Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue S-3 Year 2024 Page 121-125

A Robust Search Method Using Features To Determine Combined Keywords On Cloud Encrypted Data

Y.K. Viswanadham[1*], G Naga Lakshmi[2], G Dinesh Kumar[3], B Archana[4], B Sravanthi[5]

[1*,2,3,4,5]*Associate Professor, Department of Information Technology, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru Andhra Pradesh-521356, India.* [1]*Email: ykvnath@gmail.com*
[2]*Email: gurralanagalakshmi17@gmail.com,* [3]*Email: dineshkumargudiputi777@gmail.com,*
[4]*Email: archana.bolla2210@gmail.com,* [5]*Email: sravanthisravanthi4091@gmail.com*

***Correspondence Author:** Y.K. Viswanadham*
**Associate Professor, Department of Information Technology, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru Andhra Pradesh-521356, India. Email: ykvnath@gmail.com*

Abstract

Users are more comfortable trusting their sensitive information to the cloud as its security continues to improve. However, when there are several encrypted files, each with its own set of keywords for indexing, the storage overhead grows exponentially, and search efficiency suffers. Therefore, this work provides a technique for searching encrypted cloud data that makes use of features to match joint keywords (FMJK). Joint keywords are generated by randomly selecting a subset of the data owner's non-duplicated keywords choice among the documents' extracted keywords; together, these keywords form a keyword dictionary. Every combined keyword matches with a document's feature as well as a query keyword, making the former's result considered a dimension of a document's index with the latter's result considered a dimension about the query trapdoor. Its BM25 method is then utilized for arranging the top k results by the inner product between the document index and the trapdoor.

Key words: Encrypted cloud data, feature matching, searchable encryption, dimensionality reduction, joint keywords.

1. INTRODUCTION

With the fast growth of science and technology, organizations or individuals increasingly depend on keeping a huge number of data documents on cloud servers for the purpose to transfer data swiftly and remotely. But as use of cloud services grows, storage costs rise, search efficiency declines, and privacy safeguards become an area of study. KNN (K Nearest Neighbor) technology is utilized to construct indexes allowing cipher-text retrieval in the majority of current cipher-text sorting retrieval techniques. Most search encryption systems need a lot of space and time to implement since they are directly tied to the encrypted key, the document's index, plus the query request dimensions involved in searching encrypted material. Encrypting less data in fewer dimensions is one way to speed up the search process. The retrieval needs of a huge quantity of data are still beyond the capabilities of current research, and neither can academics sort and filter valuable data only authorized users. Since there are several types of users, it is critical to develop a system that can maintain privacy, boost retrieval efficiency, and improve query accuracy.

In this research, we offer the MRSE (Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data) technique, which is a search method that uses features to match joint keywords (FMJK) on encrypted cloud data. Initially each document's topic has to be expressed by the characteristics extracted from it. Then, d random keywords are generated from the data owner's documents' non-duplicated keywords to create a joint keyword, and the entire joint keywords constitute a keyword dictionary, and finally, the features of each document are combined with the joint keywords to generate an index. The search query keywords are entered by the allowed user, matched against the joint keywords to generate a trapdoor, and then the safe inner product between the trapdoor and index is computed to yield the top k results. The article in question makes the following contributions:
(1) Each set of d randomly chosen keywords forms a joint keyword that corresponds to a characteristic of the document that is to be mapped with a particular index dimension, thus decreasing the overall dimension size of the key, the index, as well as the trapdoor, streamlining the matrix operation throughout encryption, and enhancing search efficiency.
(2) Through the upgraded BM25 method to generate the inner product of the document index as well as the trapdoor, and this not only sorts rapidly but also assure query correctness.
(3) The encryption procedure of expanding and dividing, together with the unpredictability of the combined keywords, guarantees confidentiality.

2. LITERATURE SURVEY

In "VPSearch," a novel scheme is presented for ensuring verifiability in privacy-preserving multi-keyword searches over encrypted cloud data. By combining homomorphic MAC and multi-keyword search techniques, the scheme allows clients to efficiently verify search results without storing data locally. Experiments demonstrate minimal overhead and quick query processing, making it a promising solution for secure cloud-based searches [1].
This paper introduces a Blockchain-based Multi-Keyword Ranked Search with Fair Payment (BMFP) system to address issues of data confidentiality and trust in cloud computing. Utilizing smart contracts, it ensures the accuracy of search results and automates fair payment processes, enabling public verifiability for multi-keyword searches. The system is Ethereum-compatible and employs cost-effective verification algorithms [2].
This paper addresses the challenge of verifying secure ranked keyword search results in cloud computing, considering potential server dishonesty. It introduces a novel deterrent-based scheme with carefully devised verification data to detect misbehavior without revealing data owners' identities. Thorough analysis and experiments confirm the scheme's effectiveness and efficiency [9].
This paper addresses secure ranked keyword search in cloud computing with a focus on result verification in the face of potentially dishonest cloud servers. It presents a novel deterrent-based scheme that conceals data owner identities and data usage in the verification process, ensuring high probability detection and severe punishment for server misbehavior. Thorough analysis and experiments validate the scheme's efficacy and efficiency [10].

3. RESEARCH METHODOLOGY

The cloud server, the data owner, and the data user are three separate entities that make up the model of the searchable encryption system. The data owner begins by considering all data files like documents and then identifies relevant terms inside those files. Then, it takes those d keywords and merges them into a single joint keyword; together, these joint keywords constitute a dictionary of keywords. Third, it encrypts both the original document as the index it was based on using the same key that was used to encrypt the original document. The data owner then stores the encrypted files and indexes on a cloud server. If an authorized data user enters numerous keywords during a query, a keyword trapdoor for each will be generated and sent through the cloud server for processing.
A cloud server will send the top k encrypted documents containing the highest score out to the authorized data user following receiving a search request. This score is determined by calculating the security inner product based on the keyword trapdoor over each document index. Once the documents have been delivered, the data user will use the keys supplied by the data owner to decrypt that documents and access the necessary data.

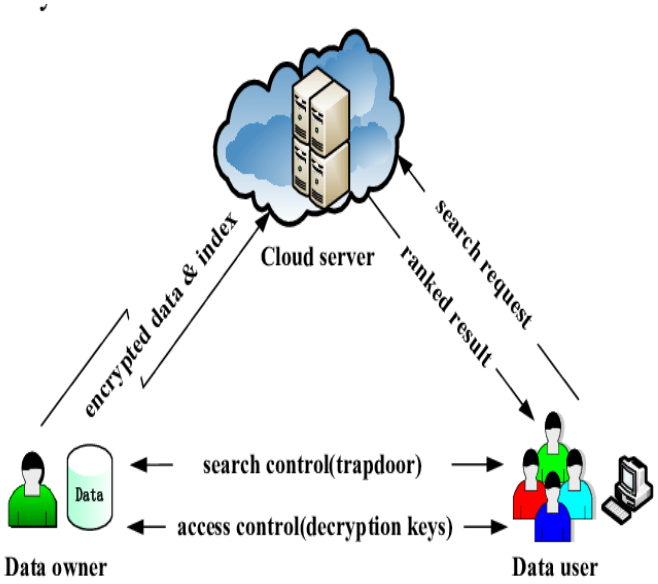

Fig 1: System Architecture

The FMJK scheme's reduction of the keyword dictionary dimension from the outset yields substantial improvements in the entire process, encompassing dimension expansion, segmentation, and index encryption. The results underscore that the FMJK scheme excels in terms of speed, mainly due to its innovative approach where every 7 keywords form a joint keyword. This strategy significantly reduces the dimension required for generating trapdoors, contributing to its remarkable speed. Furthermore, the FMJK scheme offers an invaluable advantage in ensuring query accuracy, a crucial aspect in keyword search over encrypted data. Additionally, the storage cost of trapdoors generated by query keywords remains unaffected by the number of documents in storage space, which is a considerable benefit for scalable and efficient cloud-based data retrieval. These findings affirm the FMJK scheme as a standout solution for addressing the challenges of secure, efficient, and accurate keyword search in cloud computing environments, presenting substantial advantages for data owners, cloud providers, and end-users alike.

4. RESULTS AND DISCUSSIONS

Fig 2: Home page

For Data Owners, our innovative approach safeguards your sensitive data, ensuring it remains confidential, while simultaneously streamlining the data retrieval process with our feature-based search method. By minimizing data exposure and maximizing search speed, we empower you to benefit from efficient and secure data access.

Cloud Servers, which are integral to ensuring the appeal of cloud services, can significantly enhance their efficiency and reliability with our approach. Secure keyword search becomes swift and dependable, resulting in an improved service offering.

As Data Users, you can enjoy efficient and precise data retrieval, with our method ensuring your data remains secure. Experience an optimized user journey, delivering the data you need precisely when you need it. Our approach combines efficiency and data security, providing an enhanced user experience.

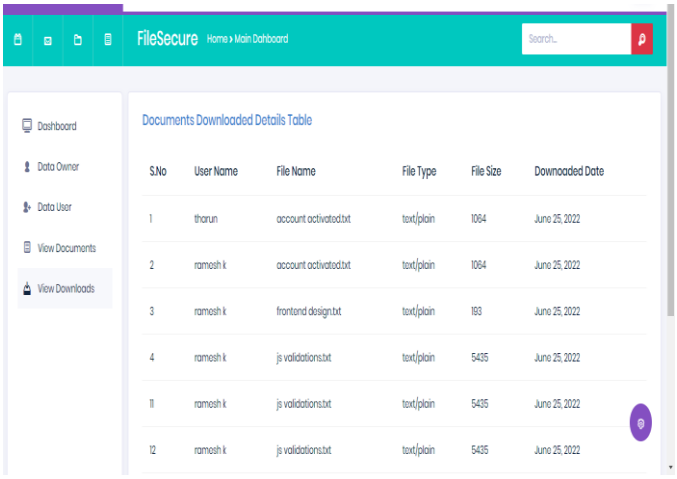

Fig 3: Downloaded information

Table with user details, including user name, filename, file type, file size, and downloaded date information, this data can be used to track and manage file downloads. This table provides valuable insights into user interactions with your system and allows you to monitor and analyze file usage.

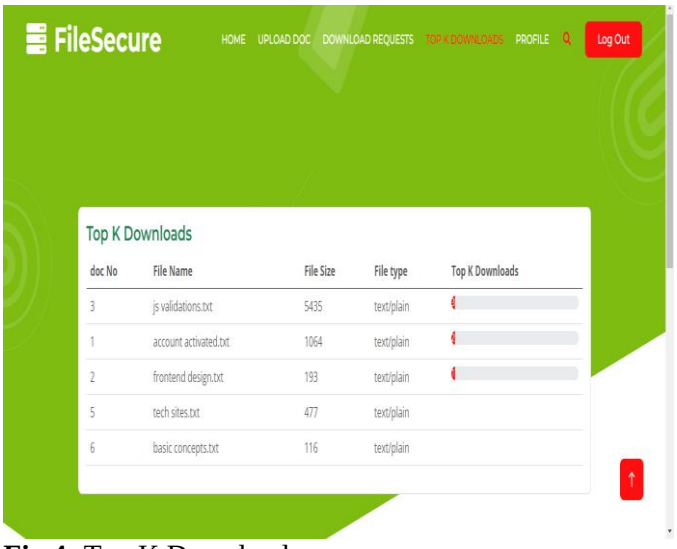

Fig 4: Top K Downloads

This comprehensive table combines document-specific information with download statistics, enabling efficient tracking, resource management, and user engagement enhancement.
Top K Downloads indicates the ranking of the file in terms of popularity based on the number of downloads.

5. CONCLUSION

In this research, we supply a search strategy for encrypted cloud data that makes use of features to match joint keywords (FMJK). Joint keywords arise by randomly selecting a subset of non-duplicated keywords retrieved from the data owner's documents; together, these keywords create a keyword dictionary, drastically decreasing its dimensionality. Since the key, index, and trapdoor dimensions are all tied to the keyword dictionary dimension, decreasing the latter helps boost search efficiency. Accurate matching between document characteristics and query keywords and the joint keywords in the keywords dictionary yields a weighted score, ensuring that only relevant documents are returned in response to a given query. Additionally, the storage space used up by indexes and trapdoors is decreased as a result of the dimension reduction. Experimental and theoretical findings demonstrate that the suggested strategy outperforms competing approaches.

REFERENCES:

1. Z. Wan and R. H. Deng, ‘‘VPSearch: Achieving verifiability for privacy preserving multi-keyword search over encrypted cloud data,’’ IEEE Trans. Depend. Secure Comput., vol. 15, no. 6, pp. 1083–1095, Nov./Dec. 2016.
2. Y. Yang, H. Lin, X. Liu, W. Guo, X. Zheng, and Z. Liu, ‘‘Blockchain based verifiable multi-keyword ranked search on encrypted cloud with fair payment,’’ IEEE Access, vol. 7, pp. 140818–140832, 2019.
3. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, ‘‘Privacy-preserving multikeyword ranked search over encrypted cloud data,’’ in Proc. IEEE INFOCOM, Apr. 2011, pp. 829–837.
4. W. K. Wong, D. W.-L. Cheung, B. Kao, and N. Mamoulis, ‘Secure kNN computation on encrypted databases,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, Jun. 2009, pp. 139–152.
5. Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, ‘‘Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing,’’ IEICE Trans. Commun., vol. E98.B, no. 1, pp. 190–200, 2015.
6. Z. Xia, X. Wang, X. Sun, and Q. Wang, ‘‘A secure and dynamic multi keyword ranked search scheme over encrypted cloud data,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 2, pp. 340–352, Jan. 2016.
7. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, ‘‘Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,’’ IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 11, pp. 3025–3035, Nov. 2014.
8. L. Liu and Z. Liu, ‘‘A novel fast dimension-reducing ranked query method with high security for encrypted cloud data,’’ Chin. J. Electron., vol. 29, no. 2, pp. 344–350, Mar. 2020.
9. W. Zhang, Y. Lin, and G. Qi, ‘‘Catch you if you misbehave: Ranked keyword search results verification in cloud computing,’’ IEEE Trans. Cloud Comput., vol. 6, no. 1, pp. 74–86, Mar. 2015.
10. Z. Guan, X. Liu, L. Wu, J. Wu, R. Xu, J. Zhang, and Y. Li, ‘‘Cross-lingual multi-keyword rank search with semantic extension over encrypted data,’’ Inf. Sci., vol. 514, pp. 523–540, Apr. 2020.
11. M. Murata, H. Nagano, R. Mukai, K. Kashino, and S. Satoh, ‘‘BM25 with exponential IDF for instance search,’’ IEEE Trans. Multimedia, vol. 16, no. 6, pp. 1690–1699, Oct. 2014.
12. D. Xiaoding Song, D. Wagner, and A. Perrig, ‘‘Practical techniques for searches on encrypted data,’ in Proc. IEEE Symp. Secur. Privacy (S&P), May 2000, pp. 44–55.
13. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, ‘‘Searchable symmetric encryption: Improved definitions and efficient constructions,’’ J. Comput. Secur., vol. 19, no. 5, pp. 895–934, Jan. 2011, doi: 10.3233/JCS-2011- 0426.
14. R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, ‘‘Efficient multi-keyword ranked query over encrypted data in cloud computing,’ Future Gener. Comput. Syst., vol. 30, no. 1, pp. 179–190, Jan. 2014, doi: 10.1016/j. future.2013.06.029.
15. J. Wang, H. Ma, T. Qiang, L. Jin, H. Zhu, S. Ma, and X. Chen, ‘‘A new efficient verifiable fuzzy keyword search scheme,’’ J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl., vol. 3, no. 4, pp. 61–71, Dec. 2012.
16. Z. Fu, X. Wu, C. Guan, X. Sun, and K. Ren, ‘‘Toward efficient multikeyword fuzzy search over encrypted outsourced data with accuracy improvement,’’ IEEE Trans. Inf. Forensics Security, vol. 11, no. 12, pp. 2706–2716, Jul. 2016, doi: 10.1109/TIFS.2016.2596138.
17. Z. Fu, F. Huang, K. Ren, J. Weng, and C. Wang, ‘‘Privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 8, pp. 1874–1884, Aug. 2017, doi: 10.1109/TIFS.2017.2692728.
18. R. Zhao, H. Li, Y. Yang, and Y. Liang, ‘‘Privacy-preserving personalized search over encrypted cloud data supporting multi-keyword ranking,’’ in Proc. 6th Int. Conf. Wireless Commun. Signal Process. (WCSP), Oct. 2014, pp. 1–6.
19. C. Chen, X. Zhu, P. Shen, J. Hu, S. Guo, Z. Tari, and A. Y. Zomaya, ‘‘An efficient privacy-preserving ranked keyword search method,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 4, pp. 951–963, Apr. 2016.
20. RFC Index. Accessed: May 25, 2020. [Online]. Available: https://www.rfc-editor.org/rfc-index-100a.html.