauthorized access. By incorporating homomorphic encryption into a system, the data remains secure even when computations or searches are performed on it. When it comes to keyword searching, there are various methodologies that can be employed. One commonly used approach is the use of search indexes or inverted indexes, which allow for efficient retrieval of documents or data based on specific keywords or terms. Another approach is using techniques like tokenization or stemming to preprocess the data and enable efficient keyword matching.

To combine homomorphic encryption and multiple keyword searching methodologies, one possible approach is as follows:

Data Encryption: Apply homomorphic encryption to the sensitive data that needs to be protected. This can be done using an appropriate homomorphic encryption scheme such as partially homomorphic encryption or fully homomorphic encryption, depending on the specific requirements.

Indexing and Tokenization: Generate indexes or inverted indexes for the encrypted data. This involves extracting keywords or terms from the encrypted data and creating a searchable index structure. Additionally, tokenization can be applied to preprocess the data and generate tokens that represent individual terms or keywords.

Keyword Search: When a search query is received, encrypt the query using the same homomorphic encryption scheme used for the data. Apply the necessary operations on the encrypted query to match it against the encrypted indexes or tokens generated in the previous step. This can involve operations like comparison, pattern matching, or similarity computations performed on the encrypted data.

Result Decryption: Once the matching process is complete, the encrypted results can be retrieved. Decrypt the encrypted results using the appropriate decryption key to obtain the original data in a readable format. It's important to ensure that the decryption process is performed securely and the decrypted results are protected from unauthorized access.

By combining homomorphic encryption with multiple keyword searching methodologies, sensitive data can be securely stored and searched without compromising its confidentiality. This ensemble security approach allows for efficient and privacy-preserving searching while maintaining the confidentiality of the underlying data.

V. FUNCTIONALITIES OF CLOUD & EXPERIMENTAL RESULTS

The experiments are conducted by using the JAVA code and cloud simulator is used to simulate the cloud. Totally 30 files are uploaded by the data owners and these files belongs to 10 data owners. These files are secured and stored in cloud server. Every file generated by the encryption key and downloaded by the decryption key. This section provides the following functionalities for the cloud users.

A. Data Owner

A data owner is a module in this case. The data owner must register with the application and then be authorized by the administration server before he can log in to the application. After successfully logging in, he can perform some operations such as encrypting files and uploading them to the cloud, viewing all uploaded files, encrypting indexes and submitting them to the Administration Server, viewing decrypt requests and sharing the decrypt key with the requested user, and logging out.

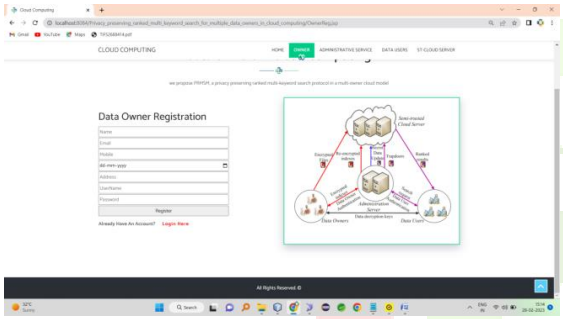

Figure 2: Data owner registration for the cloud server.

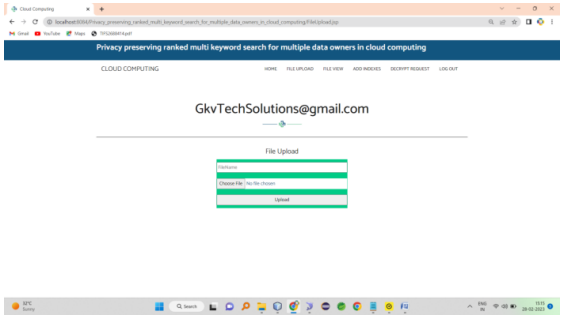

Figure 3: Upload the data file by the data owner and the file is encrypted with Enhanced Security Algorithm.

Table 1: Shows the file size and encryption time (milli seconds) for the files uploaded by the data owner.

File Name	File Size (KB)	Encryption Time (Milli Seconds)
Document-1	18.5	9.34
Document-2	21.56	13.87
Document-3	31.87	16.67
Document-4	23.34	14.56
Document-5	36.23	21.34
Document-6	27.56	17.23
Document-7	29.34	18.34

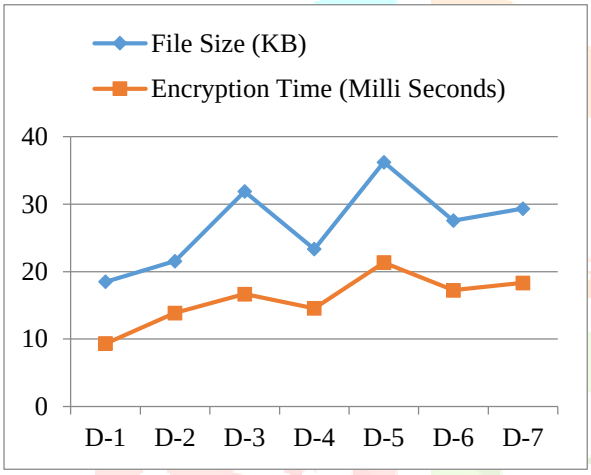

Figure 4: Performance of Enhanced Security Algorithm in terms of Encryption time for 1-7 documents.

B. Data User

Here A module is a data user. A data user must register with the application and then be authorized by the administration server before he can log in to the application. After successfully logging in, he can perform some operations such as viewing his profile, entering a search query and encrypting it before submitting it to the administration server, considering the response from the file, and sending a request for the decrypt key to the data owner, and verifying the decrypt key. If the decrypt key is successfully verified, we can decrypt the file download and view all my download files before logging out.

Table 2: Shows the file size and decryption time (milli seconds) for the files downloaded by the users.

File Name	File Size (KB)	Decryption Time (Milli Seconds)
Document-1	18.5	11.34
Document-2	21.56	14.23
Document-3	31.87	17.37
Document-4	23.34	13.45
Document-5	36.23	19.87
Document-6	27.56	16.34
Document-7	29.34	21.23

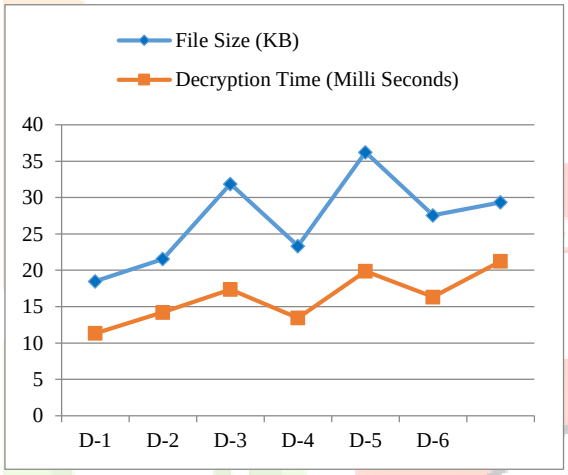

Figure 5: Performance of Enhanced Security Algorithm in terms of Decryption time for 1-7 documents.

C. Administration Server

Here, the Administration Server can directly log in with the application. After successful login, he can perform some operations such as viewing all users and authorizing them, viewing all owners and allowing them, viewing indexes and re-encrypting indexes and submitting them to the cloud server, viewing search queries submitted by data users and re-encrypting the search queries and introducing them to the cloud server, and finally logging out.

D. Semi Trusted Cloud Server

Semi-Trusted Cloud Server can first directly login with the application, and after successful login, he can perform some operations such as viewing all uploaded files, viewing all uploaded file indexes, viewing all search requests and decrypting the search query, filtering the files, and finding high ranked files to share with the user, and the cloud can check all downloaded files at the end of the session. Table 3 shows the ranking of files based on the total number of downloads from cloud server.

Table 3: Shows the ranking and total number of times the files downloaded

File Name	Ranking	Total times Downloaded
Document-1	3	23
Document-2	7	5
Document-3	5	11
Document-4	2	28
Document-5	4	19
Document-6	1	35
Document-7	6	8

V. CONCLUSION

In this paper, an ensemble multiple-keyword search with data security in a cloud environment is a practical approach to enhance search functionality while maintaining the confidentiality of sensitive information. This technique allows users to search for multiple keywords simultaneously, enabling more efficient and accurate search results. Encryption techniques such as homomorphic or searchable encryption can protect the data stored in the cloud from unauthorized access. These encryption methods ensure that the search queries and the corresponding results remain confidential, even when processed by the cloud service provider. Ensemble techniques, which combine multiple search algorithms or models, further improve search accuracy and efficiency. By leveraging the strengths of different algorithms, ensemble search can provide complete and precise results, enhancing the overall user experience. Data security is critical to any cloud-based system, especially when dealing with sensitive or confidential information. The encryption mechanisms employed in ensemble multiple-keyword searches guarantee the privacy of user queries and data stored in the cloud.

Access control measures, such as role-based or attribute-based access control, can be implemented to ensure that only authorized users can access the encrypted data. However, it is essential to note that implementing ensemble multiple-keyword search with data security in a cloud environment requires careful consideration of factors such as computational overhead, scalability, and the trade-off between security and search efficiency. These aspects should be evaluated to balance privacy protection and system performance.

IV. REFERENCES

[1] S. K. Ojha, H. Rai, P. N. Tripathi and A. Nazarov, "Security algorithm on Cloud Storage System," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 822-826, doi: 10.1109/ICACCCN51052.2020.9362792.

[2] D. Das, R. Amin and S. Kalra, "Algorithm for Multi Keyword Search Over Encrypted Data in Cloud Environment," 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020, pp. 733-739, doi: 10.1109/IWCMC48107.2020.9148472.

[3] R. Zhang, R. Xue, L. Liu and L. Zheng, "Oblivious Multi-Keyword Search for Secure Cloud Storage Service," 2017 IEEE International Conference on Web Services (ICWS), Honolulu, HI, USA, 2017, pp. 269-276, doi: 10.1109/ICWS.2017.42.

[4] D. ManJiang, C. Kai, W. ZengXi and Z. LiPeng, "Design of a Cloud Storage Security Encryption Algorithm for Power Bidding System," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 1875-1879, doi: 10.1109/ITNEC48623.2020.9085095.

[5] F. -J. Hsieh, T. -L. Chin, C. -Y. Huang, S. -H. Shen and C. -A. Shen, "Semantic Multi-Keyword Search over Encrypted Cloud Data with Privacy Preservation," 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 2019, pp. 1-5, doi: 10.1109/VTCFall.2019.8891143.

[6] D. Aritomo and C. Watanabe, "Achieving Efficient Similar Document Search over Encrypted Data on the Cloud," 2019 IEEE International Conference on Smart Computing (SMARTCOMP), Washington, DC, USA, 2019, pp. 1-6, doi: 10.1109/SMARTCOMP.2019.00020.

[7] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222-233, Jan. 2014, ISBN 1045-9219.

[8] Y. Zhang, C. Xu, J. Ni, H. Li and X. S. Shen, "Blockchain-Assisted Public-Key Encryption with Keyword Search Against Keyword Guessing Attacks for Cloud Storage," in IEEE Transactions on Cloud Computing, vol. 9, no. 4, pp. 1335-1348, 1 Oct.-Dec. 2021, doi: 10.1109/TCC.2019.2923222.

[9] Y. Zhang, C. Xu, X. Lin and X. Shen, "Blockchain-based public integrity verification for cloud storage against procrastinating auditors", IEEE Trans. Cloud Comput., pp. 1-15, 2019.

[10] Y. Yang, X. Liu, X. Zheng, C. Rong and W. Guo, "Efficient traceable authorization search system for secure cloud storage", IEEE Trans. Cloud Comput., pp. 1-14, 2018.

[11] S. K. Ocansey and C. Wang, "Search Over Encrypted Cloud Data With Secure Updates," 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia, Bulgaria, 2019, pp. 380-386, doi: 10.1109/QRS-C.2019.00077.

[12] F. Wang, T. Shi and S. Li, "Authorization of Searchable CP-ABE Scheme with Attribute Revocation in Cloud Computing," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 204-208, doi: 10.1109/ITAIC.2019.8785559.

[13] R. Chen, Y. Mu, G. Yang, F. Guo, X. Huang, X. Wang, et al., "Server-aided public key encryption with keyword search", IEEE Trans. Inf. Forensics Secur., vol. 11, no. 12, pp. 2833-2842, Dec. 2016.

[14] C. Guo, R. Zhuang, C. -C. Chang and Q. Yuan, "Dynamic Multi-Keyword Ranked Search Based on Bloom Filter Over Encrypted Cloud Data," in IEEE Access, vol. 7, pp. 35826-35837, 2019, doi: 10.1109/ACCESS.2019.2904763.

[15] Y. Lu, J. Li and Y. Zhang, "Privacy-Preserving and Pairing-Free Multirecipient Certificateless Encryption With Keyword Search for Cloud-Assisted IIoT," in IEEE Internet of Things Journal, vol. 7, no. 4, pp. 2553-2562, April 2020, doi: 10.1109/JIOT.2019.2943379.