Privacy Preserving Data Search Using Third Party in Cloud Computing

Ruturaj Desai[1], Prof. Nitin R. Talhar[2]

[1]AISSMS COE, Savitbai Phule Pune University,
Pune, India

[2]AISSMS COE, Savitbai Phule Pune University,
Pune, India

Abstract

Cloud Computing is new and most importing computing technology. Because of different advantages of cloud computing, many data owners and organizations are outsourcing there data management systems to public cloud to achieve higher flexibility and to reduce cost. The most important issues i cloud computing are privacy and security. Data privacy can be protected by encrypting sensitive data locally before outsourcing that data. But the data utilization based on single keyword search is prevented because of data encryption. Thus, enabling an encrypted cloud data search service is very important. Consider, large number of data users and files in cloud, it is important for the search service to allow multi-keyword query and provide results. Retrieving of all files having queried keywords will not be affordable in pay as per use cloud paradigm. In this paper, we proposed new scheme to solve the problem of multi-keyword search over encrypted data and data sharing using trusted third party in cloud computing. We establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. We are using different cryptographic method like AES,Base64 and BlowFish for file encryption. We are using the effective principle of coordinate matching, i.e., as many matches as possible, to capture the similarity between search query and the data file. We propose the system using the trusted third party which will allow user to share data stored on cloud without compromising data privacy. Through analysis investigating privacy and efficiency guarantee of proposed scheme is given and experiments further show proposed scheme indeed introduce low overhead on computation and communication

Keywords: Cloud Computing, Encryption, Multi-Keyword search, Coordinate Matching, Trusted Third Party.

1. INTRODUCTION

Cloud computing is the computing technique which describes the combination of logical entities like data, software which are accessible via internet. Cloud computing provides help to the business applications and functionality along with the usage of computer software by providing remote server which access through the internet. Client data is generally stored in servers spread across the globe. Cloud computing allows user to use different services which saves money that users spend on applications. Data owners and organizations are motivated to outsourced more and more sensitive information into the cloud servers, such as emails, personal documents, videos and photos, company finance data, government documents, etc.[1][2]

To provide end-to-end data security and privacy in the cloud, sensitive data has to be encrypted before outsourcing to protect data privacy. In cloud computing, effective data utilization is a very difficult task because of data encryption, also it may contain large amount of outsourced data files. Data owners may wants to share their outsourced data with other large amount of users. Users may want to only retrieve certain specific data files they are interested in during a given session. Most popular way to do so is through key word based search. The keyword based search technique allows users to selectively retrieve files of interest.[3] this technique is widely applied in plaintext search scenarios. Unfortunately the traditional plaintext search technique in encrypted data cloud because of demand of the protection of search keyword privacy and data encryption, which restrict user's ability to perform keyword search on data.

To overcome the above problem in this paper new technique is introduced which allows user to perform Ranked search on data cloud. In Ranked search, normal matching files are arranged in ranked order regarding to certain relevance criteria which greatly improves system usability. In the ”pay as-you-use” cloud paradigm it is highly desirable.[4] Ranked search elegantly eliminates unnecessary network traffic by sending back only the most relevant data. It is very important that, such ranking operation should not leak any keyword related information to protect privacy of that keyword. Single keyword search often yields far too often coarse results, so it is necessary for rank system to support multiple keyword search which will improve the user search result accuracy and enhance the user searching experience. To retrieve the most relevant data, users tend to provide a set of keywords instead of only one keyword. It is very important that the user's identity remains hidden from the Cloud Service Provider (CSP) as well as the third party user on the cloud server to provide privacy.

2. PROPOSED SYSTEM

Consider cloud service contains four different entities, as listed in fig. 1: The data owner, the data user, the trusted third party (TTP), and the cloud server. The data owner and the data user will register on the cloud for cloud computing services. The data owner will outsource the

collection of data files F which to cloud server. To provide data security, data files F must be outsourced in the encrypted form C. Before outsourcing data files to the cloud, encrypted searchable index I is generated from the file F, which will allow to improve the searching capability
over C for effective data utilization. After this, encrypted data files collection C is outsourced to the cloud serve and the encrypted index I is outsourced to the trusted third party. The authorized data user can perform search on the file collection using K keywords. Data user will perform search using K keywords. Upon receiving K in encrypted form from data user, cloud server will authenticate the user and will send those keywords to the trusted third party. Trusted third party will search all available indexes I using ”string matching” and send appropriate and most relevant results to the cloud server. To improve the searching accuracy trusted third party will rank those results. Cloud server will send those search results to the appropriate data user. The communication cost can be reduced by sending appropriate results to the data user. The access control mechanism is applied to manage to decryption capabilities given to user.

2.1 Notations

- F - The file collection, denoted as a set of m files
- F = (f_1, f_2, f_3, ..., f_m).
- C- The encrypted file collection stored in the cloud server, denoted as C = (C_1, C_2, C_3, ... , C_m).
- I- The searchable index associated with C, denoted as $I_{1,}$ $I_{2,}$ I_3 , ..., I_m where each sub-index I_i is built for F_i.
- Q - is the search , and k representing the keywords in a search request, denoted as k = (k_1, k_2, k_3, ..., k_j).
- F_Q - The id list of all files according to their relevance to Q.

2.2 Algorithms

Traditional Symmetric key cryptography i.e. AES, is used by clients for data encryption and decryption. Following are few other algorithms which are used:

1) KeyGen: By considering all security parameters this algorithm will generate two symmetric keys. First symmetric key **SK₁** is user specific key, each user has different symmetric key. and second symmetric key **SK₂** is file specific key. Public key **PK** is also used by user.

2) IndexGen(f, SK₁, SK₂, PK): This algorithm will generate searchable index from file f and will encrypt that index using **PK**. This encrypted index is outsourced to the trusted third party. After index generation, files are encrypted using combination of **SK₁, SK₂, PK** and outsourced to the cloud.

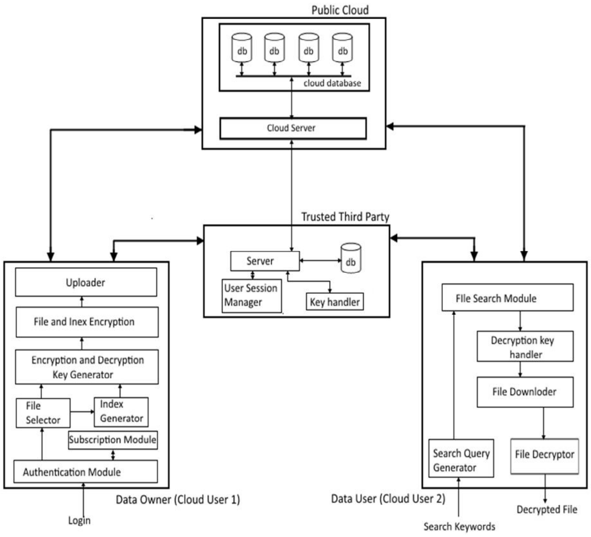

Figure 1 Architecture of Data Search Over Cloud

$$I_i \begin{cases} I'_{DF} = \sum_{i=0}^{n} (RW_i)_{k_i} & if\ f = document\ file \\ I'_{MF} = \sum_{i=1}^{n} S_i + R_i + N_i + E_i & if\ f = media\ file \\ I'_{OF} = \sum_{i=1}^{n} S_i + N_i + E_i & if\ f = other\ file \end{cases} \quad (1)$$

Where,

I'_{DF}= index for document file,
I'_{MF}= index for media file,
I'_{OF}= index for other file.

3) KeyExchange *(f, SK₁, SK₂)*:This algorithm will allow to exchange keys between Clients, Cloud and Trusted third party.

Steps:

1. If User Type=Data Owner
then, upload FileSpecific key ***(SK₂)***, Rule***(R)***, File ID into Trusted Third Party. And, upload UserSpecific key ***(SK₁)***, into the cloud.
2. If User Type=Other User
then, download the FileSpecific key ***(SK₂)***, Rule***(R)*** from the third party. And, download the UserSpecific key ***(SK₁)***, form the cloud.(For specific File ID).

4) SearchQuery*(Q)*: This algorithm allows user to perform ranked search. User will send search query ***(Q)*** to the cloud which is encrypted using public key ***(PK)***. Search query ***(Q)*** contains set of words which user wants to search. Cloud server will authenticate that request and will forward the ***Q*** to the trusted third party. Trusted third party will perform ranked search over the saved index using ***Q*** and returns the ***FQ*** the ranked id list of files similar to the ***Q***.

$$F_Q = Results(Q,k) = Q_{TP}\left(\sum_{i=1}^{n} I_i \times \sum_{s=1}^{m} k_s\right) \quad (2)$$

3. PERFORMANCE ANALYSIS

We implemented the entire secure search scheme to evaluate the overall performance of our technique using JAVA on windows with Intel Core i5 processor 3.3GHz.

We built the file set which contains document files, media files such as videos, images, audios and other types of files. In this section we present the detailed performance result. Performance analysis of this new system is done by comparing results with the existing system as follows.

3.1 Index Generation

In this section, the time required to generate indices with respect number of documents is measured. Time of index generation is measured in seconds. Index generation time of proposed system PPDS is compared with two existing system MRES-II and MRES-III. This compression show that proposed system takes less time to generate indices than existing systems.

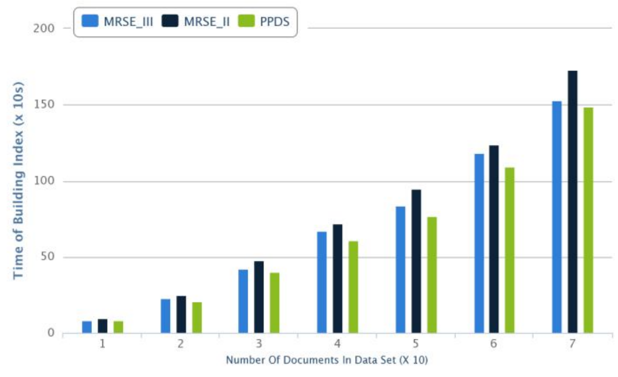

Figure 2 Time Required for Index Generation

3.2 Query Generation and Search Results

In this section, time required to generate query and search files for proposed system is measured and compared with the existing systems MRES-II and MRES-III. The compression show that the proposed system is more efficient than existing systems.

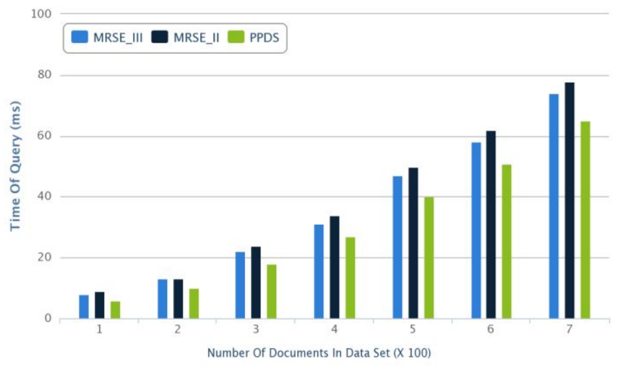

Figure 3 Time Required for Query Generation And Search Results

4. Conclusion

In this paper, a new system is proposed which solve the multi keyword search over encrypted cloud data problem in cloud computing. The efficient principle of "coordinate matching" is selected form different multi-keyword semantics, to effectively capture similarity between query keywords and outsourced documents. The proposed system perform secure search over encrypted data in cloud computing with the help of trusted third party. The proposed system allow user to perform secure search on cloud data and allows to share data with other users. This improve communication privacy and security with reduction of the communication cost. The proposed system improves the accuracy and the privacy.

References

[1] Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Louy, ”Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 1, IEEE 2014.

[2] Wenhai Sun, BingWang, Ning Cao, Ming Li,Wenjing Lou, Hou, Y.T., Hui Li,”Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking”, IEEE Transactions on Parallel and Distributed Systems, IEEE 2014.

[3] Ankatha Samuyelu, Raja Vasanthi , ”Secured Multi keyword Ranked Search over Encrypted Cloud Data”, 2012.

[4] Y.C. Chang and M. Mitzenmacher, ”Privacy Preserving Keyword Searches on Remote Encrypted Data”, Proc. Third Intl Conf. Applied Cryptography and Network Security, 2005.

[5] S. Kamara and K. Lauter, ”Cryptographic Cloud Storage”, Proc. 14th Intl Conf. Financial Cryptograpy and Data Security, Jan. 2010.

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

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

[8] CongWang, Chow, S.S.M., QianWang, Kui Ren , Wenjing Lou, ”PrivacyPreserving Public Auditing for Secure Cloud Storage”, IEEE Transactions on Computers, IEEE 2013.

[9] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, ”LT Codes-Based Secure and Reliable Cloud Storage Service ”, Proc. IEEE INFOCOM, pp. 693701, 2012.

[10] 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.

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

[12] 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.