# Survey on Multi-keyword Ranked Search Scheme over Encrypted Cloud Data

Student Name: Yogesh Singh, Suraj Sonawane, Pratik Bansode, Swapnil Wankhede

Department: Department of Computer Engineering

College Name: Sinhgad Academy of Engineering, Pune, India

Guide Name: Kanchan Jadhav

**Abstract**

Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF × IDF model are combined in the index construction and query generation. we investigate the Multi-keyword top-k search problem for big data encryption against privacy breaches, and attempt to identify an efficient and secure solution to this problem. Specifically, for the privacy concern of query data, we construct a special tree-based index structure and design a random traversal algorithm, which makes even the same query to produce different visiting paths on the index, and can also maintain the accuracy of queries unchanged under stronger privacy.

**Keywords:** Cloud computing, privacy preserving, data encryption, multi-keyword top-k search.

## Introduction

Cloud computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-as you-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data applications have outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks

Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plain text into cypher-text, which is a non-readable form to unauthorized parties. A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plain text data no longer work well over encrypted data. Existing techniques are keyword-based information retrieval, which are widely used on the plain-text data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. In order to address the above problem, researchers have designed some general-purpose solutions with fully-homomorphic encryption or oblivious RAMs. However, these methods are not practical due to their high computational overhead for both the cloud sever and user. On the contrary, more practical special-purpose solutions, such as searchable encryption (SE) schemes have made specific contributions in terms of efficiency, functionality and security. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain.

**Motivation**

- Cloud security is important for both business and personal users. Everyone wants to know that their information is safe and secure and businesses have legal obligations to keep client data secure, with certain sectors having more stringent rules about data storage. To prevent unauthorized access to our data we need to provide some security mechanism to our data. Now days the Third-party cloud service providers are increasing very fast rate uploading or using their services may lead to misuse of our data (e.g. Balance sheet, Employee details).
- To provide security to such important documents and data is our motivation behind this project.

**Related Work**

In the [1] work user suggested "Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. The area of searchable encryption has been identified by DARPA as one of the technical advances that can be used to balance the need for both privacy and national security in information aggregation systems. In the [2] work the author presented "A related issue deals with privacy of database data. There are two different scenarios: public databases and private databases, and the solutions for each are different. Private databases: In this setting a user wishes to upload its private data to a remote database and wishes to keep the data private from the remote database administrator. Later, the user must be able to retrieve from the remote database all records that contain a particular keyword.

In the [3] work author states, In cloud storage systems, data files can be stored on different platforms and geographic locations, and the management and usage for data are very convenient. This brings numerous advantages for commercial applications and prompts more and more organizations to migrate their valuable data from local storage systems to cloud storage systems. As a result, many security equipment to protect dispersive local servers are saved and the costs for organizations are greatly reduced. Meanwhile, since valuable data are stored on remote cloud servers, access control should be achieved by network and cloud servers become high value attacking targets.

In the [4] work author states the use of crypto-graphic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. These techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plain-text when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the users authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the serve.

In the [8] work author states Keyword indexes let us search in constant time for documents containing specified keywords. Unfortunately, standard index constructions such as those using hash tables are unsuitable for indexing encrypted (and presumably sensitive) documents because they leak information about the document contents (and hence break semantic security). Informally, a secure index allows users with a "trapdoor" for a word x to test the index only for x; The index reveals no information about its contents without valid trapdoors, and trapdoors can only be generated with a secret key. Data structures with such privacy guarantees can be used to safely index the contents of semantically secure ciphertexts. We note that secure indexes do not hide information such as document size that can be obtained by simply examining the encrypted documents

**Problem Statement**

A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-

keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability. We define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data.

## Proposed Method

This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the widely-used "term frequency (TF) inverse document frequency (IDF)" model are combined in the index construction and query generation to provide multi-keyword ranked search.

To resist different attacks in different threat models, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known cipher text model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.

## Architecture

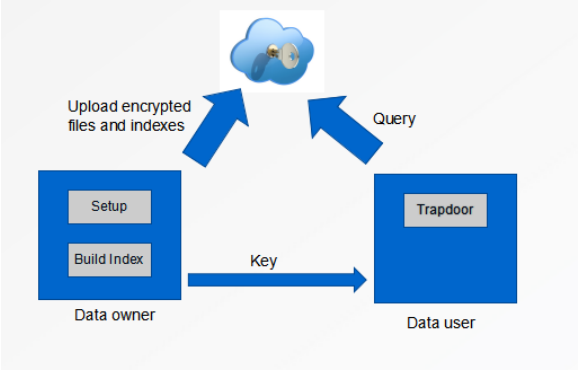


**Fig.2 System Architecture**

## Expected Result

In this system we are going to take a documents file as a input from our module i.e. data owner and he will Add keywords to so that the searching over encrypted data possible , then while sharing the details to the data users he have to process the requests for document access ,means the individual user cannot access the data unless and until Data owner approves it. And the cloud service provider will never come to know the content and keywords of the Documents that is the main motive we are going to achieve from this Project.

## Conclusion

We focus on improving the efficiency and the security of Multi-keyword top-k similarity search over encrypted data. Then, in order to improve the search efficiency, we design the group Multi-keyword top-k search scheme, which divides the dictionary into multiple groups and only needs to store the top-k documents of each word group when building index.

## References

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient con_x0002_structions," in Proceedings of the 13th ACM Conference on Computer and Communications Security.
.

[2] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology_x0002_Eurocrypt 2004.

[3] Z. Ying, H. Li, J. Ma, J. Zhang, and J. Cui, "Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating".

[4]D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted

data," in Security and Privacy, 2000. SP 2000. Proceedings.
.
[5] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003
.

[6]B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud, in INFOCOM, 2014 Proceedings IEEE, 2014.
.

[7] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data," in Distributed Computing Systems Workshops.

[8] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in INFOCOM, 2012.

[9] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data,IEEE Transactions on Parallel and Distributed Systems.
.

[10] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012 IEEE 28th International Conference on, 2012, pp. 1156–1167