International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

# Private Data Retrieval with Competent Ranked Keyword Search Algorithm Over Cloud Data Center

**Jyotsna T. Kumbhar[1], Navnath D. Kale[2]**

[1]ME Student, Department of Computer Engineering, TSSM'S, P.V.P.I.T., Bavdhan,
Pune University, Pune, Maharashtra, India, 411021

[2]Assistant Professor, Department of Computer Engineering, TSSM'S, P.V.P.I.T., Bavdhan,
Pune University, Pune, Maharashtra, India, 411021

**Abstract:** ***Cloud computing is an emerging trend in information technology. It provides different merits, but on other hand cost efficiency and privacy are two fundamental issues are addressed in cloud computing. To protect the privacy during secure searching is challenging task in this environment. There are different techniques are used to protect the data and user privacy during secure searching. But in existing system cost of communication get increase when number of users is increased in the system. Cloud provides pay as you go model so that high cost is unacceptable to the user. Proposed model helps to reduce the communication and computation cost during secure searching. Proposed system introduces flexible ranking mechanism, in that user gives keyword and rank of keyword to retrieve certain percentage of matched files on demand. This mechanism helps when there are large number of files are matched with query but user is interested in certain percentage of matched file.***

**Keywords:** Aggregate Distribute Layer (ADL), Cloud computing, Cooperative private searching protocol (COPS), Efficient Information Retrieval Query (EIRQ), Mask matrix.

## 1. Introduction

Cloud computing is a latest trend in information technology. It provides different merits like cost-effectiveness, scalability, flexibility, Security etc. so that Most of the organizations choose cloud computing to store and share their data. Data owner uploads file with different keywords. So that each stored file is represented by certain keywords; only authorized user can retrieve the file by querying cloud with the certain keywords. Cloud is operated by third party so it is necessary to provide the privacy to the user during searching and protect the data stored on cloud. Solution for that is , the data is encrypted before outsourced on the public cloud. In such a large collection of data, it becomes difficult for the user to query something out of the whole collection. There are different techniques are used for secure searching on cloud. But drawback of existing system is that when number of users is increased in system then costs of computation and communication are also get increased due to query overhead on cloud. By using Keyword searches users can actively utilize clouds to query a collection. Proposed model defines and solves the problem of secure flexible ranked keyword search over encrypted cloud data. Flexible Ranking mechanism helps to reduce the computation and communication cost, because it enables the user to retrieve more relevance result and avoid undifferentiated results. Proposed scheme uses query aggregation technique because when more than one user want to retrieve same file, the querying cost get reduced if the aggregator combines their queries to retrieve file once. In this scheme aggregator i.e. Aggregate Distribute Layer (ADL) located between the user and cloud is considered as trusted system, because it is located under the security boundary of organization. Users send query to the ADL instead of sending query directly to the cloud, then ADL aggregates queries and generate single query with the help of DISCOVER technique and sends it to the cloud. so it reduces bandwidth cost. ADL sends repeated query only once so that computation overhead on cloud get reduces. User sends keyword and rank of keyword to search certain percentage of top relevance matched file. Cloud sends result to ADL and ADL distributes results to different users.

## 2. Literature Review

There are different algorithm are proposed for the private searching in the cloud. EIRQ scheme introduced in [1] by Qin Liu et al. , allows user to retrieve certain percentage of matched file, when there are a large number of files are matched with the user query but user are interested in subset of matched file. The main drawback of this scheme is, it causes a heavy querying overhead incurred on the cloud.

In [2], R. Ostrovsky et al., introduced private searching on untrusted server, where the data is stored in the plaintext form, and Paillier cryptosystem [9] is used to encrypt the query. It is required to process the each query on every file in a collection so that it has a high computational cost. Previous work [3] by Q. Liu et al., designed a cooperate private searching protocol (COPS), where a proxy server is called the aggregation and distribution layer (ADL), is introduced between the users and the cloud to aggregate query and distribute results, so the computation cost incurred on the cloud can be largely reduced, but cloud sends all matched files even if user are interested in subset of matched file.

In [5], G. Danezis et al., presents an efficient decoding mechanism for private searching. The main drawback of the existing system is communication and computation costs are increased with users that are executing searches. Ranked searchable encryption technique enables users to retrieve the most matched files from the cloud in the case that both the query and data are in the encrypted form.

In [7], A. Boldyreva et al., introduced, Order Preserving Symmetric Encryption (OPSE) which supports single-keyword searches, in this files and queries are encrypted

using OPSE and utilizes keyword frequency to rank results.

In [9], P. Paillier introduces Paillier cryptosystem. it is a public key cryptosystem, it allows user to perform different operation like exponential and multiplication on ciphertext, and addition and multiplication to obtain plaintext. Most of the private searching techniques use Paillier cryptosystem because it provides homomorphic properties.

Work introduced in [8] by Ning Cao et al., the searching of Encrypted cloud data using Privacy-Preserving Multi-keyword Ranked Search (MRSE) method. In this co-ordinate matching technique is used to find the similarity between search query and data documents.

In [11] C. Wang et al., presents the problem and solution of secure ranked keyword search over encrypted cloud data. Relevance ranking mechanism is used to search more relevant results and avoid unnecessary results. This technique helps to protect the user privacy. Specifically, this work represents building of secure search index by using relevance score approach. It develops scheme to protect the sensitive score information using a one-to-many order-preserving mapping technique. All these algorithms address to the private searching in the cloud environment.

## 3.Motivation

Cloud security is an important problem for both industry and academia. The potential privacy leakages are the important security problems that may occur when outsourcing data to the cloud. Another thing is that, when number of users is increased computation cost get increased during secure searching, to make private search applicable in a cloud environment this work proposes a system that reduces the computational and communication costs while providing similar privacy protection as in the prior protocols.

## 4.Proposed Model

### 4.1 System Model

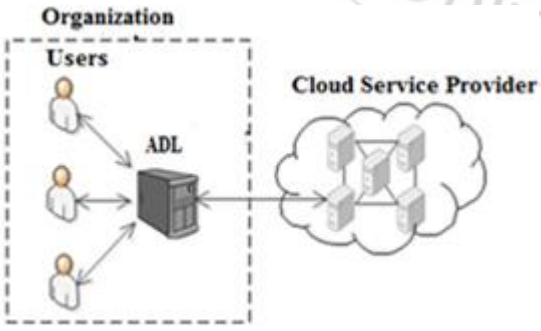


**Figure 1:** System Model

Fig. 1 shows the system model, there are three parties in the system

- User: User generates query and sends it to the ADL. Only authorized user can send query.
- ADL: Aggregate and distribute layer is a trusted middleware layer between the users and the cloud to aggregate the query and distribute the results to the users.
- Cloud Service Provider: CSP has large amount of resources and storage space which is used to manage clients data. It has capability to compute results for user query.

### 4.2 Algorithm

*Flexible Ranked Query algorithm:*

- Step 1: The user sends keywords and the rank of the Keyword to the ADL. Queries are classified into 0,1,.. r-1 ranks. here 0 is the highest rank and r is lowest rank.
- Step 2: ADL aggregates enough user queries by using DISCOVER search Technique, ADL Constructs a mask matrix to protect query and ranks. Before sending aggregate query to the cloud ADL encrypt query by the ElGamal Public Key Encryption Algorithm.
- Step 3:The cloud filters the files and generate result, after that it maps result with buffer, and returns a buffer that contains a certain percentage of matched files and keyword to the ADL. Query of Rank-i retrieves (1-i/r) percentage of matched files.
- Step 4: The ADL distributes search results to each user according with keywords and rank of keyword.

Fig. 2 shows the system architecture of proposed model

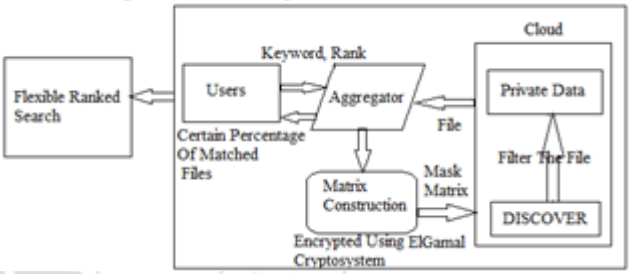


**Figure 2:** System Architecture

### 4.3 System Architecture

The drawback of EIRQ-simple is that it returns redundant files when there are files satisfying more than one ranked query .In the proposed System, instead of EIRQ scheme we are implementing DISCOVER borrows technique. It is successors query processing technique. The basic idea of DISCOVER search is in the database relations it creates a set of tuples for each subset of search terms. A candidate network is represented by a tree of tuple sets and edges represent relationships in the database schema. DISCOVER search uses breadth-first algorithm to generate candidate networks. DISCOVER creates a join expression for each candidate network, it executes the join expression to generate results, and then it ranks the results by the number of joins. DISCOVER technique is used to aggregate different queries. The score of each attribute (i.e., a document) in the tree of tuples is summed to obtain the total score. To improve scalability, DISCOVER-II creates only a single tuple set for each database relation and supports top-k query processing because users typically view only the highest ranked search results. For Privacy, the combined query sent to the cloud is encrypted under the ADLs public key with the ElGamal cryptosystem. The ElGamal Algorithm provides an alternative to the RSA for public key encryption. It has the advantage that, the same plaintext gives a different cipher text with near certainty each time it is encrypted. The query

is represented by matrix of encrypted 0s and 1s. The ElGamal cryptosystem is semantically secure, and the cipher text of every 1 or 0 is different from other 1s or 0s. So it preserves the search privacy. The cloud cannot know that what the user is searching for. Data owner uploads data on cloud for sharing, for security purpose data is encrypted before uploading on cloud using PBEWithMD5AndDES i.e. password based encryption with MD5 and Data Encryption Standard algorithm. In proposed system user uses keyword to retrieve files so dictionary attack can be possible in this case. So MD5 algorithm helps to prevent dictionary attack.

## 5.Mathematical Model

R -> Rank
Q -> Query
M -> Mask Matrix
K -> Keyword
B -> Buffer
S -> System
S={User, Q, R, M, ADL, Cloud}

- Input: Users send query to ADL Qi={Ki,Ri}
- Process: ADL aggregates different queries, Q={Q1,Q2..,Qn}ADL generates mask matrix Epk {M} which Encrypted with ADLs public key pk, it sends to Cloud. Cloud sends result to ADL B= {File, K}
- Output: ADL distributes results to users {File1, File2.,Filen}

## 6.Modules

- Cloud Service Provider

It stores all files uploaded by the authorized data owner, and allow accessing the data only to the authorized User. It accepts the query from ADL and sends the result according to the query send by ADL.

- Data Owner

The data owner is person who uploads the data. Data owner should login first to store their data on cloud. Only authorized owner can store their data on cloud and new data owner should register first.

- Cloud User

In this module, the cloud user should be authenticated so the user should register and login for cloud usage. The vendor which verifies that whether the user is authenticated or not, If he/she is unauthorized user then CSP cannot continue further processing. Once login successfully then he/she can obtain the basic information from cloud storage for data integration. Here the user can request the data to cloud using certain keywords.

- Query Search based on ranking

In this module, each user set the rank to their query and cloud searches the results on the basis of the rank. Cloud returns certain percentage of matched file to the ADL. The basic idea of this module is to protect the privacy and rank the user queries. In this module, Cloud sends the file along with the keywords.

- File Distribution

The aggregation and distribution layer (ADL), is an intermediate layer between the user and cloud. It aggregates the queries and sends combined query to the cloud. cloud sends result to the ADL, Finally the file distribution will be performed to deliver the requested files to the corresponding users.

- Performance Evaluation

Here we compare the performance of existing scheme with proposed flexible ranking mechanisms. It represents the performance analysis of the work in the form of a graph to show the variance occurred in between them.

## 7.Results and Discussion


**Figure 3:** Ranked Keyword Serching

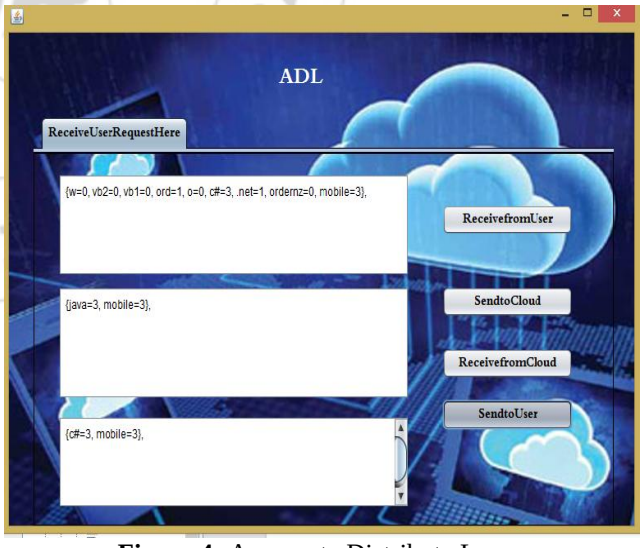


**Figure 4:** Aggregate Distribute Layer

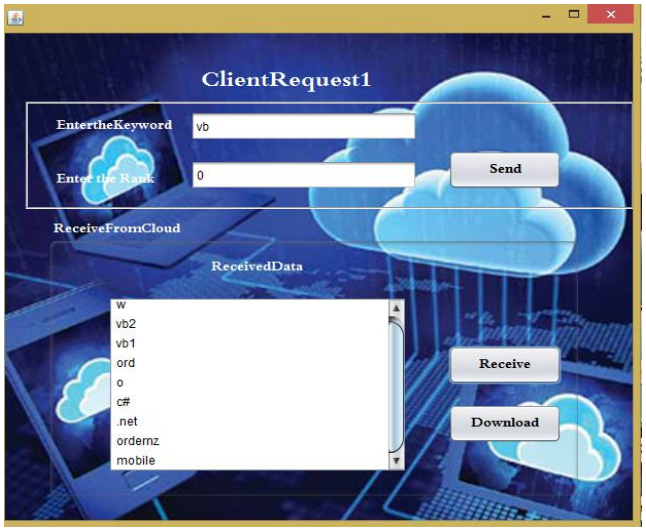


**Figure 5:** Client Request

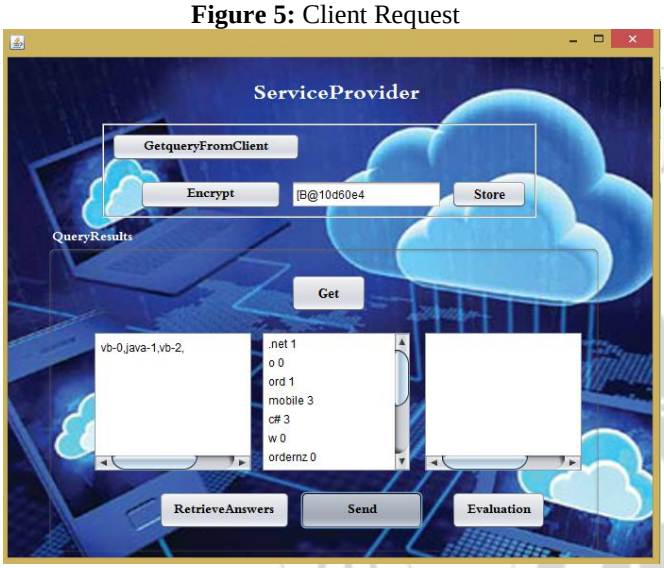


**Figure 6:** Cloud Service Provider

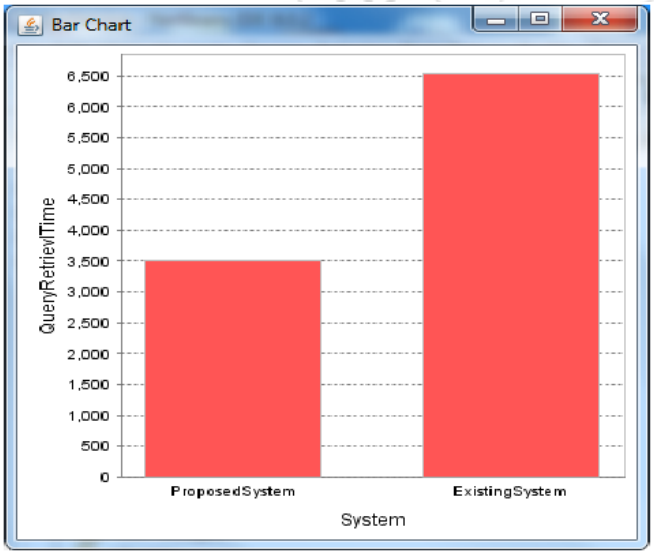


**Figure 7:** Performance Evaluation

## 8.Conclusion

The proposed flexible ranking model helps to improve the performance and security in cloud computing. In this scheme user can retrieve different percentages of matched files by specifying queries of different ranks. This technique is useful when there are a large number of matched files, but the users are interested in certain percentage of matched files. Aggregate Distribute Layer introduced between user and cloud ,aggregates the query and distributes the results to different user this helps to reduce the communication and computation cost during information retrieval in cloud computing environment. Proposed system provides security to user and data during secure searching. In future work we can use MD5 algorithm to provide more security to data. This technique helps to check the integrity of data stored on cloud.

## References

[1] Qin Liu, Chiu C. Tan, Jie Wu and Fellow ,Towards Differential Query Services in Cost-Efficient Clouds IEEE Transactions On Parallel and
[2] Distributed Systems, VOL. 25, NO. 6, JUNE 2014.
[3] Ostrovsky and W. Skeith III, Private searching on streaming data, in Proc. of ACM CRYPTO, 2005.
[4] Q. Liu, C. Tan, J. Wu, and G. Wang, Cooperative Private Searching in Clouds,J. Parallel Distrib. Comput. , vol. 72, no. 8, pp. 1019-1031, Aug.2012.
[5] Q. Liu, C. C. Tan, J. Wu, and G. Wang,Efficient information retrieval for ranked queries in cost-effective cloud environments, in Proc. of IEEE INFOCOM, 2012.
[6] G. Danezis and C. Diaz, Improving the decoding efficiency of private search,in IACR Eprint archive number 024, 2006.
[7] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, Secure ranked keyword search over encrypted cloud data, in Proc. of IEEE ICDCS, 2010
[8] Boldyreva, N. Chenette, Y. Lee, and A. Oneill,Order-preserving symmetric encryption, Advances in Cryptology-EUROCRYPT, 2009.
[9] Ning Cao, Cong Wang , Li, Ming , Kui Ren, Wenjing Lou, Privacy- Preserving Multi-keyword Ranked Search over Encrypted Cloud Data,
[10] INFOCOM, 2011 Proceedings IEEE April 2011.
[11] P. Paillier,Public-Key Cryptosystems Based on Composite Degree Residuosity Classes, in Proc. EUROCRYPT, 1999, pp. 223-238. V. Hristidis and Y. Papakonstantinou, DISCOVER: Keyword Search in Relational Databases,in Proceedings of the 29th International Conference on Very Large Data Bases, VLDB Endowment, August 2002, pp. 670681.
[12] Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou,Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Transactions On Parallel And Distributed Systems, Systems, VOL. 23, NO. 8.