IJCAT - International Journal of Computing and Technology, Volume 3, Issue 9, September 2016
ISSN : 2348 - 6090
www.IJCAT.org

Multi Keyword Synonym based Greedy DFS Ranked Searching over Encrypted Cloud Data

Narendra s. Joshi

Computer Engineering
Nashik, India

Abstract - With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. 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.

Keywords **- Cloud Computing, Searchable Encryption, Privacy-Preserving, Keyword Search, Ranked Search**.

1. Introduction

In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We 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. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication**.**

2. Ease of Use

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can 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 data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud. This is, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus, exploring privacy preserving and effective search service over encrypted cloud data is of paramount importance.

On the one hand, to meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection.

3. Objective

Following are the objectives

- To preserve data privacy on cloud using encryption technique
- To restrict non-privileged users and illegal access to data on cloud
- To trigger efficient search on cloud by using multi-keyword ranked search technique
- In search technique , preserve indexed search with equal priorities to all data

4. Problem Definition

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. 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.

Hence as per the available techniques we should design such system having security of data on cloud with efficient search.

5. Literature Survey

5.1 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 great 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 results 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, who ever put the data on Cloud Service Provider was known. This may be risky in some situations. Hence, this drawback is overcome in proposed system.

5.2 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, you 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.

5.3 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.

6. Proposed Methodology

As per our objectives and from security point of view we have to propose a system that is having high level data privacy on cloud. For data privacy system must be armed with encryption techniques. Along with the encryption of data on cloud we have to propose techniques that help to search such encrypted data. As we know cloud is following strategy like "pay-as-you-use". Hence sometimes search technique is so costly as it consider huge amount of data on cloud so at time of result fetching system must consider the large amount of data users and documents in the cloud and must consider crucial nature of searching so system should contain multi-keyword query service.

For effective bandwidth utilization proposed system must be armed with result similarity ranking service to meet effective retrieval. Also System should implement proper techniques for encryption. System should implement proper techniques for searching such as coordinate matching, inner product similarities. As a part of contribution system should contain module that hides user identity.

7. System Architecture

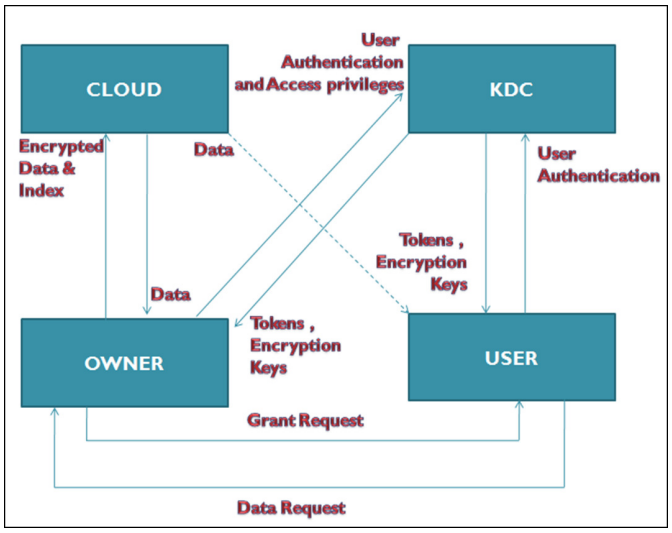

Fig 1: System Architecture

7.1 Methodology

Methodology for Registration:

Step 1: User opens URL of registration
Step 2: User provide necessary details and register on KDC (Key Distribution Cloud)
Step 3: KDC provide token to User (Via email)

Methodology to Upload File on Cloud:

Step 1: User Login
Step 2: User select files to upload
Step 3: User first get encryption key from KDC
Step 4: User get index of files
Step 5: User encrypt index
Step 6: SHA-1 algorithm to encrypt index
Step 7: Encrypt document using encryption key given by KDC

7.2 Methodology to Share document with other User

Step 1: Add user access privileges to data structure present on KDC

7.3 Methodology to Search (Search by other user)

Step 1: User login and verified by KDC
Step 2: User get key from KDC
Step 3: Generate Trapdoor for search (Trapdoor includes query words and number of ranked document)
Step 4 : Get result set

8. Conclusion

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 Multikeyword 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.

Narendra Joshi was born in Nasik, Maharashtra on 1st March 1979.He Received B.Tech Degree (Computer Technology) from Mumbai University in 2004.Recently I am completed M.E. (CSE) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.