ISSN:0975-9646



# Privacy Preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over Encrypted Cloud Data

Shiba Sampat Kale, Prof. Shivaji R Lahane

*Computer Department, University of Pune*
*GES's R. H. Sapat College of Engg Nashik ,India*

***Abstract*— The advancement in cloud computing has motivated the data owners to outsource their data management systems from local sites to commercial public cloud for great flexibility and economic savings. But people can enjoy full benefit of cloud computing if we are able to address very real privacy and security concerns that come with storing sensitive personal information. For real privacy, user identity should remain hidden from CSP (Cloud service provider) and to protect privacy of data, data which is sensitive is to be encrypted before outsourcing. Thus, enabling an encrypted cloud data search service is of great importance. By considering the large number of data users, documents in the cloud, it is important for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective need of data retrieval search and not often differentiate the search results. In this system, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to be implemented in real.**

**We first propose a basic idea for the Multi-keyword Ranked Search over Encrypted cloud data (MRSE) based on secure inner product computation and efficient similarity measure of coordinate matching, i.e., as many matches as possible, in order to capture the relevance of data documents to the search query, then we give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Assignment of anonymous ID to the user to provide more security to the data on cloud server is done. To improve the search experience of the data search service, further extension of the two schemes to support more search semantics is done.**

***Keywords*- Cloud computing, searchable encryption, privacy-preserving, keyword search, ranked search Anonymization, MRSE.**

## I. INTRODUCTION

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect privacy of data and oppose unsolicited accesses in the cloud and beyond it, sensitive data, for instance, e-mails, personal health records, photo albums, tax documents, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The insignificant solution of downloading all the data and decrypting locally is clearly impractical, due to the large amount of bandwidth cost in cloud scale systems. Images also contain useful and important information, so proposed system also provides image tagging in MRSE scheme [1]. Moreover, aside from eliminating the local storage management, storing data into the cloud doesn't serve any purpose unless they can be easily searched and utilized. Hence, exploring privacy-preserving and effective search service over encrypted cloud data is of great importance.

Considering potentially huge number of on-demand data users and large amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability. Document ranking is provided for fast search, but the priorities of all the data documents is kept same so that the cloud service provider and third party remains unaware of the important documents, thus, maintaining privacy of data.

Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. Besides, to improve search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keyword search, as single keyword search often yields far too coarse results. As a common practice indicated by today's web search engines (ex. Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. Along with the privacy of data and efficient searching schemes, real privacy is obtained only if the user's identity remains hidden from the Cloud Service Provider (CSP) as well as the third party user on the cloud server.

## II. LITERATURE SURVEY

*A. Secured Multi-keyword Ranked Search over Encrypted Cloud Data*: In cloud computing data possessor are goaded to farm out their complex data management systems from local sites to the commercial public cloud for greater flexibility and economic savings. To ensure safety of stored data, it is must to encrypt the data before storing. It is

necessary to invoke search with the encrypted data also. The specialty of cloud data storage should allow copious keywords in a solitary query and result the data documents in the relevance order. In [1], main aim is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of "coordinate matching" (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically "inner product similarity", i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm.
The main limitation of this paper was, the user's identity (ID) is not kept hidden. Due to this, whoever puts the data on Cloud Service Provider was known. This may be risky in some situations where confidentiality of data need to be maintained. Hence, this drawback is overcome in the proposed system.

*B. Privacy Preserving Keyword Searches on Remote Encrypted Data*: Consider the problem: a user *U* wants to store his files in an encrypted form on a remote file server *S*. Later the user *U* wants to efficiently retrieve some of the encrypted files containing specific keywords, keeping the keywords themselves secret and not to endanger the security of the remotely stored files. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In [2], solutions for this problem under well-defined security requirements are offered.
The schemes are efficient as no public-key cryptosystem is involved. Indeed, the approach is independent of the encryption method chosen for the remote files. They are incremental too. In that, user *U* can submit new files which are secure against previous queries but still searchable against future queries. From this, the main theme taken is of storing data remotely on other server and retrieving that data from anywhere via mobile, laptop etc.

*C. Cryptographic Cloud Storage:* When the benefits of using a public cloud infrastructure are clear, it introduces significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality and integrity of data. In [3], an overview of the benefits of a cryptographic storage service, for example, reducing the legal exposure of both customers and cloud providers, and achieving regulatory compliance is provided. Besides this, cloud services that could be built on top of a cryptographic storage service such as secure back-ups, archival, health record systems, secure data exchange and e-discovery is stated briefly.

*D. Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data:* On one hand, users who do not necessarily have prior knowledge of the encrypted cloud data, have to post process every retrieved file in order to find ones most matching their interest; On the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data [4]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. For the first time, the paper has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our proposed system in order to enhance the security of data on Cloud Service Provider.

*E. Providing Privacy Preserving in Cloud Computing:* Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust and needs to be considered at every phase of design. The [5] paper tells the importance of protecting individual's privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. Paper tells that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data. From this paper, main theme taken is of preserving privacy of data. This paper only describes privacy of data but doesn't allow indexed search as well as doesn't hide user's identity. Thus, these two drawbacks are overcome in our proposed system.

*F. Privacy Preserving Data Sharing With Anonymous ID Assignment:* In this paper, an algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. In [6], existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements. These new algorithms are built on top of a secure sum data mining operation using Newton's identities and Sturm's theorem. The main idea taken from this paper is of assigning anonymous ID to the user on the cloud.

*G. Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing:* In this paper, main idea is to formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy [7]. This basic idea is taken but it is for multi-keyword raked search (MRSE scheme) in our proposed system. In [8], design of secure cloud storage

service which addresses the reliability issue with near-optimal overall performance is proposed.

*H. Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing:* Achieving fine-grainedness, scalability, and data confidentiality of access control simultaneously is a problem which actually still remains unresolved. The paper [9] addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine-grained data access control to untrusted cloud servers without disclosing the underlying data contents. In [10], authors have proposed a privacy-preserving public auditing system for data storage security in Cloud Computing scheme is proposed. It utilizes the homomorphic linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which eliminates the burden of cloud user from the tedious and possibly expensive auditing task, it also alleviates the user's fear of his/her outsourced data leakage.

## III. PROPOSED SYSTEM

Considering a cloud data hosting service involving three different entities, the data owner, the data user along with his ID, and the cloud server. The data owner first registers on cloud using anonymity algorithm for cloud computing services. Before saving user registration information to database present on cloud anonymous algorithm process the data and then anonymous data is saved to registration database. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable searching capability over C for effective data utilization, the data owner, will first build an encrypted searchable index I from F before outsourcing , and then outsource both the index I and the encrypted document collection C to the cloud server. The work deals with efficient algorithms for assigning identifiers (IDs) to the users on the cloud in such a way that the IDs are anonymous using a distributed computation with no central authority. Given are N nodes, this assignment is essentially a permutation of the integers {1….N} with each ID being known only to the node to which it is assigned. Our main algorithm is based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data. To search the document collection for given keywords, an authorized user having an ID acquires a corresponding trapdoor T through search control mechanisms, for example, broadcast encryption. On receiving T from a data user, cloud server is responsible to search the index I and then returns the corresponding set of encrypted documents. In order to improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching) and assigning anonymous ID [6] to the user on cloud in order to make the data on cloud more secure. Moreover, to reduce the cost of communication the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. At last, the access control mechanism is employed in order to manage decryption capabilities given to users and the data collection can be updated in terms of inserting new documents, updating existing ones, and deleting the existing documents.

TABLE I

REVIEW SUMMARY

| Sr. No. | Paper Title | Objective |
|---|---|---|
| 1 | Secured Multi-keyword Ranked Search over Encrypted Cloud Data | Main focus is on the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. |
| 2 | Privacy Preserving Data Sharing With Anonymous ID Assignment | Main objective is to assign user an anonymous ID |
| 3 | Providing Privacy Preserving in Cloud Computing | The main idea is protecting individuals' privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. |
| 4 | Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data | This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data. |
| 5 | Privacy Preserving Keyword Searches on Remote Encrypted Data | Main objective is to to get the access to user's data which is stored remotely from anywhere according to user's convenience |
| 6 | Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing | Main idea is to formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. |

## IV. CONCLUSIONS

The previous work [1] mainly focused on providing privacy to the data on cloud in which using multi-keyword ranked search was provided over encrypted cloud data using efficient similarity measure of co-ordinate matching. The previous work [4] also proposed a basic idea of MRSE using secure inner product computation. There was a need to provide more real privacy which this paper presents. In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the CSP and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained.

## REFERENCES


[1] Ankatha Samuyelu Raja Vasanthi ," Secured Multi keyword Ranked Search over Encrypted Cloud Data", 2012

[2] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," *Proc. Third Int'l Conf. Applied Cryptography and Network Security*, 2005.

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

[4] Y. Prasanna, Ramesh . "Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data", 2012.

[5] Jain Wang, Yan Zhao , Shuo Jaing, and Jaijin Le, "Providing Privacy Preserving in Cloud Computing",2010.

[6] Larry A. Dunning, Ray Kresman ," Privacy Preserving Data Sharing With Anonymous ID Assignment",2013.

[7] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," *Proc. IEEE INFOCOM*, Mar. 2010.

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

[9] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," *Proc. IEEE INFOCOM*, 2010.

[10] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," *Proc. IEEE INFOCOM*, 2010.

[11] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "Privacy preserving Query over Encrypted Graph-Structured Data in Cloud Computing," *Proc. Distributed Computing Systems (ICDCS*), pp. 393-402, June, 2011.