Searching Encrypted Data on Cloud

Muhammad Sajid Khan[1], Chengliang Wang[2] , Ayesha Kulsoom[3], Zabeeh Ullah[4]

[1, 2] School of Software Engineering, Chongqing University, Shapingba, Chongqing , 400044, PR.CHINA
[3] School of Computer Science , Chongqing University, Shapingba, Chongqing , 400044, PR.CHINA
[4] Department of Computer Science , University of Ballarat , Victoria, 3353 , Australia

Abstract

Encryption is an optimistic way to preserve the secrecy of the outsource data. On other hand, performing search operation on encrypted data is a complicated job. A lot of encryption techniques have been suggested by researchers to avoid improper usage of unstructured data on the cloud. But still by using the existing searchable encryption techniques, searching data on cloud server becomes difficult. In this paper, Data Encryption on cloud as well as corresponding security issues has been addressed. The proposed method incorporates two main phases: indexing and searching. Trapdoor and codeword are the two security parameters applicable in this technique. Simulated results demonstrate that it provides fast and efficient ranking sentence search for unstructured data in original documents on cloud server. The proposed technique reduces the overhead of decryption thereby minimizing the search time to a considerable extent.

Keywords: *SEDC, Ranked Encryption Search, Indexing unstructured data, searchable Encryption.*

1. Introduction

Data owners are encouraged to handover their traditional database management system like spatio-temporal DBMS system and Raster Data Management are shifted from a local server to cloud for its easy access, high availability, infinite scalability, cost saving and great performance . [1][2].NIST is the supporting community for cloud computing that gives the definition of cloud computing as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[3].

The huge amount of data storage on local area network (LAN) is highly expensive. To reduce the capacity of storage, cloud server is a viable option as it is very cheap, possessing unlimited data storage and is easily accessible via Internet ubiquitously. Traditional hosting providers IBM and HP are transferring their hosting solution to the cloud [4][5].

Security and privacy of data are major issues of cloud computing as all the information and data is accessible by a service provider. For this purpose, CSA, ENISA and NIST offer some concerns for clouds to secure data from hackers [6].The vast usage of cloud introduces different level of security threats to the cloud users. To minimize security threats, users are using different encryption techniques to store their confidential data on cloud. Data encryption allows the authorized and authentic user to access the data in order to keep information in form of codeword, not easy to be disclosed by entrusted users [7]. As per David Simms survey 2013, 95% people were reported to use cloud for storage [8]. The main reason is management flexibility, compute on demand, online sharing and improvement on customer value added services.

Existing encryption techniques for encrypted data on cloud are considerably complicated and time consuming. Therefore, a lot of work has been done by researchers to meet the search criteria but still these techniques are not much efficient and accurate. The traditional search mechanism for unstructured data is the linear search in which user has to decrypt documents for searching. Documents decryption requires huge amount of time for searching [9].Sentence search on encrypted cloud data is impossible until now. This research presents searching technique for encrypted unstructured data. The technique proposed in this paper is secure, efficient and accurate.

The rest of paper is organized as follows: Section 2. provides the research objective. Section 3. Presents the Methodology of proposed technique section 4. is providing experimental work and analysis and Section 5. draws the conclusion and discusses future work.

2. Objective

The aim of this research is to provide a sentence search mechanism which does not decrypt the documents to perform search operation. Trapdoor and codeword are two security parameters that have been achieved. Data passed

through these two security levels cannot be hacked or predicted by cloud.

3. Methodology

First of all Indexing is done on all the documents of data holders. After uploading documents on the server, the Encrypted index is created. Secondly Bloom filters (BFAH) are used to make the codewords more secure and confidential.

Finally after choosing the data from indexes, encryption and matches with searched sentence in the selected document's data and algorithm complete its cycle. Cloud server is unfamiliar with the content of those documents which are in the encrypted form. Cloud server cannot learn any information about the searched sentences during searching. Searched sentences are sent to cloud after applying security parameters so that server or user remains unaware of searching.

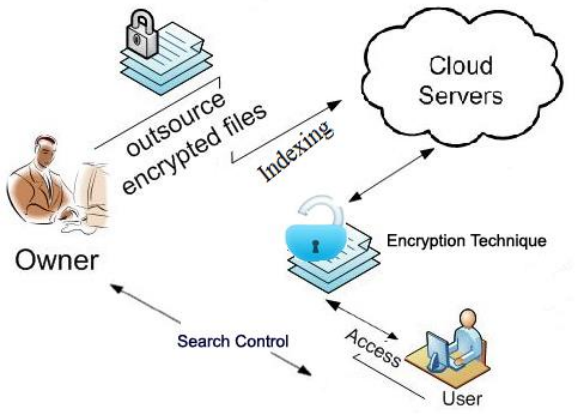

Fig. 1 Pictorial representation of searching on Encrypted Data

3.1 Encryption on Unstructured Data

Firstly all documents given by data owner are divided into set of words. Then key generation algorithm is used to create Master Key, which avoids the unusable words to increase the efficiency and make the indexing fast. In the proposed technique 512-bit hash (SHA-512) of the input password is used to create the master private key, M_k, which is then divided into thirty two 2-byte keys (r =32).

$$M_k = SHA512(password) \quad (1)$$

M_k is then split into a set of sixteen keys. These sixteen split keys are used in trapdoor generation step to secure the trapdoors [10].

$$M_k : k_1, k_2, k_3, k_3, k_4, k_5, k_6, k_7, k_8 \quad (2)$$

Trapdoor generation is the first security parameter to secure the keywords from document. Hash algorithm can be used by data owner for generating Trapdoor. The trapdoor obtained from above technique is represented in formula as shown below.

$$Trapdoor: f(k_1 + word), f(k_2 + word) \ldots f(k_8 + word) \quad (3)$$

The generated codewords step is second security parameter to enhance the security of keyword's trapdoor generated in the above step. The trapdoor is sent to bloom filter (BFAH) to get codewords. Codewords are bit location obtained for each trapdoor when the crc32 hash algorithm is applied on it. The codewords generation formula is as follows [12].

$$codeword = 5\ hash\ algos\ (Trapdoor) => 5\ bit\ location \quad (4)$$

Finally code word is added to Bloom filter (BFAH) in order to achieve the index of document. The index created through this stage is used as the index of documents. Practical results of the above method are depicted below in Table 1.

EXPERIMENTAL RESULTS		
WORD	**TRAPDOOR**	**CODEWORD**
Introduction	aQLix3or1GPwlUZqWJm65RaiGOXHz9Tfo5+KY+gvd bEiWiANBCAaHgoywQKS=	67497393932618 56899922281
Method	aQLix3or1GPwlUZqWJm65RaiGOXHz9Tfo5+KY+gvd bEiWiANBCAaHgoywQKS=	30121297822519 66995220043
Result	aQLix3or1GPwlUZqWJm65RaiGOXHz9Tfo5+KY+gvd bEiWiANBCAaHgoywQKS=	86511928562814 520562317
Discussion	9gHUliEXr1H9KBSLDanrrggHcABC3y1L7bz/32BBu9a mjqhr8B6b7RBTlBooZ7TkJW/bJhBGDxWgXmhAHA GXC5NL	16863643634878 03489216941
References	9gHUliEXr1H9KBSLDanrrggHcABC3y1L7bz/32BBu9a mjqhr8B6b7RBTlBooZ7TkJW/bJhBGDxWgXmhAHA GXC5NL	43276691162396 91226546612
Determine	cwBl+c4Zr1GYRvbp6twdTYkRf58t1d2guZeCk/6+04J Z9CJBq/8zU92OJimBGAOHcS2H3WMhlWWrZVPkrrl +7fbOgo4hOSl6eMkZ	40245291664004 61715536485

Table 1: Generation of Trapdoor and Codeword from the given input

All keywords have been converted to codewords and codeword will have the same location in document which its original keyword has. Now encrypted document index is uploaded. As all the data is in encrypted form, the cloud is unable to identify any document content or indexes. This makes the index and documents secure and private.

3.2 Searching on Encrypted data

Searching is very common and well-known operation. User likes to find out any documents stored on the cloud and also does not like to wait longer for retrieving the results back.

For search after the input sentence the processes up to Generate Codeword are same as describe in section 3.1. This illustrates that security levels have been achieved so that sentence or keywords searched should not be predicted by cloud server or malicious users.

Fig. 2 Flow Chart of searching encrypted data process

Fig.2 represents the steps after Generating Codeword, related documents are selected form the document indexes in Select Query step. The selected documents are then arranged in Arrange Selection step. All possible combinations of the arranged data are found in Find Combinations Step. The documents are ranked by given formula.

$$Rank = 0.4 \times SD - 0.6 \times TF \quad (5)$$

To fetch the matching documents in ranked order the most relevant document lies at top.

4. Experimental Results and Analysis

In this section, we present the experimental evaluation of the proposed technique. A set of 150 documents are taken for experimental evaluation .The results are calculated on basis of Indexing time, Searching time, Indexing time of ranked keyword search and searching time of ranked keyword search. The comparison of time taken to preprocess the document and to generate the encrypted indexes in proposed searching technique and ranked keyword search technique is shown in Fig.4. It can be seen that indexing time of proposed searching technique is more than ranked search due to more database queries for insertion of codewords to the database. The time of indexing in proposed searching technique can be reduced with minimizing the database queries.

.

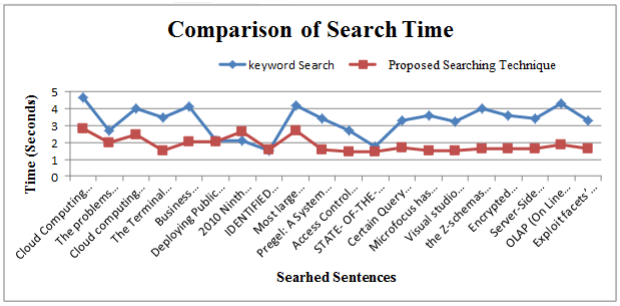

Fig. 3 Graphical representation of comparison of Indexing time

Search time comparison of proposed searching technique and ranked keyword search technique is shown in Fig. 5 It can be seen that proposed searching technique takes less time for sentence searching than ranked keywords search technique. The same data set is used to calculate the searching time of proposed searching technique and ranked keyword search. Whereas the searching time of proposed searching technique is less than ranked keyword search as the algorithm of searching in the proposed technique is efficient, faster and accurate than ranked keyword search.

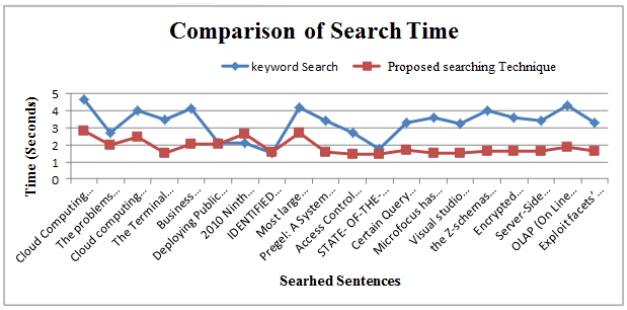

Fig. 4 Graphical representation of comparison of Search time.

5. Conclusion and Future Work

This paper concentrates on safe and secure searching of unstructured data in cloud. To accommodate secure encryption on cloud, a novel method of searching encrypted data on cloud is proposed. Thus user does not need to hesitate about security issues and shall be willing to place more data on cloud. In this technique trapdoor and codeword are two security levels for both indexing and searching.

Bloom Filter (BFAH) has been used to make the codeword more secure and confidential. The evaluation of experimental results indicates that searching encrypted data on Cloud proposed is secure and protective from hackers. The method is proposed for cloud environment where a large amount of unstructured data is stored in encrypted form.

It is not restricted to cloud environment only. This can be used on individual server within an organization for data security. As a future work, it can be extended to enable case sensitivity and sub match searches. Ranking formula can be improved and other parameters need to be considered like document length and number of index words.

Acknowledgment

The authors are grateful to Professor Chenglian Wang of Chongqing University, China for his invaluable support and contribution in this paper. The authors are also grateful to the reviewers who, through numerous comments made this paper more readable.

REFERENCES

[1] Jerry Archer. "Security guidance for critical areas of focus in cloud computing," December 2009.

[2] Minqi Zhou. "Services in the Cloud Computing Era" Published by IEEE, Beijing China 2010.

[3] Tyrone Grandson, E. Michael Maximilien, Sean S. E. Thorpe and Alfredo Alba. "Towards a Formal Definition of a Computing Cloud". Published by IEEE 6th World Congress on Services 2010.

[4] Platform-As-A-Service Is Here: How To Sift Through The Options, Forrester, April 2009-06-18

[5] Market Trends: Application Development, Worldwide, 2008-2013, Gartner, January 2009

[6] SeongHan Shin, Kazukuni Kobara and Hideki Imai. "A Secure Public Cloud Storage System 6th International Conference on Internet Technology and Secured Transactions", Abu Dhabi, United Arab Emirates 11-14 December 2011

[7] Cong Wang, Qian Wang, and Kui Ren."Towards Secure and Effective Utilization over Encrypted Cloud Data".31st International Conference on Distributed Computing Systems Workshops 2011.

[8] David Simms and Solange Ghernaouti. "Structured and unstructured data in the Cloud" by conference 2013 27th International Conference on Advanced Information Networking and Applications Workshops

[9] Yao Hanbing, Xiang Dong, Peng Dewei, Huang Jing. An "Approach for Searching on Encrypted Data Based on Bloom Filter "published by 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science

[10] Shay Gueron. "New Generations SHA-512/256".Published by Eighth International Conference on Information Technology, USA 2011

[11] Mahmood Ahmadi and Stephan Wong. "A Memory-optimized Bloom Filter using An Additional Hashing Function".

[12] Yao Hanbing, Xiang Dong, Peng Dewei, Huang Jing "An Approach for Searching on Encrypted Data Based on Bloom Filter".