Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Privacy Preserving Data Search in Cloud Computing

Ruturaj Desai, Nitin R. Thalhar
Dept. Computer Engineering, AISSMS COE,
Pune, India

Abstract— Privacy and security are the most important issues in cloud computing. To achieve high flexibility and to reduce cost, many data owners are outsourcing their data management system to public cloud. Data must be encrypted locally before outsourcing to protect data privacy. The data encryption reduces the data utilization based on simple keyword search. Consider large number of documents are outsourced on cloud by large number of cloud users. It is mandatory for the search service to provide results similarity ranking to provide the accurate search results. Retrieving of all the files having queried keyword will not be affordable in pay as per use cloud paradigm.

In this paper, we propose new scheme to solve the problem of multi keyword search over encrypted data using trusted third party in cloud computing. User will encrypt their data locally. Before encrypting data, the index will be created. Trusted third party will use all these indexes to search data similar to the search query of user. Using these search results, cloud server will send encrypted document to the user.

Keywords--- Cloud Computing, Privacy Preserving, Trusted Third Party, Keyword Search, Encryption.

I. INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., network, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort of service provides interaction. Cloud computing means a remote server that access through the internet which helps in business applications and functionality along with the usage of computer software[3]. Cloud computing saves money that users spend on annual or monthly subscription. Due to advantage of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records, private videos and photos, company finance data, government documents, etc.

So to provide end-to-end data confidentiality assurance in the cloud, confidential data has to be encrypted before outsourcing to protect data privacy.

Data encryption makes effective data utilization a very difficult task given that there could be a large amount of outsourced data files[2][3]. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. This keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios[1][2]. Unfortunately, data encryption, which restricts user's ability to perform keyword search and father demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Ranked search greatly improves system usability by normal matching files in ranked order regarding to certain relevance criteria.

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. 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.

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 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.

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 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[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.

III. PROPOSED SYSTEM

Consider the Cloud data hosting service contains four different entities, as listed in fig. 1: the data owner, the data user, the trusted third party, and the cloud server. Consider data owner will registers on cloud for cloud computing service. Anonymous algorithm is used to process the registration information of user and then saves anonymous data to registration database. The data owner has a collection of data documents D to be outsourced to the cloud server in the encrypted form E. Before outsourcing, the data owner will first build an encrypted searchable index I from D to enable searching capability over E for effective data utilization. The data owner will outsource the encrypted document collection D to the cloud server and encrypted index to the trusted third party. The trusted third party will check the integrity of outsourced data without violating user privacy policies. Anonymous identifiers are assigned to user using efficient algorithms. The data user send the encrypted search query to the cloud server along with his session ID. This encrypted search query is transferred to the trusted third party for processing by cloud server. The trusted third party will search index using "string matching" and sends the search results to the cloud server which returns the corresponding set of encrypted documents to the data user.

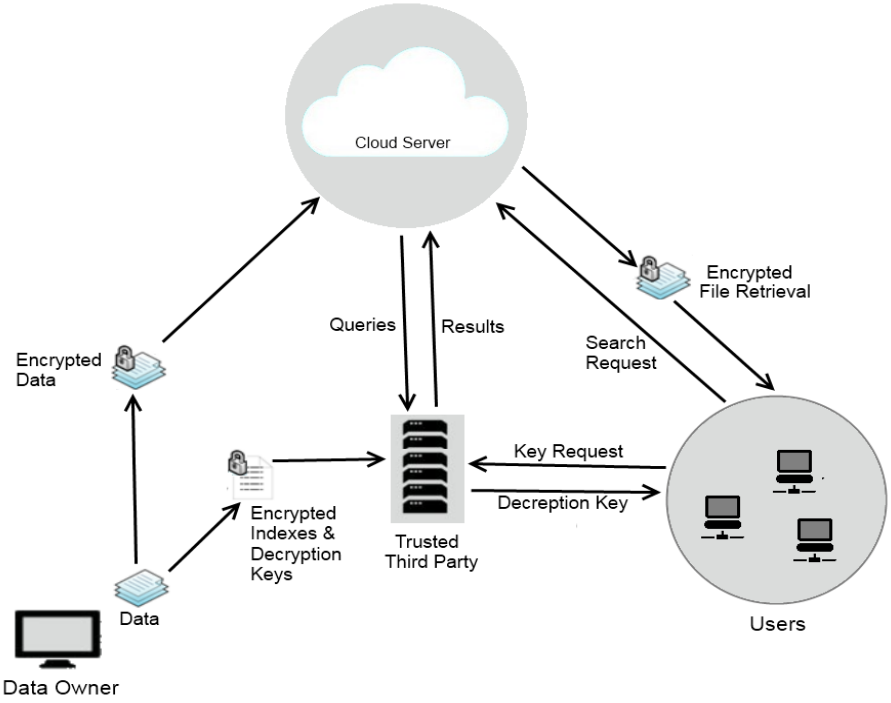

Fig.1 Architecture of search over encrypted data cloud

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 data set and the index, and to meet privacy requirements.

Efficiency. Above goals on functionality and privacy should be achieved with low communication and computation overhead.

To improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria. Finally, the access control mechanism is employed to manage decryption capabilities given to users and the data collection can be updated in terms of inserting new documents, updating existing documents, and deleting existing documents.

IV. CONCLUSIONS

The previous work 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 also proposed a basic idea of MRSE using secure inner product computation for multi-keyword ranked search. There was a need to provide more real privacy which this paper presents. In this paper, the method is proposed to perform the multi-keyword ranked search over cloud data. the proposed system will perform secure search over encrypted data in cloud computing.

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, Bing Wang, 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 Int'l Conf. Applied Cryptography and Network Security, 2005.

[5] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l 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] Cong Wang, Chow, S.S.M., Qian Wang, Kui Ren , Wenjing Lou," Privacy-Preserving Public Auditing for Secure Cloud Storage ", IEEE Transactions on Computers, IEEE 2013.

