A NOVEL APPROACH FOR MULTI-KEYWORD SEARCH WITH ANONYMOUS ID ASSIGNMENT OVER ENCRYPTED CLOUD DATA

U.Pandi Priya[1], R.Padma Priya[2]

[1]Research Scholar, Department of Computer Science and Information Technology,
Nadar Saraswathi College of Arts and Science, (India)
[2]Assistant Professor, Department of Computer Application,
Nadar Saraswathi College of Arts and Science, (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 multikeyword 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 multikeyword 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, Keyword Search, Mrse, Privacy Preserving, Ranked Search Anonymization, Searchable Encryption,

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, personalhealth 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 searches, 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. MODELS

2.1 System Model

Considering a cloud data hosting service involving three different entities, the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index I from F, and then outsource both the index I and the encrypted document collection C to the cloud server. To search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control Mechanisms, e.g., broadcast encryption. Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. 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, as will be introduced shortly). Moreover, to reduce the communication cost, 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. Finally, the access control mechanism is employed to manage decryption capabilities given to users.

2.2 Threat Model

The cloud server is considered as "honest-but-curious" in our model, which is consistent with related works on cloud security [24], [25]. Specifically, the cloud server acts in an "honest" fashion and correctly follows the designated protocol specification. However, it is "curious" to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information. Based on what information the cloud server knows, we consider two threat models with different attack capabilities as follows.
Known Cipher text Model. In this model, the cloud server is supposed to only know encrypted dataset *C* and searchable index *I*, both of which are outsourced from the data owner. Known Background Model In this stronger model, the cloud server is supposed to possess more knowledge than what can be accessed in the Known Cipher text Model. Such information may include the correlation relationship of given search requests (trapdoors), as well as the dataset related statistical information. As an instance of possible attacks in
This case, the cloud server could use the known trapdoor information combined with document/keyword frequency to deduce/identify certain keywords in the query.

2.3. Design Goals

To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, our System design should simultaneously achieve security and performance guarantees as follows.
• Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide Result similarity ranking for effective data retrieval, instead of returning undifferentiated results.
• Privacy-Preserving: To prevent the cloud server from learning additional information from the dataset and the index, and to meet privacy requirements.
• Efficiency**:** Above goals on functionality and privacy should be achieved with low communication and computation overhead.

2.4 Preliminary on Coordinate Matching

As a hybrid of conjunctive search and disjunctive search, "coordinate matching" [4] is an intermediate similarity measure which uses the number of query keywords appearing in the document to quantify the relevance of that document to the query. When users know the exact subset of the dataset to be retrieved, Boolean queries perform well with the precise search requirement specified by the user. In cloud computing, however, this is not the practical case, given the huge amount of outsourced data. Therefore, it is more flexible for users to
Specify a list of keywords indicating their interest and retrieve the most relevant documents with a rank order.

III. FRAMEWORK AND PRIVACY REQUIREMENTS FOR MRSE

In this section, we define the framework of multi-keyword ranked search over encrypted cloud data (MRSE) and establish various strict system-wise privacy requirements for such a secure cloud data utilization system.

3.1MRSE Framework

For easy presentation, operations on the data documents are not shown in the framework since the data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data. With focus on the index and query, the MRSE system consists of four algorithms as follows.
• **Setup (1$^{\ell}$)** Taking a security parameter ℓ as input, the data owner outputs a symmetric key as SK.

• **Build Index (*F, SK*)** Based on the dataset F, the data owner builds a searchable index I which is encrypted by The symmetric key SK and then outsourced to the cloud server. After the index construction, the document collection can be independently encrypted and outsourced.

• **Trapdoor (f*W*)** with t keywords of interest in fW as input, this algorithm generates a corresponding trapdoor TfW.

• **Query(*TfW, k, I*)** When the cloud server receives a query request as (TfW, k), it performs the ranked search on the index I with the help of trapdoor TfW, and finally returns FfW, the ranked id list of top-k documents sorted by their similarity with fW.

Neither the search control nor the access control is within the scope of this paper. While the former is to regulate how authorized users acquire trapdoors, the later is to manage users' access to outsourced documents.

3.2. Privacy Requirements for MRSE

The representative privacy guarantee in the related literature, such as searchable encryption, is that the server should learn nothing but search results. With this general privacy description, we explore and establish a set of strict privacy requirements specifically for the MRSE framework. As for the *data privacy*, the data owner can resort to the traditional symmetric key cryptography to encrypt the data before outsourcing, and successfully prevent the cloud server from prying into the outsourced data. With respect to the *index privacy*, if the cloud server deduces any association between keywords and encrypted documents from index, it may learn the major subject of a document, even the content of a short document. Therefore, the searchable index should

Be constructed to prevent the cloud server from performing such kind of association attack. While data and index privacy guarantees are demanded by default in the related literature, various *search privacy* requirements involved in the query procedure are more complex and difficult to tackle as follows.

IV. PROPOSED ALGORITHM

4.1Homo Morphic Token Pre-computation

Homomorphism encryption is a form of encryption that allows computations to be carried out on ciphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. This is sometimes a desirable feature in modern communication system architectures. Homomorphic encryption would allow the chaining together of different services without exposing the data to each of those services Homomorphic encryptions allow complex mathematical operations to be performed on encrypted data without compromising the encryption.

4.2 Algorithm

1: **procedure**

2: Choose parameters l, n and function f;

3: Choose the number t of tokens;

4: Choose the number r of indices per

Verification;

5: Generate master key

6: **for** vector G(j), j ← 1, n **do**

7: **for** round i← 1, t **do**

8: Derive _i = fkchal (i) and k(i) prp
from KPRP .
9: Compute v (j)
i =Pr
q=1 * G(j)[_k(i)prp (q)]
10: **end for**
11: **end for**
12: Store all the vis locally.
13: **end procedure**

V. 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 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.

VI. CONCLUSION

The previous work mainly focused on providing privacy to the data on cloud in which using multi-keyword ranked search. In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data using Homomorphism token Pre-Computation and establish a variety of privacy requirements. 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 Searchon 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 ssignment",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. Shiba Sampat Kale et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7093-7096 www.ijcsit.