International Journal of Research In Science & Engineering e-ISSN: 2394-8299
Volume: 3 Issue: 2 March-April 2017 p-ISSN: 2394-8280

# A DYNAMIC MULTI-KEYWORD RANKED SEARCH SCHEME OVER ENCRYPTED CLOUD DATA USING ADL

**Mr. Anilkumar H.[1], Mr.Nareshkumar R.M[2], Mr.Ravikumar Bajaj[3], Mr. Jitendrakumar Jha [4]**
[1, 2]*Assistant Professor, Dept. of Computer Engineering, DYPIEMR, Akurdi*
[1]*anil.hulsure@email.com* [2]*nareshkumarmustary@gmail.com*
[3, 4]*Student, Dept. of Computer Engineering, DYPIEMR, Akurdi*
[3] *ravi.bajaj00007@yahoo.in* [4]*indrakantjha25@email.com*

**ABSTRACT**

***Nowadays, more and more people are motivated to outsource their local data to public cloud servers for great convenience and reduced costs in data management. But in consideration of privacy issues, sensitive data should be encrypted before outsourcing, which obsoletes traditional data utilization like keyword-based document retrieval. In this paper, we present a secure and efficient Information Retrieval for Ranked Query Using Aggregation and Distribution Layer (ADL), which additionally supports dynamic update operations like deletion and insertion of documents. Specifically, we construct an index tree based on vector space model to provide multi-keyword search, which meanwhile supports flexible update operations. Besides, cosine similarity measure is utilized to support accurate ranking for search result. Moreover, to protect the search privacy, we propose a secure scheme to meet various privacy requirements in the known cipher text threat model.***

**Keywords*: ADL, Cloud, Multi keyword Ranked Search, Privacy Preserving.***

---

## 1. INTRODUCTION

Cloud computing as an emerging technology is expected to reshape information technology processes in the near future. Due to the overwhelming merits of cloud computing, e.g., cost-effectiveness, flexibility and scalability, more and more organizations choose to outsource their data for sharing in the cloud. As a typical cloud application, an organization subscribes the cloud services and authorizes its staff to share files in the cloud. Each file is described by a set of keywords, and the staff, as authorized users, can retrieve files of their interests by querying the cloud with certain keywords [1]. In such an environment, how to protect *user privacy* from the cloud, which is a third party outside the security boundary of the organization, becomes a key problem. The objective of this research paper is to protect on data duplication for removing duplicate copies of data and used on cloud to save storage space and increase bandwidth [14].

User privacy can be classified into *search privacy* and *access privacy* [2]. Search privacy means that the cloud knows nothing about what the user is searching for, and access privacy means that the cloud knows nothing about which files are returned to the user. When the files are stored in the clear forms, a naıve solution to protect user privacy is for the user to request *all* of the files from the cloud; this way, the cloud cannot know which files the user is really interested in. While this does provide the necessary privacy, the communication cost is high.

Private searching was proposed by Ostrovsky et al. [3,6] (referred to as the Ostrovsky scheme in this paper), which allows a user to retrieve files of interest from an untrusted server

without leaking any information. However, the Ostrovsky scheme has a high computational cost, since it requires the cloud to process the query (perform holomorphic encryption) on *every* file in a collection. Otherwise, the cloud will learn that certain files, without processing, are of no interest to the user. It will quickly become a performance bottleneck when the cloud needs to process thousands of queries over a collection of hundreds of thousands of files. We argue that subsequently proposed improvements, like [6], also have the same drawback. Commercial clouds follow a *pay-as-you-go* model, where the customer is billed for different operations such as bandwidth, CPU time, and so on. Solutions that incur excessive computation and communication costs are unacceptable to customers.

To make private searching applicable in a cloud environment, our previous work [7] designed a cooperate private searching protocol (COPS), where a proxy server, called the aggregation and distribution layer (ADL), is introduced between the users and the cloud as shown in system architecture (Fig. 1). The ADL deployed inside an organization has two main functionalities: aggregating user queries and distributing search results. Under the ADL, the computation cost incurred on the cloud can be largely reduced, since the cloud only needs to execute a combined query *once*, no matter how many users are executing queries. Furthermore, the communication cost incurred on the cloud will also be reduced, since files shared by the users need to be returned only once. Most importantly, by using a series of secure functions, COPS can protect user privacy from the ADL, the cloud, and other users.

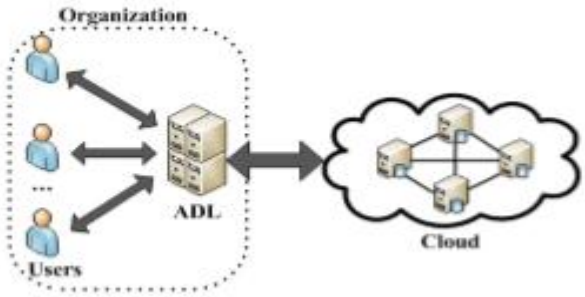



**Fig. 1. System Architecture**

### 1.1 Existing System

With the prevalence of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records, private videos and photos, company finance data, government documents, etc. To protect data privacy and combat unsolicited accesses, sensitive data has to be encrypted before outsourcing so as to provide end-to-end data confidentiality assurance in the cloud and beyond. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data.

***Disadvantages of Existing System***

1. For each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead.
2. Invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm.

### 1.2 Objective

- Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the cloud.

- It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors.

- When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant.

## 2. RELATED WORK

Searchable encryption schemes enable the clients to store the encrypted data to the cloud and execute keyword search over cipher text domain. Due to different cryptography primitives, searchable encryption schemes can be constructed using public key based cryptography or symmetric key based cryptography [2]. Song et al. proposed the first symmetric searchable encryption (SSE) scheme, and the search time of their scheme is linear to the size of the data collection. Goh proposed formal security definitions for SSE and designed a scheme based on Bloom filter. The search time of Goh's scheme is O(n), where n is the cardinality of the document collection. Curtmola et al. proposed two schemes (SSE-1 and SSE-2) which achieve the optimal search time. Their SSE-1 scheme is secure against chosen-keyword attacks (CKA1) and SSE-2 is secure against adaptive chosen-keyword attacks (CKA2).

These early works are single keyword boolean search schemes, which are very simple in terms of functionality. Afterward, 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, and multi-keyword ranked search [5],[4],[8] etc. Multi-keyword boolean search allows the users to input multiple query keywords to request suitable documents. Among these works, conjunctive keyword search schemes only return the documents that contain all of the query keywords. Disjunctive keyword search schemes [9] return all of the documents that contain a subset of the query keywords. Predicate search schemes [9],[10] are proposed to support both conjunctive and disjunctive search. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. Ranked search can enable quick search of the most relevant data. Sending back only the top-k most relevant documents can effectively decrease network traffic. Some early works [7] have realized the ranked search using order-preserving techniques, but they are designed only for

single keyword search. Cao et al. [7],[10] realized the first privacy-preserving multi-keyword ranked search scheme, in which documents and queries are represented as vectors of dictionary size. With the "coordinate matching", the documents are ranked according to the number of matched query keywords. However, Cao et al.'s scheme does not consider the importance of the different keywords, and thus is not accurate enough. In addition, the search efficiency of the scheme is linear with the cardinality of document collection.

Sun et al. [11] and Zheng et al. [12] proposed secure attribute-based keyword search schemes in the challenging scenario where multiple owners are involved. However, applying CPABE in the cloud system would introduce problems for data user revocation, i.e., the cloud has to update the large amount of data stored on it for a data user revocation [13]. Additionally, they do not support privacy preserving ranked multi-keyword search. Our paper differs from previous studies regarding the emphasis of multiple data owners in the system model. This paper seeks a solution scheme to maximally relax the requirements for data owners and users, so that the scheme could be suitable for a large number of cloud computing users.

## 3. PROPOSED SYSTEM

This paper proposes a secure Efficient Information Retrieval for Ranked Query Using Aggregation and Distribution Layer (ADL), which supports multi keyword ranked search and dynamic operation on the document collection. In order to obtain high search efficiency, we construct a tree-based index structure and propose a "Greedy Depth-first Search" algorithm based on this index tree. Due to the special structure of our tree-based index, the proposed search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors.

***Advantages of Proposed System***

1. We provide not only multi-keyword query and accurate result ranking, but also dynamic update on document collections.
2. Improved searching efficiency with privacy preserving.

## 4. MODULES DESCSRIPTION

- **Data Owner:**

The data owner is responsible for the update operation of his documents stored in the cloud server. While updating, the data owner generates the update information locally and sends it to the server.

- **Data User:**

Data users are authorized ones to access the documents of data owner. He fetches encrypted documents from cloud server, and then he can decrypt the documents with the shared secret key.

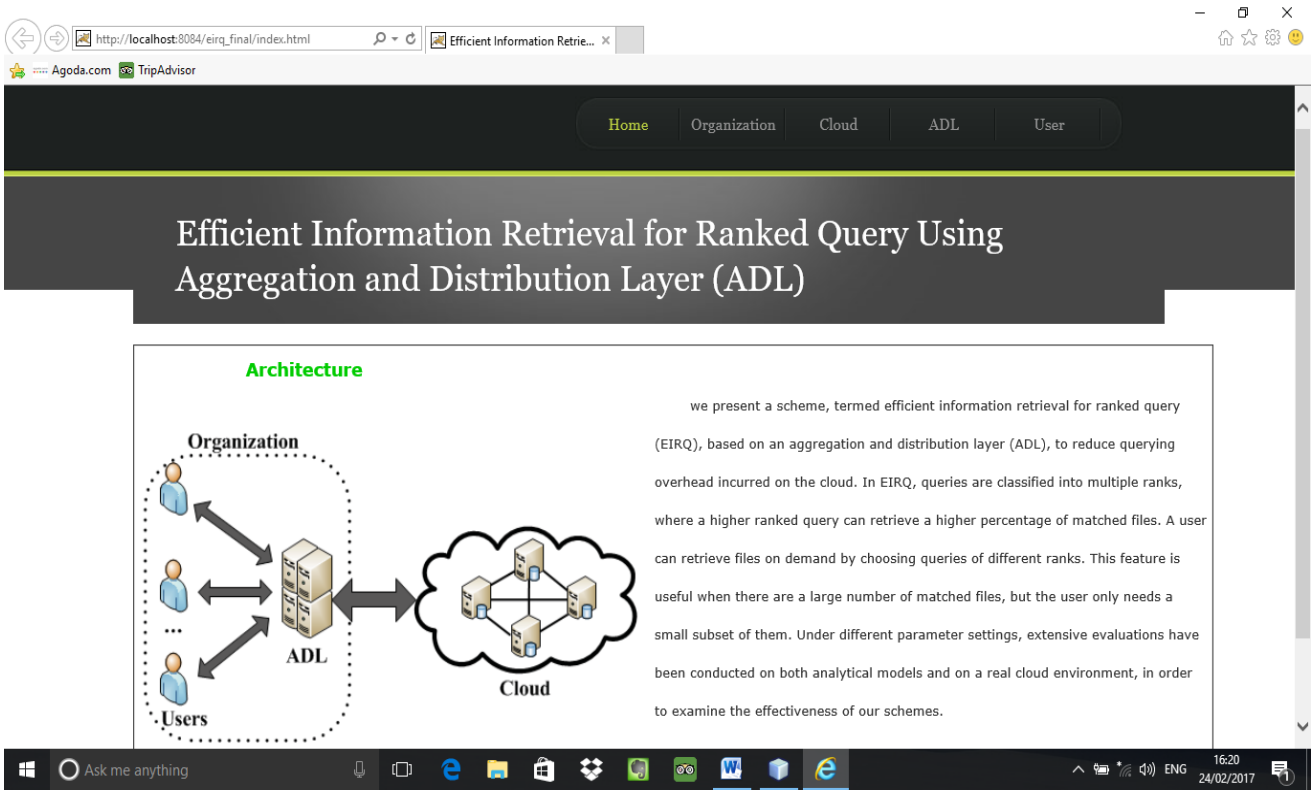
- **Semi-Trusted Cloud Server:**

Cloud server stores the encrypted document collection and the encrypted searchable tree index for data owner.

## 5. RESULTS

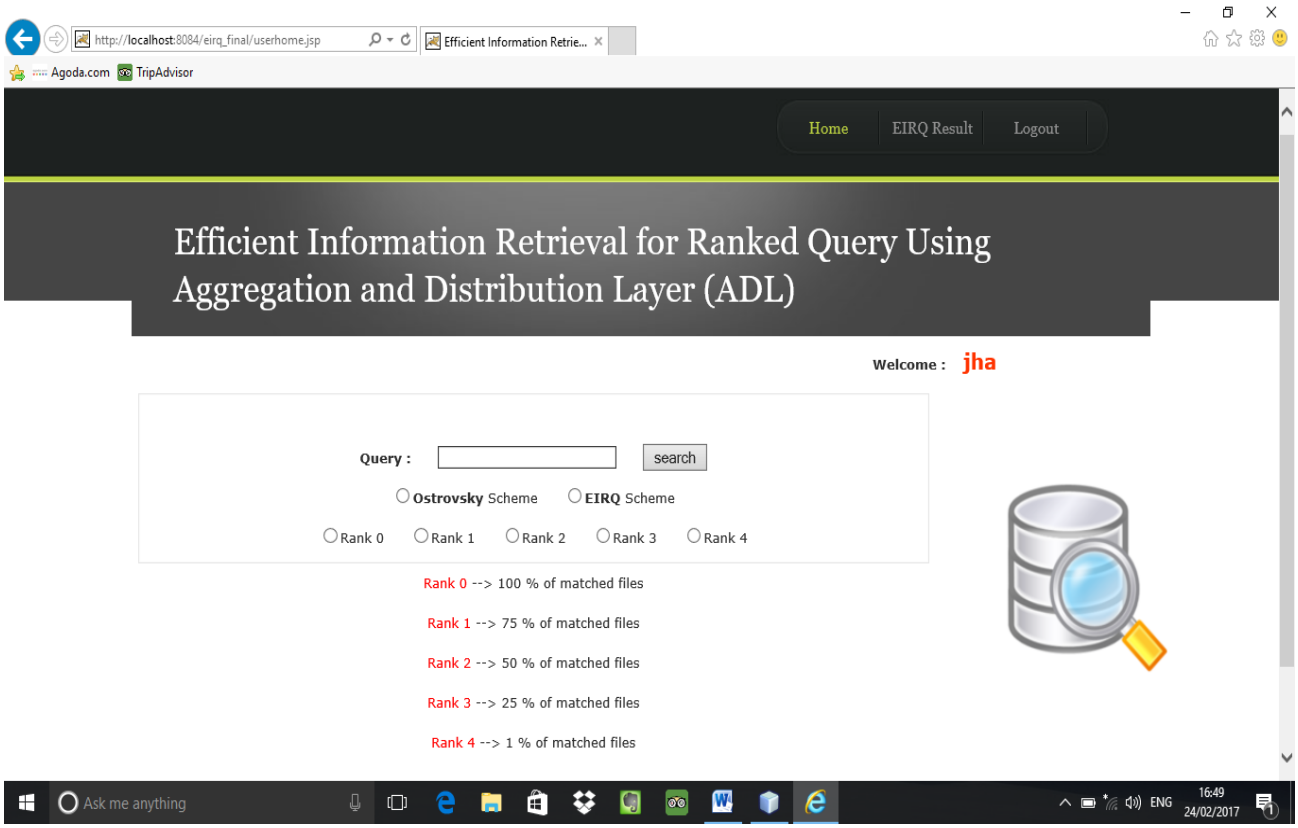
### 5.1 Home Page

In this page, it describes different modules which are going to perform.




### 5.2 EIRQ Page

In this page, we can prove that EIRQ is more efficient in retrieval of information using aggregation & distribution layer as compared to ostrovsky.




## 6. CONCLUSION

We proposed three EIRQ schemes based on an ADL to provide differential query services while protecting user privacy. By using our schemes, a user can retrieve different percentages of matched files by specifying queries of different ranks. By further reducing the communication cost incurred on the cloud, the EIRQ schemes make the private searching technique more applicable to a cost-efficient cloud environment. However, in the EIRQ schemes, we simply determine the rank of each file by the highest rank of queries it matches. For our future work, we will try to design a flexible ranking mechanism for the EIRQ schemes.

## ACKNOWLEDGEMENT

It gives us an immense pleasure and satisfaction in submitting this paper. In this endeavour of preparing this paper many people gave a helping hand, we would thank them all. Specially, we heartily pay gratitude to our faculty Prof.Anilkumar Hulsure & Mr.Nareshkumar R.M who gave us this opportunity to work upon this topic.

## REFERENCES

[1]. K.Ren, C.Wang, Q.Wangetal.,"Security challenges for the public cloud" IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.
[2]. S. Kamara and K. Lauter, "Cryptographic cloud storage" in Financial Cryptography and Data Security. Springer, 2010, pp. 136– 149.

[3]. C. Gentry, "A fully homomorphic encryption scheme" Ph.D. dissertation, Stanford University, 2009.
[4]. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology Eurocrypt 2004. Springer, 2004, pp. 506–522.
[5]. D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows peer queries," in Advances in Cryptology-CRYPTO 2007. Springer, 2007, pp. 50–67.
[6]. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data" in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44– 55.
[7]. Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proceedings of the Third international conference on Applied Cryptography and Network Security. Springer-Verlag, 2005, pp. 442–455.
[8]. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 79–88.
[9]. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in INFOCOM, 2010 Proceedings IEEE. IEEE, 2010, pp. 1–5.
[10]. B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi keyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, 2014.
[11]. W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud," in Proc. IEEE INFOCOM'14, Toronto, Canada, May 2014, pp. 226–234.
[12]. Q. Zheng, S. Xu, and G. Ateniese, "Vabks: Verifiable attributebased keyword search over outsourced encrypted data," in Proc. IEEE INFOCOM'14, Toronto, Canada, May 2014, pp. 522– 530.
[13]. J. Hur, "Improving security and efficiency in attribute-based data sharing, " IEEE TRANSACTIONSON KNOWLEDGE AND DATA ENGINEERING, vol. 25, no. 10, pp. 2271–2282, 2013.
[14]. Nareshkumar Mustary et. Al., "Smart Distributed Deduplication with secured Reliability Mechanism", in International Journal of Research In Science & Engineering, Volume: 1 Issue: 6, Dec-2015, pp. 115-120.