# Ranking Based Search in the Encrypted Cloud Environment

Vignesh. P
M. Tech-Information Technology
SRM University
Chennai, Tamilnadu
*e-mail:vignesh.gars12@gmail.com*

Mr. P. Rajasekar
Assistant Professor, Department of Information Technology
SRM University
Chennai, Tamilnadu
*e-mail:rajasekar.p@ktr.srmuniv.ac.in*

***Abstract:*** Cloud computing is emerging as a promising technology for outsourcing of data and quality of data services. However, information which is sensitive when upload on cloud eventually cause privacy problems. Data encryption provides security of data to some level, but at the cost of compromised efficiency. This paper focus on addressing data privacy problems. For the first time, the privacy issue is formulated from the aspect of similarity relevance of data and scheme robustness. Privacy of data is not assured if Server-side ranking based on order-preserving encryption is maintained. For the assurance of data privacy, multi-keyword ranked search over encrypted data in cloud computing (MRSE) scheme is proposed which supports top-k multi keyword retrieval. In MRSE, vector space model and Homomorphic encryption were employed. The vector space model helps to provide accuracy sufficient search of data and the Homomorphic encryption enables users to involve in the encryption of data. The majority of computing work is done on the server side. As a result, leakage of information can be eliminated and data security is ensured.

***Keywords***- *encrypted cloud; vector model; ranked search; multi-keyword; MRSE;*

*****

## I. INTRODUCTION

Cloud computing is an emerging technology for data outsourcing and high-quality data services. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. There are many cloud platforms like Google Drive, Sky Drive, Amazon S3, Drop Box, and Microsoft Azure which provides storage services. Privacy problems may occur if sensitive data such as tax reports, health reports, bank statements, etc., upload on cloud.

There are several mechanisms like firewall have been used by cloud provider but these are not enough to protect data. Encrypting the data provides data security to certain level, but at the cost of compromised efficiency. Data user have to search for the data in cloud which results in irrelevant search of data. Downloading all the data and decrypting the data is not possible. It needs huge bandwidth. Searchable symmetric encryption (SSE) scheme allows retrieval of encrypted data over cloud. This project focusses on addressing data privacy problems using SSE. For the first time, this project formulate the privacy problems from the aspect of similarity relevance and scheme robustness. This technique is helpful to search encrypted data as documents and it allows the user to search securely through multi-keyword. Ranking scheme provides data which has more vector value which user searched for more times is helpful for searching relevant data that provides usage of less bandwidth and data efficiency. Generally, data privacy is untrusted when server-side ranking based on order-preserving encryption (OPE) is employed. To maintain data privacy, this project propose a multi-keyword ranked search over encrypted data in cloud computing (MRSE) scheme that supports top-k multi keyword retrieval. MRSE employs vector space model and Homomorphic encryption. The vector space model is used to provide vector values for data based on keyword which helps for sufficient search accuracy and the Homomorphic encryption enables data users to involve in the encrypting of the data when upload to or download from the cloud while the majority of computing work is done on the server side. As a result, leakage of the information can be eliminated and security of data is ensured. Thorough analysis of security and performance show that the high security and practical efficiency is guaranteed through the proposed scheme.

## II. EXISTING SYSTEM

Keyword-based retrieval is a common data service which is widely applied in plaintext scenario. Single-keyword search without ranking system provides searching data with single keyword. Boolean- keyword search without ranking provides true or false scenario but without ranking of data. Single-keyword search with ranking provides ranking of data with single keyword. Traditional SSE schemes provides data users to securely retrieve the cipher-text data from cloud, but these schemes support only Boolean keyword search algorithm. To meet the data retrieval efficiency, the large amount of documents demand the cloud server to perform relevance ranking. Undifferentiated result might be occurred when performing irrelevant search. Ranked based search can eliminate unnecessary network traffic elegantly by sending back only the most relevant data, which is highly desirable in the “pay-as-you use” policy.

*Disadvantages*

1. Undesirable security and privacy risks of data.
2. Only allowed for Single keyword Search

3. Top-k multi-keyword has been used but only allow for Boolean search.
4. Loss of Bandwidth and Less Security.
5. Downloading all the data and Decrypting is impractical.

## III. PROPOSED SYSTEM

This project introduce the concepts of relevance of similar data and scheme robustness which formulates the privacy problems by proposing a multi-keyword ranked search over encrypted data in cloud computing (MRSE). Homomorphic encryption and vector space model were employed for better encryption and efficient ranking. The computing work is done on the cloud while the user takes part in ranking module, which ensures top-k multi keyword retrieval over encrypted cloud data with high security and practical efficiency. MRSE scheme is proposed in this project which fulfills the secure multi keyword top-k retrieval over encrypted cloud environment.

*Advantages*

- The new scheme guarantees high data privacy.
- Provide heavy security for storage
- Lightweight cost of computation and communication.

## IV. MRSE FRAMEWORK

For easy understanding, data operations are not shown since the data owner could implement the traditional symmetric key cryptography easily to encrypt the data and then outsource data. With great focus on the query and index, the MRSE system has four algorithms [3] as follows

1. Setup (ℓ) Security parameter ℓ is taken as input, the data owner outputs a symmetric key which is noted as SK.

2. BuildIndex (F, SK): With dataset F, the data owner creates a searchable index I which is then encrypted by the symmetric key SK and then it is outsourced to the cloud server. After the construction of index, the document collection can be encrypted independently and outsourced.

3. Trapdoor (fW): With list of keywords t in fW as input, this algorithm generates a trapdoor TfW correspondingly.

4. Query(TfW, k, I): When the cloud server receives a request of query as (TfW, k), ranked search is performed on the index I with trapdoor TfW help, and finally returns the value FfW, which is the ranked id list of top-k documents which is sorted by their similarity with fW.

## V. MODULES DETAILS

1. Index Creation Module
2. Data Encryption Module
3. Vector Space Module
4. Top- k Rank Provide Module
5. MRSE-Query Process Module

### *1. Data Owner - Index Creation Module:*

The data owner who has a collection of n files $C = \{f1, f2, \dots \dots . fn\}$ to outsource onto the cloud server in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other users who are authorized. To meet this, the data owner wants to build a searchable index I from a collection of keywords extracted out of cloud and then outsources both the encrypted index I' file on to the cloud server.

### *2. Data Encryption Module:*

The encryption module guarantee the security and operability at the same time on cloud server. Homomorphic encryption provides certain types of computations to be carried out on the corresponding cipher text. Homomorphic encryption allows computation of cipher text without knowing anything about the plaintext to get the correct encrypted result. Employing the vector space model to top-k retrieval can reduce the original homomorphism in a full form to a simplified form that only supports operations on integers, which assures more efficiency than the full form does.

On basis of property of Homomorphism, the encryption technique can be described as four stages: KeyGen, Encrypt, Evaluate, and Decrypt.

### *3. Vector Space Module:*

The vector space model to identify the score on multi keyword search over cloud. The vector space model is an algebraic model for representing vector value for a single file. Each and every dimension of the vector corresponds to a different terms, i.e., if any term occurs in the file, the value in the vector is declared to nonzero, otherwise is zero. The vector space module supports multi term and non-binary presentation. It allows computing a continuous degree of similarity between file and queries, and then relevant files are ranked accordingly. A query is also represented as a vector while each dimension of the vector is assigned with 0 or 1 according to whether this term is queried. The score of file f on query is deduced by the inner product of the two vectors: When the scores are provided, files can be ordered by the rank and most files which are relevant can be found.

### *4. Top- k Rank Provide Module*

SSE schemes employ server-side ranking based on OPE to improve the efficiency of retrieval over encrypted cloud environment. However, ranking in server-side based on OPE does not secure the privacy of information which are sensitive. This is considered as uncompromised in the scenario of security-oriented third party cloud computing, i.e., security cannot be give up for better efficiency.

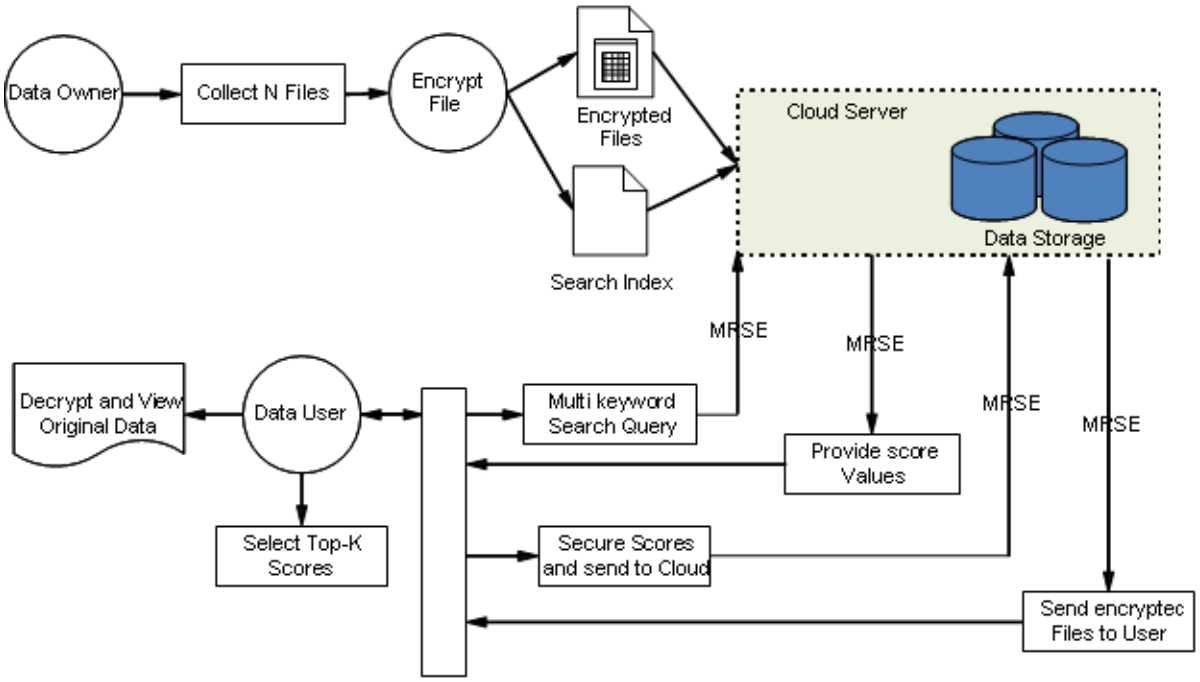
Traditional user-side techniques load high communication and heavy computational burden overhead on the data user side, which is due to the interaction between the user and the

server including searchable index return and calculation of score for ranking. Thus, the user-side ranking techniques are challenged by means of practical use. A more server-side scheme can be a better solution for privacy problems.

### 5. *MRSE- Query Process Module*

The cloud server receives a query consisting of multi keywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. The data user decrypts the scores and picks out the top-k highest scoring files identifiers to request to the cloud server which has encrypted cloud data. The retrieval of data takes a two-round communication between the data user and the cloud server.

## VI. DATA FLOW DIAGRAM




## VII. CONCLUSION

This project ensures to search data using multi keyword over encrypted cloud environment and it also provides ranking of the files which helps to download particular files by means of top-k retrieval.

In future, many enhancement can be done like admin option can be enable to maintain encrypted files over cloud, users can be authorized via OTP and so on.

## REFERENCES

[1] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[2] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" IEEE transactions on parallel and distributed systems, vol. 25, no. 1, January 2014.

[4] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS '10), 2010.

[5] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.

[6] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFO-COM, pp. 693-701, 2012.

[7] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[8] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467- 1479, Aug. 2012.