International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Secure Data Searching and Sharing Using Third Party in Data Cloud

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

[1]AISSMS College of Engineering, Savitribai Phule Pune University, Pune, India

[2]Professor, AISSMS College of Engineering, Savitribai Phule Pune University, Pune, India

Abstract: ***Data privacy and security are the most important issues in cloud computing. To achieve higher flexibility and to reduce the cost, many data owners are outsourcing their data management systems to public cloud. Data privacy can be protected by encrypting sensitive data locally before outsourcing. The data encryption prevents the data utilization based on simple keyword search. 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 "payas-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. Related Work

A. Privacy Preserving multi-keyword ranked search over encrypted cloud data: Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper[1], for the first time, they define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). To improve the search result and privacy, they introduce different MRSE schemes. they first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

B. Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking: In this paper[2], they have proposed tree-based index structure methods for multi-dimensional algorithm to improve the search efficiency. They have proposed two new secure index schemes to meet the stringent privacy requirements under strong threat models.

C. Secured Multi-keyword Ranked Search over Encrypted Cloud Data: In this paper [3], 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.

D. 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 [4], 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.

E. 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 [5], 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.

F. 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 [6] paper tells the importance of protecting individuals 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 users identity. Thus, these two drawbacks are overcome in our proposed system.

G. 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 payas-you-use cloud paradigm. This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data[7]. 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.

H. Privacy-Preserving Public Auditing for Secure Cloud Storage: Cloud storage is widely used now days by user to outsource their data.[8]The large size of outsource data makes the data integrity protection in Cloud Computing a very challenging and potentially formidable task, especially for users with constrained computing resources and capabilities. In this paper, third part auditing (TPA) is introduced. we utilize the public key based homomorphic authenticator and uniquely integrate it with random mask technique to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind.

I. 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 [9]. This basic idea is taken but it is for multi-keyword raked search (MRSE scheme) in this proposed system. In [10], design of secure cloud storage service which addresses the reliability issue with near optimal overall performance is proposed.

J. Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing: Achieving scalability and data confidentiality of access control simultaneously is a problem which actually still remains unresolved. The paper [10] 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 [11], 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 users fear of his/her outsourced data leakage

3. Proposed System

3.1 System Model

Consider cloud service contains four different entities, as listed in figure 1: The data owner, the data user, the trusted third party (TTP), and the cloud server. At the beginning, the data owner and the data user will register on the cloud for cloud computing services. The data owner has collection of data files F which is to be outsourced 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

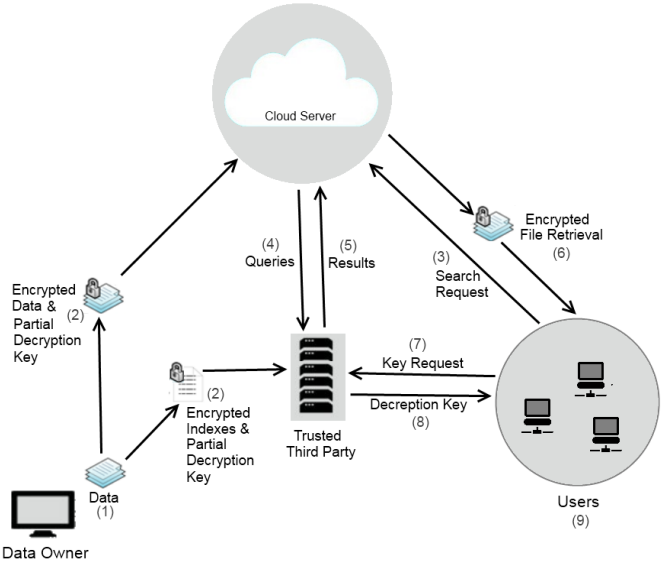

Figure 1: Framework of the search over encrypted cloud data

appropriate results to the data user. The access control mechanism is applied to manage to decryption capabilities given to user.

3.2 Threat Model

Cloud server is considered as "honest-but-curious" in our model, which is consistent with the most related works on searchable encryption. Specifically, cloud server acts in an "honest" fashion and correctly follows the designated protocol specification. However, it is "curious" to infer and analyze data in its storage and message flows received during the protocol so as to learn additional information. based on what information cloud server knows, we consider two levels of threat models as follows.

Known Ciphertext Model In this model, cloud server is supposed to only know encrypted dataset C and searchable index *I*, both of which are outsourced from data owner.

Known Background Model In this stronger model, cloud server is supposed to possess some backgrounds on the dataset, such as the subject and its related statistical information, in addition to what can be accessed in known ciphertext model. As an instance of possible attacks in this case, cloud server could utilize document frequency or keyword frequency to identify keywords in the query

3.3 Design Goals

For effective utilization of outsourced data, our system should achieve security and performance guarantee as follows:

- Secure keyword search: to explore different mechanisms for designing effective keyword search schemes based on the existing searchable encryption framework.
- Secure Data Sharing: to allow user to share data over the cloud without losing privacy.
- Security guarantee: to prevent cloud server from learning the plaintext of either the data files or the searched keywords, and to achieve the as strong as possible security

strength compared to existing searchable encryption schemes.
- Efficiency: above goals should be achieved with minimum communication and computation overhead.

3.4 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.

3.5 Preliminary on Coordinate Matching

As a hybrid of conjunction search and disjunction search, "coordinate matching" is an intermediate approach which uses the number of query keywords appearing in the document to quantify the similarity 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 rank order.

3.6 Algorithms Used

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 $\boldsymbol{SK_1}$ is user specific key, each user has different symmetric key. and second symmetric key $\boldsymbol{SK_2}$ is file specific key. Public key ***PK*** is also used by user.

2) *IndexGen(f, SK_1, SK_2, 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 $\boldsymbol{SK_1}$, $\boldsymbol{SK_2}$, ***PK*** and outsourced to the 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, SK_1, SK_2)*: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 $\boldsymbol{(SK_2)}$, Rule***(R)***, File ID into Trusted Third Party. And, upload UserSpecific key $\boldsymbol{(SK_1)}$, into the cloud.

2. If User Type=Other User
then, download the FileSpecific key $\boldsymbol{(SK_2)}$, Rule***(R)*** from the third party. And, download the UserSpecific key $\boldsymbol{(SK_1)}$, 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***.

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

4. 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.

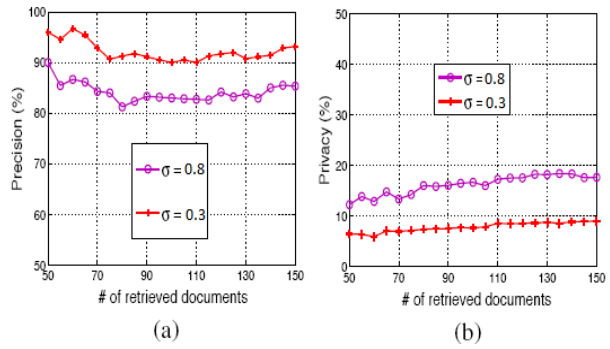

Figure 2: With different choice of standard deviation σ for the random variable ε, there exists tradeoff between (a) Precision, and (b) Rank Privacy.

4.1 Precision and Privacy

To evaluate the impact on accuracy of search result, we adopt the definition of "precision" in [1]. Namely, the precision of a top -k search is defined as $P_k=k'/k$ *where* k' is the number of the real top-k files that are returned by the cloud server. Figure 2(a) shows that the precision in our scheme is evidently affected by the standard deviation σ of the random variable ε. From the consideration of effectiveness, standard deviation σ is expected to be smaller so as to obtain high precision indicating the good purity of retrieved files. On the other hand, we evaluated the "rank privacy", whose definition is also adopted from [1], i.e., the

rank privacy at point k is calculated as $\widetilde{P_k} = \sum \widetilde{p_i} / k$ for every file i in the returned top-k files. Figure 2(b) shows the rank privacy at different points with two standard deviations $\sigma = 0.8$ and $\sigma = 0.3$ respectively. From these two figures, we say that our scheme provides a balance parameters for data users to satisfy their different requirements on precision and rank privacy.

4.2 Index Generation

Figure 3(a) shows number of distinct keyword in dataset. Figure 3(b) shows that with the same documents set, the index construction time is proportional to the number of keywords in the dictionary. On the other hand, considering the massive storage cost in the cloud, the storage overhead is practical and completely affordable.

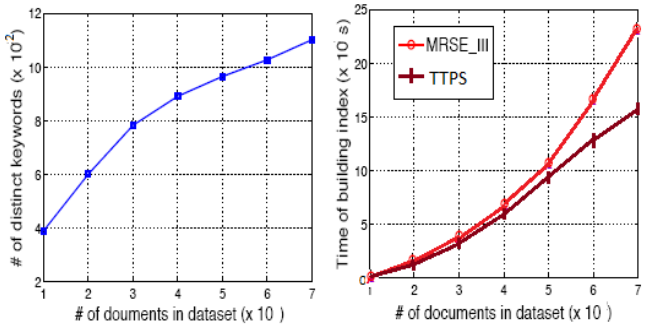

Figure 3: Relationship between number of documents in dataset and (a) Number of distinct keywords in dataset, and (b) Time cost for building searchable index.

4.3 Query Generation

Figure 4(a) demonstrates that when $|T_i|$ is fixed, the time cost for generating an encrypted query is only linear to the number of keyword present in query. While the computation is linear with the number of query keyword in other multi-keyword search schemes [1], [2]. our proposed scheme enjoy the constant overhead in the query which makes it more practical in the cloud paradigm as shown in figure 4(b).

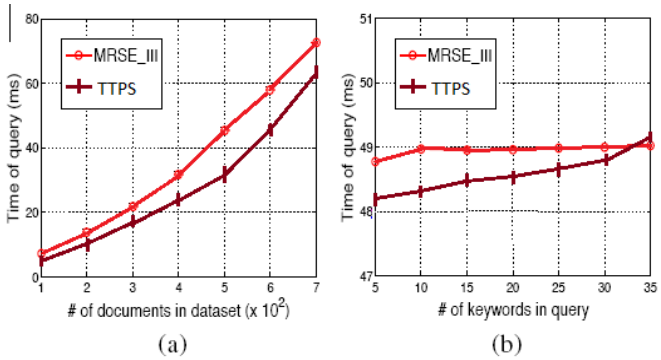

Figure 4: Time cost of query. (a) For the same number (10) of keywords in a query within different number of documents in dataset. (b) For different number of keywords in a query within the same number (50) of documents in dataset.

5. Conclusion

In this paper we proposed new system which will solve the multi keyword search over encrypted cloud data problem in cloud computing. We chose ”coordinate matching” for accurate search results. The proposed system will perform secure search over encrypted data in cloud computing with the help of trusted Third Party. This proposed system will allow user to perform secure search on cloud data and allows to share data with other users. This will improve communication privacy and security and will reduce the communication cost. The proposed system improves the accuracy and the privacy. In future scope, we will expend up to sky computing and will provide security and privacy in multi-user environment.

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.