International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 3, March 2015
Available at http://internationaljournalofresearch.org

# Restoration of protected Top-k Multi-Keyword over Encrypted Cloud Data

**[1]Gitta. Bharath, [2]Dr. P. Venkateswarlu, [3]S. Sree Hari Raju**

[1] M.Tech (CSE), Department of Computer Science & Engineering Nagole Institute of Technology & Science, Kuntloor (V), Hayathnagar (M), RR District, Hyderabad, India.
E-mail id: bharathgitta@gmail.com
[2] Professor & HOD, Department of Computer Science & Engineering.
E-mail id: venkat123.pedakolmi@gmail.com.
[3] Assistant Professor, Department of Computer Science & Engineering.
E-mail id: rvs2raju@gmail.com

## Abstract:

*Cloud computing has as of late developed as another stage for conveying, overseeing, and provisioning extensive scale benefits through an Internet-based foundation. On the other hand, concerns of weak data on cloud conceivably cause protection issues. Information encryption ensures information security to some degree, however at the expense of bargained productivity. Searchable symmetric encryption (SSE) permits recovery of encoded information over cloud. Here, spotlight on tending to information protection issues utilizing SSE. Shockingly, figure the protection issue from the part of similitude pertinence and plan strength. We observe that server-side positioning in light of request protecting encryption (PPE) unavoidably spills information protection. To wipe out the spillage, propose a two-round searchable encryption (TRSE) plan that backings top-k multi keyword recovery. In TRSE, utilize a vector space model and Homomorphic encryption. The vector space model serves to give sufficient hunt exactness, and the Homomorphic encryption empowers clients to include in the positioning while the dominant part of figuring work is carried out on the server side by operations just on figure content. Subsequently, data spillage can be wiped out and information security is guaranteed. Exhaustive investigation demonstrates that proposed arrangement appreciates "as-solid as could be expected under the circumstances" security surety contrasted with past SSE plans, while effectively understanding the objective of watchword hunt. Far reaching exploratory results exhibit the proficiency of the proposed agreement.*

## Keywords –

Cloud; data privacy; relevance score; similarity relevance; Homomorphic encryption; vector space model

## I Introduction

Cloud computing, a critical pattern for advanced data service, has become a necessary feasibility for data users to outsource data. Controversies on privacy, however, have been incessantly presented as outsourcing of sensitive Information including e-mails, health history and personal photos is explosively expanding. Reports of data loss and privacy breaches in cloud computing systems appear from time to time. Cloud computing is a long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud as to enjoy the on-demand high-quality application

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 3, March 2015
Available at http://internationaljournalofresearch.org

and services from a shared pool of configurable computing resources. The main threat on data privacy roots in the cloud itself. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully

## 1.1 Existing System

Besides, in order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to users. A series of searchable symmetric encryption schemes have been proposed to enable search on cipher text. Traditional SSE schemes enable users to securely retrieve the cipher text, but these schemes support only Boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage. However, the limited computational power on the user side and the high computational overhead precludes information security.

## 1.2 Problem Statement

The main threat on data privacy roots in the cloud itself. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully. To ensure privacy, users usually encrypt the data before outsourcing it onto cloud, which brings great challenges to effective data utilization. However, even if the encrypted data utilization is possible, users still need to communicate with the cloud and allow the cloud operates on the encrypted data, which potentially causes leakage of sensitive information. A series of searchable symmetric encryption (SSE) schemes have been proposed to enable search on cipher text. Traditional SSE schemes enable users to securely retrieve the cipher text, but these schemes support only Boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. To solve the problem and to improve security without sacrificing efficiency top-k multi I -keyword retrieval was done over encrypted cloud data. The cloud server is considered as "honest-but curious", a model extensively used in SSE and characterized by that the cloud server will honestly follow the designed protocol but is curious to analyze the hosted data and the received queries to learn extra information.

## 1.3 PROPOSED SYSTEM

In this paper, we introduce the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval community are employed, including Homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top k multi-keyword retrieval over encrypted cloud data with high security and practical efficiency. Contributions can be summarized as follows: 1. We propose the concepts of similarity relevance and scheme robustness. We, thus, perform the first attempt to formulate the privacy issue in searchable encryption, and we show server-side ranking based on order-preserving encryption (OPE)

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 3, March 2015
Available at http://internationaljournalofresearch.org

inevitably violates data privacy. 2. We propose a TRSE scheme, which fulfills the secure multi-keyword top-k retrieval over encrypted cloud data. Specifically, for the first time, we employ relevance score to support multikeyword top-k retrieval. 3. Thorough analysis on security demonstrates the proposed scheme guarantees high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization. The rest of this paper is organized as follows: We provide scenario and related background in Section 2, and then we give the security definitions and problems with existing schemes in Section 3. In Section 4, we present the detailed description of the proposed searchable encryption scheme Section 5 concludes this paper. 2

## 2. PRELIMINARIES

2.1 Scenario We consider a cloud computing system hosting data service, as illustrated in Fig. 1, in which three different entities are involved: cloud server, data owner, and data user. The cloud server hosts third-party data storage and retrieve services. Since data may contain sensitive information, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Any kind of information leakage that would affect data privacy are regarded as unacceptable Fig1 Scenario of Retrieval of Encrypted Cloud Data The data owner has a collection of n files (f 1, f 2,....., 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 authorized users. To achieve this, the data owner needs to build a searchable index I from a collection of (w1,w2,.....,wn) extracted out=I keywords w of c , and then outsources both the encrypted index I and encrypted files onto the cloud server. The data user is authorized to process multi-keyword retrieval over the outsourced data. The computing power on the user side is limited, which means that operations on the user side should be simplified. The authorized data user at first generates a query REQ. For privacy consideration, which keywords the data user has searched must be concealed. Thus, the data user encrypts the query and sends it to the cloud server that returns the relevant files to the data user. Afterward, the data user can decrypt and make use of the file In this scheme, the data owner encrypts the searchable index with homo-morphic encryption. When 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. Next, the data user decrypts the scores and picks out the top-k highest scoring files, identifiers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user, thus, name the scheme the TRSE scheme, in which ranking is d one at the user side while scoring calculation is done at the server side

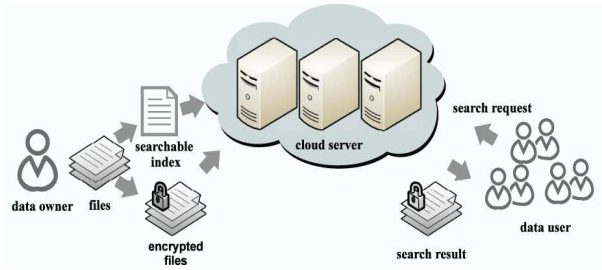


Fig 1. Scenarioo of Retrival of Encrypted Cloud Data

### 2.2 Relevance

Scoring some of the multi-keyword SSE schemes support only Boolean queries, i.e., a file either matches or does not match a query. Considering the large number of data users and documents in the cloud, it is necessary to allow multi-keyword in the search query and return documents in the order of their relevancy with the queried keywords. Scoring is a natural way

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 3, March 2015
Available at http://internationaljournalofresearch.org

to weight the relevance. Based on the relevance score, files can then be ranked in either ascendingly or dissentingly. Several models have been proposed to score and rank files in IR community. Among these schemes, adopt the most widely used one idf ) weighting− idf ) weighting. The (tf −(tf involves two attributes: Term frequency and inverse document frequency. Term frequency denotes the number of occurrences of term t in file f . Document frequency refers to the number of files that contains term t, and the inverse document frequency is defined as: log(N / df ).Where N denotes the total=idf number of files. By introducing the IDF factor, the weights of terms that occur very frequently in the collection are diminished and the weights of terms that occur rarely are increased. 2.3 Vector Space Model idf) depicts the weight of a−While (tf single keyword on a file, employ the vector space model to score a file on multi-keyword. The vector space model is an algebraic model for representing a file as a vector. Each dimension of the vector corresponds to a separate term, i.e., if a term occurs in the file, its value in the vector is nonzero, otherwise is zero. The vector space model supports multi-term and non-binary presentation. Moreover, it allows computing a continuous degree of similarity between queries and files, and then ranking files according to their relevance. It meets needs of top-k retrieval. A query is also represented as a vector ~ q , while each dimension of the vector is assigned with 0 or 1 according to whether this term is queried. Given the scores, files can be ranked in order and, therefore, the most relevant files can be found.

## 3. TRSE DESIGN

Existing SSE schemes employ server-side ranking based on OPE to improve the efficiency of retrieval over encrypted cloud data. However, server-side ranking based on OPE violates the privacy of sensitive information, which is considered uncompromisable in the security oriented third party cloud computing scenario, i.e., security cannot be tradeoff for efficiency. To achieve data privacy, ranking has to be left to the user side. Traditional user-side schemes, however, load heavy computational burden and high communication overhead on the user side, due to the interaction between the server and the user including searchable index return and ranking score calculation. Thus, the user-side ranking schemes are challenged by practical use. A more server siding scheme might be a better solution to privacy issues. We propose a new searchable encryption scheme, in which novel technologies in cryptography community and IR community are employed, including Homomorphic encryption and the vector space model. In the proposed scheme, the data owner encrypts the searchable index with Homomorphic encryption. When 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. Next, the data user decrypts the scores and picks out the top-k highest scoring files' identifiers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user. We, thus, name the scheme the TRSE scheme, in which ranking is done at the user side while scoring calculation is done at the server side.

## 4. Framework of TRSE

The framework of TRSE includes four algorithms: Setup, Index Build, Trapdoor Gen; Score Calculate, and Rank. ). The data owner generates theλSetup (secret key and public keys for the homomorphic isλencryption scheme. The security parameter taken as the input, the output is a secret key SK and a public key set PK. Index Build (C, PK) . The data owner builds the secure searchable index from the file collection C. Technologies from IR community like stemming are employed to build searchable index I from C, and then I is encrypted into I'

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 3, March 2015
Available at http://internationaljournalofresearch.org

with PK , output the secure searchable index I' .TrapdoorGen ( REQ, PK ).The data user generates secure trapdoor from his request. Vector T ' is built from user's multi-keyword request REQ and then encrypted into secure trapdoor T '' with public key from PK PK, output the secure trapdoor T'. Score Calculate (T ' ; I'). When receives secure trapdoor T', the cloud server computes the scores of each files in I' with T ' and returns the encrypted result vector V back to the data user. Rank ((V, SK,K) .The data user decrypts the vector V with secret key SK and then requests and gets the files with top-k scores. Note that Ỹ is only involved in the Setup algorithm, and the Setup algorithm needs to be processed only once by the data owner, ¥ thus, is a constant integer for one individual application instance. The whole framework can be divided into two phases: Initialization and Retrieval. The Initialization phase includes Setup and Index Build. The Setup stage involves the secure initialization, while the Index Build stage involves operations on plaintext. For security concerns, the vast majority of work should only be done by the data owner. Moreover, for convenience of retrieve, we modify the original vector space model by adding each vector Vi a head node idi at the first dimension of Vi to store the identifier of fi. In this way, the correspondence between scores and files is established The Retrieval phase involves Trapdoor Gen, Score Calculate, and Rank, in which the data user and the cloud server are involved. As a result of the limited computing power on the user side, the computing work should be left to server side as much as possible. Meanwhile, the confidentiality privacy of sensitive information cannot be violated.

## 5 CONCLUSIONS

In this paper, we inspire and take care of the issue of secure multi-essential word top-k recovery over scrambled cloud information. We characterize closeness pertinence and plan power .Based on OPE imperceptibly releasing touchy data; we devise a server-side positioning SSE plan. We then propose a TRSE plan utilizing the completely Homomorphic encryption, which satisfies the security necessities of multi-magic word top-k recovery over the scrambled cloud information. By security examination, we demonstrate that the proposed plan ensures information protection. As indicated by the effectiveness assessment of the proposed plan over a genuine information set, far reaching trial results show that our plan guarantees commonsense proficiency. The framework is outlined to take care of the issue of supporting proficient positioned catchphrase hunt down accomplishing compelling usage of remotely put away scrambled information in Cloud Registering

## REFERENCES:

[1]International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3,March2013

[2]R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky,"SearchableSymmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer andComm. Security (CCS), 2006

[3] 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),2010

[4] Cong Wang†, Ning Cao‡, Jin Li†, Kui Ren†, and Wenjing Lou‡ †Department of ECE, Illinois Institute of Technology, Chicago, IL 60616 ‡Department of ECE, Worcester Polytechnic Institute, Worcester, MA 01609 Secure Ranked Keyword Search over Encrypted Cloud Data

[5] A. Singhal, Modern information retrieval: A brief overview, IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 3543, 2001.

[6] R. Ananthakrishna, S. Chaudhuri, and V. Ganti, Eliminating Fuzzy Duplicates in DataWarehouses, Proc. 28th Intl Conf. Very Large Data Bases, pp. 586-597, 2002.

[7] R. Baeza-Yates and B. RibeiroNeto, Modern Information Retrieval. ACM Press,1999.

[8] I. H. Witten, A. Moffat, and T. C. Bell, Managing gigabytes: Compressing and indexing documents and images, Morgan Kaufmann Publishing, San Francisco, May 1999.

[9] E.-J. Goh, Secure indexes, Cryptology ePrint Archive, 2003, http://eprint.iacr.org/2003/216.

[10] D. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data,‖ in Proc. of IEEE Symposium on Security and Privacy'00, 2000.

[11] E.-J. Goh, Secure indexes,‖ Cryptology ePrint Archive, Report 2003/216, 2003, http://eprint.iacr.org/.

[12] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, Public key encryption with keyword search,‖ in Proc. of EUROCRYP'04, volume 3027 of LNCS. Springer, 2004.

[13] Y.-C. Chang and M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data,‖ in Proc. of ACNS'05, 2005.

[14] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, Searchable symmetric encryption: improved definitions and efficient constructions,‖ in Proc. of ACM CCS'06, 2006.

## ABOUT THE AUTHOR

**Gitta. Bharath** pursuing M.Tech (CSE) from Department of Computer Science & Engineering, Nagole institute of technology & science", Hyderabad. India and received B.Tech Degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad, India