An Efficient Verification of Data Integrity and Keyword Search for Multi Data Owners by Preserving Privacy in Cloud

Jincy Easow[#1], Prof. Jisha P Abraham[*2]
[#1]*Department of Computer Science & Engineering*
Mar Athanasius College of Engineering, Kothamangalam, Kerala, India
[#2]*Associate Professor, Department of Computer Science & Engineering*
Mar Athanasius College of Engineering, Kothamangalam, Kerala, India

Abstract

Cloud is an emerging technology which is widely used today for various applications. By going day by day, the size of the data to be processed, stored, and manipulated also increases frequently. Such giant amount of data from various areas are necessary to store in a perfectly secured platform. It is impractical to maintain such huge amount of data in a single machine or hard disk. In such situation, cloud is the platform in which can store any amount of data. Major advantage of using cloud is that the stored data can be access by the users at anytime, anywhere and from any device. But security of the residing data is the major problem with the cloud. Due to this reason, most of the users are reluctant to outsource their documents to the cloud in spite of the benefits. This problem can be avoided by encrypt the data before outsource it to the cloud. By doing so, the security of the documents can be improved to a level best. While doing so, searching over the encrypted file is another problem faced by the users. The proposed system introduced a novel approach to search data over encrypted files without affecting the privacy and security of the file to be outsourced by multiple data owners. For the effective implementation of encrypted searching of data by using paillier homomorphic encryption algorithm, which provides the ability to make computations over the encrypted data without decryption.

Keywords

Cloud Computing, Paillier Homomorphic Encryption.

I. INTRODUCTION

The cloud is a term referring to accessing computer, information technology (IT), and software applications through a network connection, often by accessing data centre using wide area networking (WAN) or Internet connectivity. Cloud computing has many advantages. It's often faster to provision the service, and in many case you can gain access to it instantly. Remote users can access cloud resources from wherever they have a connection, rather than being limited by physical geography. It is often divided into three categories: private, public, and hybrid, referring to who has access to the services or infrastructure. Public-cloud services are made available to anybody that wants to purchase or lease the services. Private-cloud services are built by enterprises for use by their employees and partners only. Hybrid-cloud services combine the two.

The concept has grown over time, to include just about any service that can be provisioned quickly via a network connection, often using the Web or mobile applications. For example, a customer could order up Web hosting in the cloud (Amazon or Rackspace), or consume digital media services such as movies and music on demand in the cloud (Apple iTunes, Amazon, and Netflix), storage (Dropbox or Google Drive), email (gMail), or even contract for housing and transportation services (AirBnB or Uber). Business software such as Microsoft Outlook, once predominately reached on local networks or computer, is migrating to applications accessible by the cloud.

There are four entities are involved in this scheme data owner, administration server, cloud server and data user. Data owner wants to upload the files containing sensitive data into the cloud. Due to privacy concerns, files to be encrypted before upload it into the cloud. With the approval from the administration server, the file is to be split and encrypted each block using standard symmetric encryption algorithm. Before encrypt the files, keywords are extracted from the file and also encrypted and forwarded them to store in the cloud server. These encrypted keyword index made the searching operation easier. Extracted keywords are encrypted homomorphically. [1]When the data user wants to retrieve the files from the cloud of his interest. Firstly user send a request to the administration server. The request is in the form of multiple keywords that are given to the administration server. [2][3]Upon receiving the request from the data user administration server authenticate the user. If the

user is authorized, then administration server encrypt the received keywords homomorphically and send to the cloud server. Otherwise discard the request. The searching operation to retrieve the files are done by the cloud server. If the server found a matching between the encrypted keywords stored with the requested keywords, the server obtain the file id and retrieve it in a combined form and forward to the requested data user. But the received file which are in an encrypted form. In order to access the file in the readable form only by through the secret key used by the data owner used to encrypt it. So the user send a request to the data owner to access the key. If the user received the key from the owner, then user can download the file and use it.

II. PROPOSED SYSTEM

Cloud is the most promising medium to store and retrieve large amount of data, which can be accessed anywhere and at any time and from any device. Here, it benefits this property of cloud, while the security of the data in the cloud is the problem with this. In order to increase the privacy of the data in the cloud, which is to be stored in an encrypted form. The data can be encrypted using any standard symmetric encryption algorithm like AES or DES. In this work AES encryption algorithm is used to encrypt the data files by the data owners. By doing so, the searching over the encrypted data files is a tedious operation. The previous work was to perform the searching operation over the unencrypted data which is a simple task while the security become overruled. To overcome the problem with searching over the encrypted data, here homomorphic encryption algorithm is used to encrypt the keywords to be stored and to be searched. The Homomorphic encryption algorithm is used because of the reason that it helps to perform operations over the encrypted data without decrypted. Hence it improves the security of the files stored in the cloud and also make the searching operation easier. Figure 1 shows the system architecture.

The proposed system consist of four entities like data owner, administration server (AS), cloud server and the data user. Data owner is the person who upload files to the cloud. Administration server which is an intermediate server between the owner/user and the cloud server. Uploading and requesting files are only through the administration server. Data user made request to access and update the files to the cloud server. Cloud server ultimately store the files and perform the searching operation and provide the matched requested files to the data user. The different modules involved in this scheme are given below:

1. Processing Module

Four different entities are involved in cloud computing environment: Cloud server, Data owner, Administration server and Data user. Data owner at first has to register with the cloud. After that the user has to wait until the administration server approves him. The data owner can upload the documents. Third-party data storage and retrieval services are hosted by the cloud server. As the uploaded data may contain sensitive, data is needed to be encrypted before outsourcing. Data owner send the file to the AS. After getting the approval from the AS, the file is to be split into multiple blocks and encrypted each block using standard symmetric encryption algorithm AES, forward the encrypted blocks to the cloud server.

2. Keyword Extraction Module

Before split the files into blocks, the AS extract keywords from the file using rapid keyword extraction algorithm[11]. To effectively search documents, it is necessary to build a searchable secure index for all the documents being uploaded into the cloud. In this algorithm, extraction process is done by four phases.

i. Preprocessing
ii. Word Co-occurrence Graph
iii. Calculate word score
iv. Keyword Extraction

Preprocessing

During preprocessing phase, which uses the stop words and phrase delimiters to partition the document text into candidate keywords. Candidate keywords are sequence of content words in the text. It is done by parsing the text into a set of candidate keywords. Firstly, text is split into an array of words by the word delimiters. Array is then split into sequence of contiguous words at phrase delimiters and stop word positions. Words within a sequence are assigned the same position in the text and together are considered as candidate keyword.

Word Co-occurrence Graph

After the preprocessing stage, the word co-occurrence graph is drawn using the candidate keywords from the previous stage. The X and Y axis represents the candidate keywords. From the graph, the frequency of each word and how it is related to other words are identified.

Calculate Word Score

From the word co-occurrence graph, frequency and the degree of each word is obtained. Frequency of the word is the number of times that particular keyword occur in the document. Word degree is the words that occur often and in longer candidate keywords. Score value of each candidate keyword is computed by taking the ratio of the degree and the frequency of the words. If the candidate keyword is a longer keyword, then its score value is the sum of its member word scores. By this method, score value of all the candidate keywords are calculated.

Keyword Extraction

Among the score value of all the candidate keywords, one-third of the words in the graph having highest score is ultimately taken as the actual keywords which are used to create the searchable index.

3. Encryption Module

The documents before uploading into the global space are encrypted using any encryption scheme, whereas for the secure index homomorphic encryption is applied. The AES symmetric encryption algorithm is used to encrypt the document to be uploaded. Homomorphic encryption maps the operations in cipher text and plain text domain. The additive and multiplicative property of homomorphic encryption is utilized here. Thus Paillier homomorphic encryption is applied to the keyword index. By using Paillier homomorphic encryption, computations can be performed over encrypted data. Paillier homomorphic encryption is work over pure integers. Hence the keywords need to be encrypted is firstly converted into integers. Paillier homomorphic encryption is done as follows:

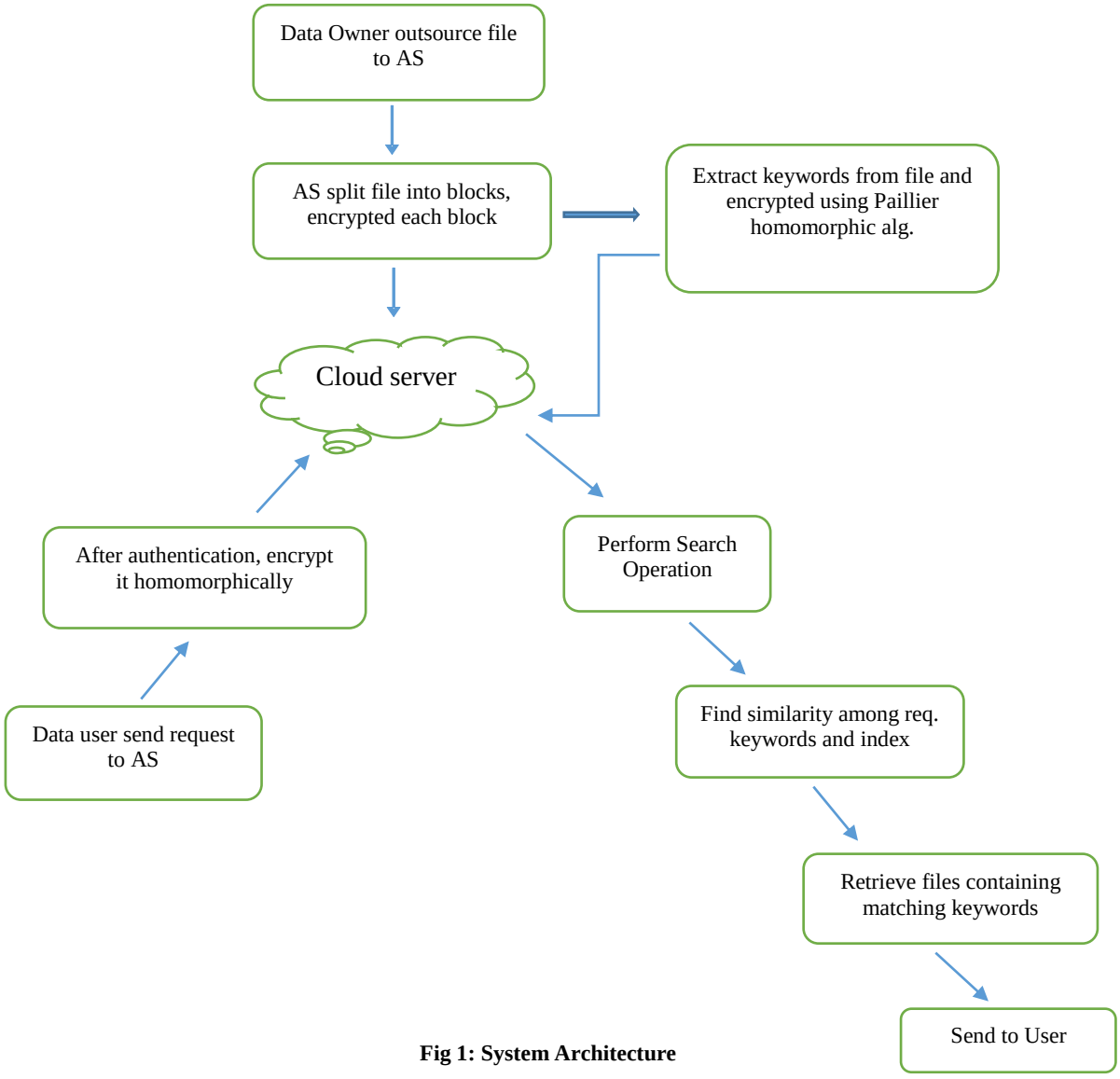

Fig 1: System Architecture

1) Key Generation:

Choose 2 large prime numbers p, q
Compute n = pq, n^2
Select random integer g, g = (1+n)
Carmichael's Function, λ = lcm(p-1,q-1)
$\varphi(n) = (p-1, q-1)$

$\mu = \varphi(n)^{-1} \bmod n^2$
Public Key: (n, g)
Private key: (p, q, λ)

2) Encryption:

Plaintext m, m<n

Find a random r

Ciphertext c = $g^m . r^n \ mod \ n^2$

4. Searching Operation

Search operation is done by the cloud server. Homomorphically encrypted keywords are stored in the cloud server. When the data user give request to the administration server by multiple keywords. After authentication, authentication server encrypt the requested keywords homomorphically and forward them to the cloud server. In the cloud server, both the stored keywords and the requested keywords are in the integer form by applying paillier homomorphic encryption on them. Then the searching can be performed by subtract both integer values ie, stored keyword value and requested keyword value[12]. Search operation can be done between the ciphers as follows:

Difference, d = $\frac{E_n[a, r]\, E_n[b, r]^{-1} - 1 \ mod \ n^2}{n}$ mod n

If the subtraction result is zero, then the ciphers are identical otherwise they are not. Through this method, identical keywords can be found. From the matched keywords, the corresponding file id can be obtained. If the matching file is found, then combine the splitted blocks of that particular file and send to the data user. The corresponding encrypted files are send to the requested data users by the cloud server. While the secret key used for encrypting file is required by the data user to download the contents of the file. For that purpose, the data user send request which is forwarded to both AS and the data owner to get the permission to download the received file. Data owner provide permission in the form of secret key and AS again cross check the timestamps and the requested nature of the data user. If it valid, then give the approval to download the file. The file can be downloaded only after get the permission from both the owner and the AS.

The downloaded file can be updated by the data user with the permission of the corresponding data owner. For that the user send an edit request to the corresponding data owner. If he approves, user can make changes in the file. Changes made by the user can be viewed by the owner. If the owner disagree with the changes on the contents of the file, then the owner can block the user from further operation to be done. Hence, the user no longer remains a data user in that cloud environment.

Result and Performance Evaluation

The proposed system brings much better performance than the existing systems, which perform the search operation using cosine similarity (MKSCS). In MKSCS, indexed keywords are obtained by using porter stemmer method and which is encrypted using any standard encryption algorithm during outsource to the cloud. While matching operation is perform over unencrypted keywords, instead encrypted. On searching, the encrypted keywords are decrypted and perform matching operation and find the suitable files. The performance of the proposed system is evaluated as compared with the existing MKSCS. The result of the performance evaluation is shown below:

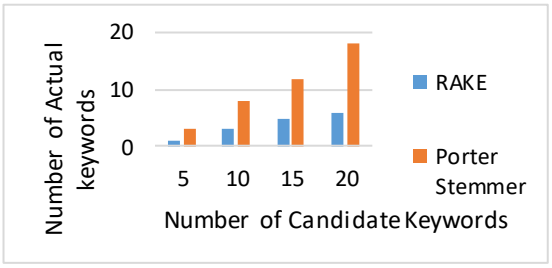

Fig 2: Number of candidate keywords Vs actual keywords

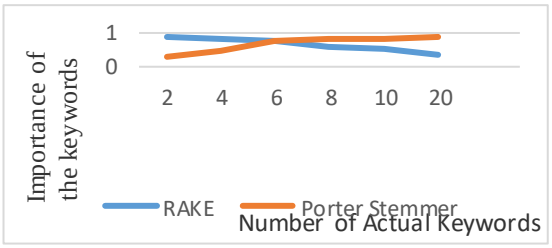

Fig 3: No. of actual keywords Vs importance

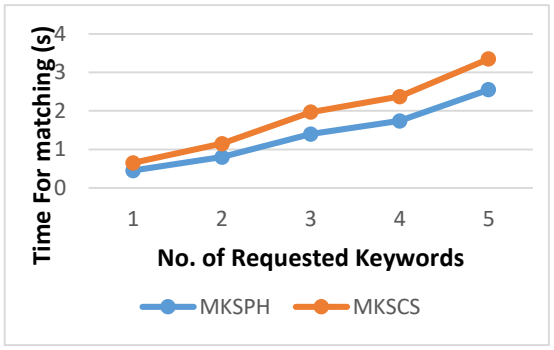

Fig 4: No. of requested keywords Vs time for matching operation

III. CONCLUSION

Now a day it become common to access global storage space over the Internet. The main problem of storing data in a trusted third party is regarding security. Even data is encrypted before outsourcing; effective secure multi-keyword search over encrypted cloud data retrieval is a big challenge. To overcome these challenges, proposed this scheme enables data owners to upload encrypted data files into global storage and allow several authorized users to perform search and retrieval over them. By

using homomorphic encryption algorithm, computation can be performed without decryption. It helps to improve the security of the data as the server doesn't know the actual data to be stored by the data owner and requested by the data user.

REFERENCES

[1] CongWang, Ning Cao, Jin Li, Kui Ren andWenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data",in Proc. IEEE Distrib. Comput. Syst., Genoa, Italy, Jun. 2010, pp. 253-262.

[2] Anuradha Meharwade, G. A. Patil, "Efficient Keyword Search Over Encrypted Cloud data", International Conference on Information Security and Privacy, Dec. 2015, vol. 28, pp.37-73.

[3] Quin Liu, Guojun Wang and Jie Wuz, "Secure and Privacy Preserving Keyword Search on Encrypted Cloud Data", ELSEVIER Journel of Network and Computer Application, March 2011

[4] Wenhai Sun, Bing Wang, and Ning Cao,"Privacy Preserving Muli-Keyword Text Search in Cloud Supporting Similarity based Ranking", ACM Symposium on Information and Computer security, May 2013.

[5] Zhang Xu, W kang, Rin Li, K Yow and C Xu, "Efficient Multi-keyword Ranked Query on Encrypted Cloud Data", IEEE Int. Conf. Parallel Distribution System,vol. 25, pp. 222{233, December 2012.

[6] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage", IEEE Trans. Comput., Feb 2013, pp. 10-12

[7] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data", in Proc. IEEE Int. Symp. Security Privacy, Nagoya, Japan, vol. 52, no. 7, pp. 1218-1226, Jan. 2000.

[8] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions" in Proc. 13th ACM Conf. Comput. Commun. Security,Oct. 2006.

[9] Ayantika Chatterjee and Indranil Sengupta, "Searching and Sorting of Fully Homomorphic Encrypted Data on Cloud", Commun. ACM, vol. 53, no. 3, pp. 97 - 105, Mar. 2010.

[10] F. Baldimtsi and O. Ohrimenko, "Sorting and searching behind the curtain," in Financial Cryptography and Data Security - 19th International Conference, FC 2015, San Juan, Puerto Rico, January 26-30, 2015, Revised Selected Papers, 2015, pp. 127-146.

[11] Stuart Rose, Dave Engel, Nick Cramer and Wendy Cowley, "Automatic keyword extraction from individual documents", Research Gate Text Mining:Applications and Theory, Oct. 2017.

[12] Tanyaporn Sridokmai and Somchai Prakanchareon, "The Homomorphic Other Property of Paillier Cryptosystem", IEEE Int. Conf. Science and Technology, Jan. 2016.