ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Encrypted Multi-Keyword Search with Data Chunking and TPA Verification for Secured Cloud Storage

[1]D.Sterlin Rani,[2]R.Annie Jeyakumari,[3]L.Annamari Thangam,[4]B.Fathima Jameera

[1]Assistant Professor, [2,3,4]U.G Students Dept of CSE ,Kings Engineering College,
, Dept of CSE, Kings Engineering College.

Abstract- ***Using Cloud Computing, individuals can store their data on remote servers and allow data access to public users through the cloud servers. As the outsourced data are likely to contain sensitive privacy information, they are typically encrypted before uploaded to the cloud. This, however, significantly limits the usability of outsourced data due to the difficulty of searching over the encrypted data. In this paper ,we address this issue by developing the multi-keyword search schemes over encrypted cloud data. The data is encrypted, splitted and stored in separate servers .To achieve the security and data confidentiality, we introduce a new technique called TPA verification with data chunking and Replica implementation for secured cloud storage.***

Index terms :Cloud computing ,data chunking, Multi-keyword, Replica, TPA.

1.INTRODUCTION

The cloud computing treats computing as a utility and leases out the computing and storage capacities to the public individuals. In such a framework, the individual can remotely store her data on the cloud server, namely data outsourcing, and then make the cloud data open for public access through the cloud server. This represents a more scalable, low- cost and stable way for public data access because of the scalability and high efficiency of cloud servers, and therefore is favorable to small enterprises .Note that the outsourced data may contain sensitive privacy information. It is often necessary to encrypt the private data before transmitting the data to the cloud servers. The data encryption, however, would significantly lower the usability of data due to the difficulty of searching over the encrypted data. Simply encrypting the data may still cause other security concerns. For instance, Google Search uses SSL (Secure Sockets Layer) to encrypt the connection between search user and Google server when private data, such as documents and emails, appear in the search results .However, if the search user clicks into another website from the search results page, that website may be able to identify the search terms that the user has used .On addressing above issues, the searchable encryption has been recently developed as a fundamental approach to enable searching over encrypted cloud data, which proceeds the following operations. Firstly, the data owner needs to generate several keywords according to the outsourced data. These keywords are then encrypted and stored at the cloud server. When a search user needs to access the outsourced data, it can select some relevant keywords and send the cipher text of the selected keywords to the cloud server. The cloud server then uses the cipher text to match the outsourced encrypted keywords, and lastly returns the matching results to the search user. To achieve the similar search efficiency and precision over encrypted data as that of plaintext keyword search, an extensive body of research has been developed in literature. Unfortunately, due to using order- preserving encryption (OPE) to achieve the ranking property, the proposed scheme cannot achieve unlink ability of trapdoor. Later, they propose a multi-keyword text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query. After, they propose a multi- keyword top-k retrieval scheme which uses fully homomorphic encryption to encrypt the index/trapdoor and guarantees high security. Cao et propose a multi-keyword ranked search (MRSE),which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords. Although many search functionalities have been developed in previous literature towards precise and efficient searchable encryption, it is still difficult for searchable encryption to achieve the same user experience as that of the plain text search, like Google search. This mainly attributes to following two issues. Firstly, query with user preferences is very popular in the plain text search. It enables personalized search and can more accurately represent user's requirements, but has not been thoroughly studied and supported in the encrypted data domain. Secondly, to further improve the user's experience on searching, an important and fundamental function is to enable the multi-keyword search with the comprehensive logic operations, i.e., the "AND", "OR" and "NO" operations of keywords

This is fundamental for search users to prune the searching space and quickly identify the desired data. Later, they propose the coordinate matching search scheme(MRSE) which can be regarded as a searchable encryption scheme with “OR” operation. Zhang et al propose a conjunctive keyword search scheme which can be regarded as a searchable encryption scheme with “AND” operation with the returned documents matching all keywords. However, most existing proposals can only enable search with single logic operation, rather than the mixture of multiple logic operations on keywords, which motivates our work.

In this work, we address above two issues by developing Multi-keyword Search (MS) schemes over encrypted cloud data. Our original contributions can be summarized in three aspects as follows:

- We introduce the relevance scores and the preference factors of keywords for searchable encryption. The relevance scores of keywords can enable more precise returned results, and the preference factors of keywords represent the importance of keywords in the search keyword set specified by search users and correspondingly enables personalized search to cater to specific user preferences. It thus further improves the search functionalities and user experience.
- We realize the “AND”, “OR” and “NO” operations in the multi-keyword search for searchable encryption. Compared with schemes in the proposed scheme can achieve more comprehensive functionality and lower query complexity.
- We employ the classified sub-dictionaries technique To enhance the efficiency of the above two schemes. Extensive experiments demonstrate that the enhanced scheme scan achieve better efficiency in terms of index building, trapdoor generating and query in the comparison with schemes.

2. SYSTEM MODEL,THREAT MODEL AND SECURITY REQUIREMENTS

2.1 System Model

We consider a system consist of three entities.

Data Owner:The data owner outsources her data tothe cloud for convenient and reliable data access to thecorresponding search users. To protect the data privacy,the data owner encrypts the original data throughsymmetric encryption. To improve the search efficiency,the data ownegenerates some keywords for eachoutsourced document. The corresponding index is then created according to the keywords and a secret key. After that, the data owner sends the encrypted documents and the corresponding indexes to the cloud, and sends the symmetric key and secret key to search users.

Cloud server: The cloud server is an intermediate entity which stores the encrypted documents and corresponding indexes that are received from the data owner, and provides data access and search services to search users. When a search user sends a keyword trapdoor to the cloud server, it would return a collection of matching documents based on certain operations.

Search user: A search user queries the outsourced documents from the cloud server with following three steps. First, the search user receives both the secret key and symmetric key from the data owner. Second, according to the search keywords, the search user uses the secret key to generate trapdoor and sends it to the cloud server. Last, she receives the matching document collection from the cloud server and decrypts them with the symmetric key.

2.2 THREATMODEL AND SECURITY REQUIREMENTS

In our threat model, the cloud server is assumed to be “honest-but-curious”, which is the same as most related works on secure cloud data search. Specifically, the cloud server honestly follows the designated protocol specification. However, the cloud server could be “curious” to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information. We consider two threat models depending on the information available to the cloud server.

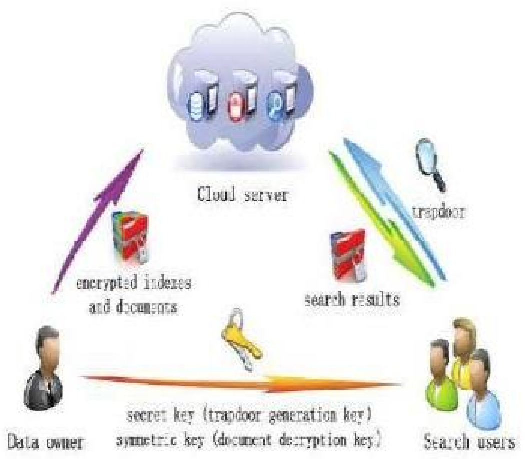

Figure 2.1 System Architecture

We assume search users are trusted entities, and they share the same symmetric key and secret key. Search users have pre-existing mutual trust with the data owner. For ease of illustration, we do not consider the secure distribution of the symmetric key and the secret key between the data owner and search users; it can be achieved through regular authentication and secure channel establishment protocols based on the prior security context shared between search users and the data owner [18]. In addition, to make our presentations more focused, we do not consider following issues, including the access control problem on managing decryption capabilities given to users and the data collection's updating problem on inserting new documents, updating existing documents, and deleting existing documents, are separated issues.

Based on the above threat model, we define the security requirements as follows:

• ***Confidentiality of documents:*** The outsourced documents provided by the data owner are stored in the cloud server. If they match the search keywords, they are sent to the search user. Due to the privacy of documents, they should not be identifiable except by the data owner and the authorized search users.

• ***Privacy protection of index and trapdoor:*** As discussed in Section 2.1, the index and the trapdoor are created based on the documents' keywords and the search keywords, respectively. If the cloud server identifies the content of index or trapdoor, and further deduces any association between keywords and encrypted documents, it may learn the major subject of a document, even the content of a short document. Therefore, the content of index andtrapdoor cannot be identified by the cloud server.

• ***Unlinkability of trapdoor:*** The documents stored in the cloud server may be searched many times. The cloud server should not be able to learn any keyword information according to the trapdoors, e.g., to determine two trapdoors which are originated from the same keywords. Otherwise, the cloud server can deduce relationship of trapdoors, and threaten to the privacy of keywords. Hence the trapdoor generation function should be randomized, rather than deterministic. Even in case that two search keyword sets are the same, the trapdoors should be different.

3.PROPOSED SCHEMES

We proposed a new technique called TPA verification with Data Chunking and Replica implementation for data confidentiality and security.

3.1 *Index building:* The data owner firstly utilizes symmetric encryption algorithm to encrypt the document collection with the symmetric key. Then the data owner generates number of keywords indicates whether the keywords contained in the encrypted document. The keywords are generated using Stemming algorithm. It removes the stop words in the document and extracts the main keyword. The keywords are listed in the index as the pointers to the documents in the server.

3.2 *Data Chunking:* The data is splitted and stored in the separate cloud server. When user searches the keywords in the documents, the main server grouped all the documents from the sub server and displayed to the user according to the user query.

3.3 *Replica Server:* We'll maintain the separate Replica Cloud server. If suppose the data in the data server was lost, then the Main Cloud server will contact the Replica Cloud server and get the data from the Replica Cloud Server. By using this concept, we can get the data if any data loss occurs.

3.4 *TPA Verification:* The data will be given to the Trusted Party Auditor. The Trusted Party Auditor will generate the signature using change and response method. The data will be audited in this module, if any changes occurs it will provide the intimation regarding the changes.

4. SECURITY ANALYSIS

In this section, we analyze the main security properties of the proposed schemes. In particular, our analysis focuses on how the proposed schemes can achieve confidentiality of documents, privacy protection of index and trapdoor, and unlinkability of trapdoor. Other security features are not the focus of our concern.

4.1 Confidentiality of Documents

In our schemes, the outsourced documents are encrypted by the traditional symmetric encryption algorithm (e.g., AES). In addition, the secret key is generated by the data owner and sent to the search user through a secure channel. Since the AES encryption algorithm is secure, any entity cannot recover the encrypted documents without the secret key. Therefore, the confidentiality of encrypted documents can be achieved.

4.2 Privacy Protection of Index and Trapdoor

Using multi-keyword search through index, the user query does not reveal to the server.

The server will search the encrypted keyword so it will find the original content of search results. Thus, the content of index and trapdoor cannot be identified. Therefore, privacy protection of index and trapdoor can be achieved.

4.3 Unlinkability of Trapdoor

To protect the security of search, the unlinkability of trapdoor should be achieved. Although the cloud server cannot directly recover the keywords, the linkability of trapdoor may cause leakage of privacy, e.g., the same keyword set may be searched many times, if the trapdoor generation function is deterministic, even though the cloud server cannot decrypt the trapdoors, it can deduce the relationship of keywords. We consider whether the trapdoor can be linked to the keywords. We prove our schemes can achieve the unlinkability of trapdoor in a strong threat model.

5. RELATED WORK

There are mainly two types of searchable encryption in literature, Searchable Public-key Encryption (SPE) and Searchable Symmetric Encryption (SSE).

5.1 SPE

SPE supports single keyword search on encrypted data but the computation overhead is heavy. In the framework of SPE, conjunctive, subset, and range queries on encrypted data. Hwang et al propose a conjunctive keyword scheme which supports multi-keyword search. Zhang et al propose an efficient public key encryption with conjunctive subset keywords search. However, these conjunctive keywords schemes can only return the results which match all the keywords simultaneously, and cannot rank the returned results. Qin et al propose a ranked query scheme which uses a mask matrix to achieve cost-effectiveness. Yu et al propose a multi-keyword top-k retrieval scheme with fully homomorphic encryption, which can return ranked results and achieve high security. In general, although SPE allows more expressive queries than SSE, it is less efficient, and therefore we adopt SPE in the work.

5.2 SSE

The concept of SSE is first developed by Song et al. Wang et al develop the ranked keyword search scheme, which considers the relevance score of a keyword. However, the above schemes cannot efficiently support multi-keyword search which is widely used to provide the better experience to the search user. Later, Sun et al propose a multi-keyword search scheme which considers the relevance scores of keywords, and it can achieve efficient query by utilizing the multidimensional tree technique. A widely adopted multi-keyword search approach is multi-keyword ranked search (MRSE). This approach can return the ranked results of searching according to the number of matching keywords. Li et al utilize the relevance score and k-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. Within this framework, they leverage an efficient index to further improve the search efficiency, and adopt the blind storage system to conceal access pattern of the search user. Li et al also propose an authorized and ranked multi- keyword search scheme (ARMS) over encrypted cloud data by leveraging the cipher text policy attribute-based encryption (CP-ABE) and SSE techniques. Security analysis demonstrates that the proposed ARMS scheme can achieve collusion resistance. In this paper, we propose FMS (CS) schemes which not only support multi-keyword search over encrypted data, but also achieve the fine-grained keyword search with the function to investigate the relevance scores and the preference factors of keywords and, more importantly, the logical rule of keywords. In addition, with the classified sub-dictionaries, our proposal is efficient in terms of index building, trapdoor generating and query.

6. CONCLUSION

In this paper, we have investigated on the multi-Keyword Search (MS) issue over encrypted cloud data, and proposed two MS schemes. The MS I includes both the relevance scores and the preference factors of keywords to enhance more precise search and better users' experience, respectively. The MS II achieves secure and efficient search with practical functionality, i.e., "AND", "OR" and "NO" operations of keywords. Furthermore, we have proposed the enhanced schemes supporting classified sub- dictionaries (SCS) to improve efficiency. For the future work, we intend to further extend the proposal to consider the extensibility of the file set and the multi-user cloud environments. Towards this direction, we have made some preliminary results on the extensibility and the multiuser cloud environments. Another interesting topic is to develop the highly scalable searchable encryption to enable efficient search on large practical databases.

REFERENCES

1) H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng,"An smdpbased service model for interdomain resource allocation in mobile cloud networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5,pp. 2222–2232, 2012.

2) M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspotlocatingattack in wireless sensorn networks," *IEEE Transactions on Parallel and Distributed Systems*, vol.23, no. 10, pp. 1805–1818, 2012.

3) Christo Ananth, S.Shafiqa Shalaysha, M.Vaishnavi, J.Sasi Rabiyathul Sabena, A.P.L.Sangeetha, M.Santhi, "Realtime Monitoring Of Cardiac Patients At Distance Using Tarang Communication", International Journal of Innovative Research in Engineering & Science (IJIRES), Volume 9, Issue 3,September 2014,pp-15-20

4) T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L.Zhang, "Privacypreserving data aggregation without secure channel: multivariate polynomial evaluation," in*Proceedings of INFOCOM*. IEEE, 2013, pp. 2634–2642.

5) Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Securedynamic searchable symmetric encryption withconstant document update cost," in *Proceedings of GLOBCOM*. IEEE, 2014, to appear.

6) N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 222–233, 2014.