International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Multiple Keyword Search in Cloud Data

Satyam R. Thigale, Parth M. Kulkarni, Pradnyesh P. Shinde, Mohit H. Tathed.

Guide: Mr.R. P. Kumawat, Assistant Professor

Department of Cloud Computing and Big Data , Padmashri dr.vithhalrao vikhe patil inst. of tech. and engi. polytechnic loni,rahata

Abstract : Cloud computing is often regarded as a great innovation due to the increasing demand for endless storage space and reliable retrieval services. Several works have been developed on ranked multi-keyword search over cloud data with numerous data owners concerned about privacy. However, most of these methods can be broken by a clever attacker using a combination of a keyword guess and an equivalency test. Also, when searching across data from many data owners, they can't reliably return the top k results to the users. The unreliability of search results and the potential for the exposure of critical keyword information are two obvious drawbacks. To address this issue, we first propose a novel efficient ranked multi-keyword retrieval scheme with keyword privacy for multiple data owners, allowing the cloud server to perform multikeyword search over the cloud data and then return the top-k ranked search results to data users without leaking any keyword and trapdoor information. Furthermore, we demonstrate through stringent security analysis that our method is safe from both internal and external threats. Finally, the results of the performance evaluation show that our scheme outperforms other ranked multi-keyword search systems.

Keywords: cloud computing, multi-keyword retrieval, multiple data owner model, privacy sensitive, ranked.

INTRODUCTION

In this day and age of cloud computing, by moving locally installed apps and service provisioning to the cloud. Concerns about data security and privacy are the most significant barriers to the broad adoption of cloud computing. This is due to the fact that data owners are unable to fully complete real-time monitoring of the data that they have outsourced. The sensitive data should be uploaded to the server side in ciphertext form, as this is the way that is most usually used to safely preserve data privacy. On the other hand, data encryption makes it impossible for data consumers to obtain encrypted data from a server in a timely and effective manner. A simplistic strategy, such as downloading and decrypting all of the encrypted data locally in order to search for the desired information, would significantly reduce and degrade the utility of the cloud. As a consequence of this, the question of how to search for encrypted target data in a quick and effective manner at a low cost has become an important one that needs to be resolved as soon as possible. Any user in the system has the ability to transfer encrypted data to a cloud server so that it may be retrieved by receivers. This feature, known as searchable encryption (SE), was developed with the goal of making the trade-off between data encryption and data use more manageable. When users delegate the corresponding keyword trapdoor, cloud servers are able to search encrypted data for specific keywords if the keyword trapdoor corresponds with keyword indexes. This allows users to search encrypted data that contains certain terms. After that, the Search Engine is further enhanced to allow single, conjunctive, semantic, and Boolean keyword searches. Nevertheless, neither the top-k search query nor the multi-keyword search is supported by any of these techniques. ranked multi-keyword search strategies are a proposition made by for the purpose of enhancing the functionality of SE. whereas all of these programs take place in an environment with a single proprietor. A privacy-preserving ranked multi-keyword search system is providing an efficient ranked multi-keyword search scheme for multiple data owners as a means of achieving ranked multi-keyword search in multiple data owner scenarios. This is necessary in order to achieve ranked multi-keyword search. Nevertheless, we will be addressing three problems that can be found in both of these. To begin, all of the top-k search results that were brought back from the cloud server were from the same data owner. In actual use, the results of the top k searches come primarily from separate data owners. Second, by initiating keyword guessing attacks on both ciphertext and trapdoor, the privacy-sensitive keyword information is easily subject to eavesdropping by an honest but interested cloud server as well as other hostile adversaries. This leaves the information exposed to potential harm. Third, the cloud server is capable of executing equivalency test attacks in order to detect whether or not any two delegated trapdoors include the same group of keywords. This paper will be published in a forthcoming edition of the journal, as it has been accepted for publication. With the exception of the pagination, the content is finalized in its current form. Anyone who is in possession of the receiver's public key has the ability to construct the target cipher text or trapdoor for any arbitrary keyword on their own, which is one of the two types of assaults that can be launched against a cipher text or trapdoor. In this paper, we suggest a novel ranked multi-keyword retrieval that offers privacy protection for multiple data owners. This is done in order to address the flaws that have been discussed previously. The linear splitting strategy is utilized in order to conceal the keyword behind some degree of randomization so as to withstand the keyword guessing attacks that are made by the cloud server. As a direct consequence of this, the cloud server is unable to extract any privacy-sensitive keyword information from either the ciphertext or the trapdoor. In addition, this method of splitting can be utilized to prevent the cloud server from deciding whether or not the two trapdoors include the same group of keywords. A modified keyword balanced binary tree, also known as a KBB-tree, must first be developed before a cloud server can be made capable of performing ranked multi-keyword searches over encrypted data belonging to a variety of data owners and returning the top-k most relevant search results to data users. The "Depth-First-Search" (DFS) technique is then utilized in order to locate the top k most relevant multi-keyword ciphertexts for each data owner. The "Heap-Sort" (HS) algorithm is used by the cloud server to search for and return the final top-k search results to data users. This is done after the cloud server has obtained the k2 relevant search results that belong to the k data owners.

Problem Definition

To enhance the process of multi key word search the proposed model uses the AWS bucket and AWS RDS to search the query using k-top result functionality efficiently.

LITERATURE SURVEY

GUILAN CHEN et al [1], explained A block chain-based multi-keyword corticated less searchable public key authenticated encryption technique is proposed. We employ corticated less cryptosystem to encrypt keywords, avoiding corticated administration and key escrow in traditional and identity-based cryptosystems. We offer multi-keyword search to exactly find encrypted _les and retrieve them. We upload the true encrypted _les to the cloud server and place the encrypted indexes in block chain for anti-tampering, integrity, and traceability. Users can get accurate search results without third-party interference thanks to block chain anti-tampering. We also track monetary rewards with smart contracts to facilitate fair transactions between data owners and users without a third party.

YUANBO CUI et al [2], narrates a secure search service called attribute-based multiple keyword search (ABMKS), which extends searchable encryption. In current ABMKS methods, encrypted keyword index generation computations are time-consuming modular exponentiation, and the number is linearly rising with m. Here m is the number of keywords in _le. This work proposes an ABMKS with only multiplication operations in encrypted keyword index construction to reduce computing costs. Thus, encrypted keyword index generation computation is more efficient than existing approaches.

Debasis Das et al [3], focused on a semantic multi-keyword ranked search scheme for document retrieval on cloud data that is encrypted. The proposed scheme returns not only the documents containing terms that match with our query terms but also some more documents that contain terms which are semantically similar to the query keywords. Our experimental result shows that our technique is more precise than TF-IDF/VSM (Vector Space Model) models that focus only keyword matching while simultaneously achieving faster retrieval times than other tree based TF-IDF/VSM approach as our algorithm runs on reduced dimensions

Qin Liu et al [4], propose a prime inner product encoding (PIPE) scheme, which makes use of the indecomposable property of prime numbers to provide efficient, highly accurate, and flexible multi-keyword fuzzy search. Our main idea is to encode either a query keyword or an index keyword into a vector filled with primes or reciprocals of primes, such that the result of vectors’ inner product is an integer only when two keywords are similar. Specifically, we first construct PIPE0 that is secure in the known ciphertext model. Unlike existing works that have difficulty supporting AND and OR semantics simultaneously, PIPE0 gives users the flexibility to specify different search semantics in their queries. Then, we construct PIPES that subtly adds random noises to a query vector to resist linear analyses. Both theoretical analyses and experiment results demonstrate the effectiveness of our scheme

Yinbin Miao et al [5], propose Basic TMS approach for cloud-based group data sharing uses Shamir's secret sharing technique to achieve threshold multi-keyword search, threshold decryption, and short record cipher text size. Enhanced TMS adds threshold result verification and traceability to this basic TMS. Additionally, the updated TMS supports public result verification and dynamic operations with the public verifier and improved hash tables. Our formal security study shows that basic and improved TMS are semi-adaptively secure and can withstand Chosen-Keyword Attack. Our theoretical analysis and empirical experiments show both strategies' promise.

Lei Shang et al [6], describes a keyword query intention of mixed probability model, the model can be recommended result of query words together with the query intention fusion, and thus to minimize noise on the result of recommended keywords. This article also adds tags to key words, as the query keywords supplement. This method can not only make up for the noise interference of multiple keywords, but also calculate the similarity between the tag and the tag that has been marked on the web page, so as to judge the relevance between the user's query intention and the web page content more accurately. Finally, the results that are more in line with the user query interest are ranked in the front position. Experimental results show that the proposed method is more effective.

UA DAI et al [7], explained MRSE-HC is a privacy-preserving multi-keyword ranked search strategy for encrypted hybrid cloud data. A bisecting k-means clustering-based keyword partition technique balances document keyword dictionaries. Keyword partition-based bit vectors are used for documents and queries as search indexes. The public cloud uses the trapdoor to determine the candidate results after the private cloud filters out candidate documents using keyword partition-based bit vectors. The EMRSE-HC augmentation scheme adds a complete binary pruning tree to MRSE-HC to improve search performance. MRSE-HC and EMRSE-HC are privacy-preserving multi-keyword ranked search methods for hybrid clouds that beat FMRS in search efficiency.

Jianfei Sun et al [8], propose a novel efficient ranked multi-keyword retrieval scheme with keyword privacy for multiple data owners, which empowers the cloud server to perform multi keyword search over the cloud data and then, return the top-k ranked search results to data users without leaking any keyword and trapdoor information. Additionally, we show through rigorous security analysis that our scheme is secure against the attacks launched by inside attackers and outside attackers. Finally, the performance evaluation indicates that our scheme has more satisfactory features than the existing ranked multi-keyword search schemes.

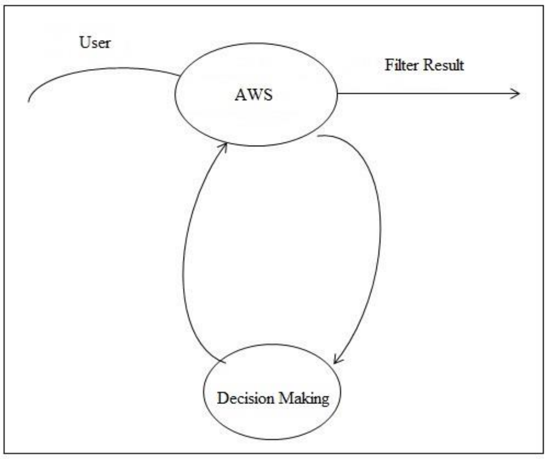
Peiming Xu et al [9], Describe a practical PEKS scheme named as public-key multi-keyword searchable encryption with hidden structures (PMSEHS). It could enable e-mail receivers to do the multi-keyword and Boolean search in the large encrypted email database as fast as possible, without revealing more information to the cloud server. We also give comparative experiments, which demonstrate that our scheme has a higher efficiency in multi-keyword search for encrypted emails.

Snehal N et al [10], develop an efficient data group sharing and multi-keyword ranked search method for encrypted cloud data collection in this research work. The developed system is implemented using the El-Gamal cryptography algorithm to provide security through effective key generation techniques and encryption strategy. Here, a multi-owner data setting is used instead of a centralized data owner setting; each member of the system in one particular group gets equal rights for both searching and sharing functionality and this may increase system usability. By taking into consideration lots of data in the cloud, the vector space model and TF-IDF model are utilized and according to the cosine similarity score, the method generates a ranked multi-keyword search result to deliver effective query result from numerous data and enhance secrecy in the situation of numerous data owners. In this system searching efficiency is improved by developing an index-based search.

Proposed Methodology with relevant Diagrams and Figures:

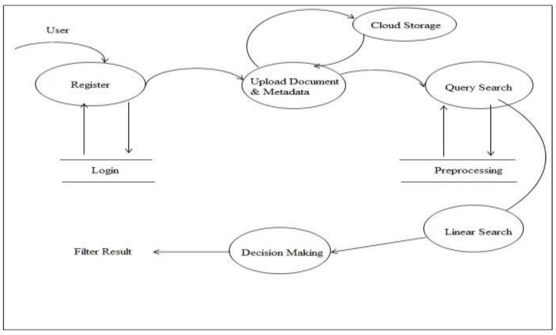
Data Flow Diagrams

DFD level 0



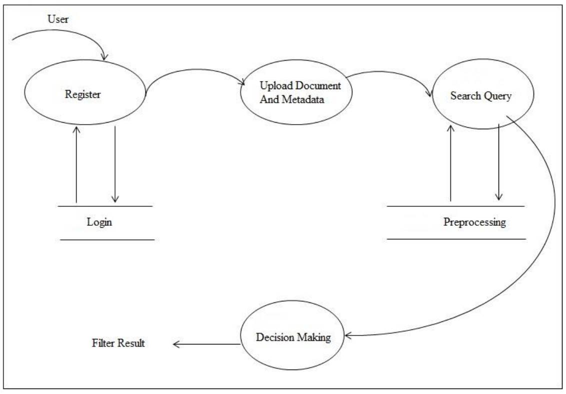
The DFD 0 diagram for the data flow diagrams describes the flow of the approach. The DFD diagram provides the simplest flow where user provide data AWS Cloud is preprocessed stopword, stemming Tokenization implemented in Multi-Keyword Search Cloud.

DFD level 1



The DFD 1 diagram provides even more details where in the user provides the Metadata which is effectively preprocessed. The Decision Making and provided to the AWS cloud which are used to store the uploaded metadata and Decision Making to achieve the Filtered Result

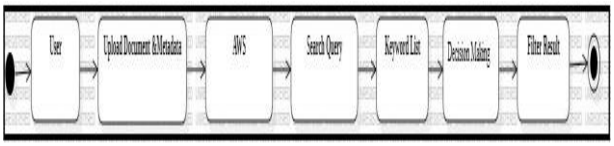
DFD level 2



The DFD 2 diagram is the most detailed where in the user provides the input metadata which is extracted and preprocessed through linear search and then the provided to the AWS cloud processing. The Decision Making is then implemented to achieve the Filtered result.

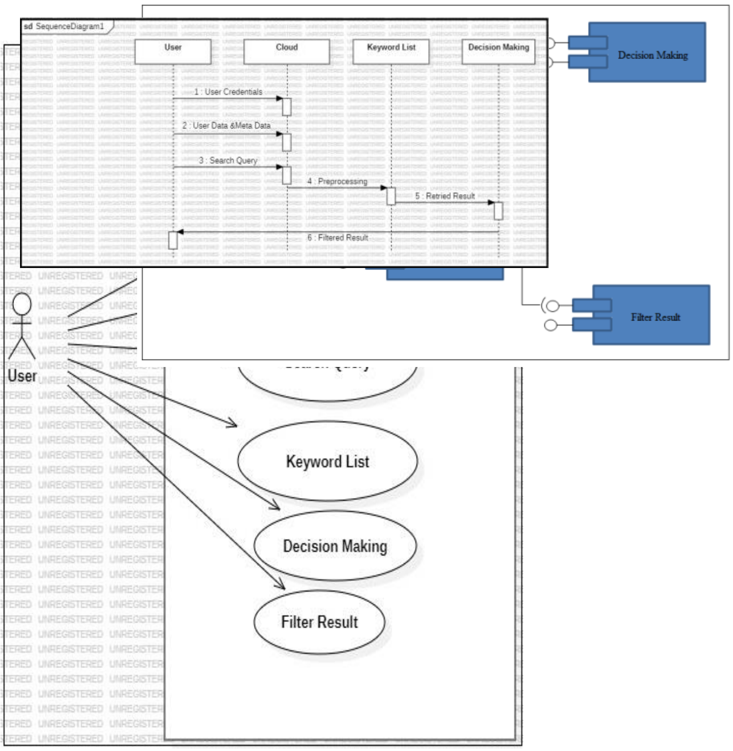
Activity Diagram:

The activity diagram lists the various activities that are performed in the proposed methodology, the start state is User and the upload document metadata, provide the AWS cloud for process is evaluated.



Usecase Diagram :

The Use case Diagram depicts the various use cases that are performed by the user in the proposed model. The use cases feeding the uploaded document, and used the AWS cloud for storing Metadata In here access the user all cases view the filtered result.



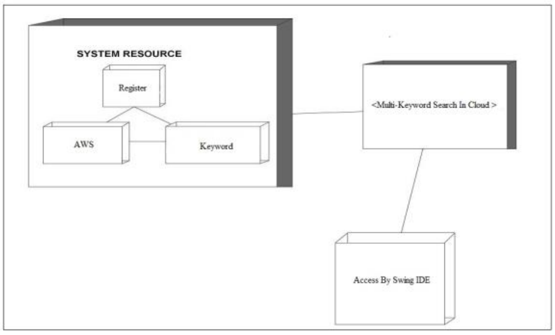
Sequence Diagram:

The sequence role diagram provides a sequence of the approaches as well as the various roles performed in the intermediate. In this approach the input document which is preprocessed stop-word stemming & Tokenization .In here provide the AWS cloud . The Decision Making is then implemented to achieve to get the filtered results.

Component Diagram

The component diagram illustrates the important components in the proposed system. In our approach the important components consist of the input document and metadata which is interlinked with AWS cloud, these two modules are further linked to the Decision Making and the Filtered Result module.

Deployment Diagram



The deployment diagram illustrates the important resources that are utilized for the deployment purposes. In our approach the system resources consist of metadata, Input document and the IDE along with the filter result s and the access to the system using the Swings GUI.

Minimum Hardware Specification:

CPU: Intel Core i5

RAM: DDR 8 GB

HDD: 500 GB

Software Specification:

Coding Language: Java

Development Kit: JDK 8.2

Front End: Swing Framework

Development IDE: NetBeans 8.2

Conclusion

The research presented in this paper addresses the critical need for efficient multi-keyword search in cloud data while ensuring privacy and security for multiple data owners. Cloud computing has revolutionized data storage and retrieval, but concerns about data privacy and security remain significant barriers to its widespread adoption. Our work contributes to overcoming these challenges by proposing a novel efficient ranked multi-keyword retrieval scheme with keyword privacy for multiple data owners.

The introduction of cloud computing has led to a paradigm shift in how data is stored and accessed. However, this shift has also brought forth new challenges, particularly regarding data security and privacy. Data owners are often unable to monitor their outsourced data in real-time, leading to concerns about unauthorized access or leakage of sensitive information. Traditional encryption methods, while effective in preserving data privacy, can hinder efficient data retrieval and utilization.

To address these challenges, our research focuses on developing a ranked multi-keyword search system that allows cloud servers to perform efficient searches over encrypted data while preserving keyword and trapdoor privacy. Our approach employs a linear splitting strategy to conceal keywords and resist keyword guessing attacks by cloud servers. Additionally, we utilize a modified keyword balanced binary tree (KBB-tree) to enable ranked multi-keyword searches across data from multiple owners.

The proposed methodology leverages AWS cloud services, including AWS bucket and AWS RDS, to enhance the efficiency of multi-keyword search queries and ensure reliable top-k search results. By utilizing the power of cloud computing and secure encryption techniques, our system provides data owners with a robust solution for searching and retrieving encrypted data without compromising privacy.

The literature survey highlights the existing research in the field of multi-keyword search over encrypted data, including blockchain-based approaches, attribute-based search methods, semantic search schemes, and privacy-preserving ranked search strategies. Our work builds upon these advancements and addresses key limitations to offer a comprehensive and efficient solution.

The proposed methodology is supported by relevant diagrams and figures, including Data Flow Diagrams (DFD), Activity Diagrams, Use Case Diagrams, Sequence Diagrams, Component Diagrams, and Deployment Diagrams. These diagrams illustrate the workflow, components, interactions, and deployment architecture of our system, providing a clear understanding of its functionality and design.

In conclusion, our research presents a robust and efficient solution for ranked multi-keyword search over encrypted cloud data, addressing critical privacy and security concerns for multiple data owners. The proposed methodology, supported by rigorous security analysis and performance evaluation, demonstrates the feasibility and effectiveness of our approach in real-world applications.