International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Encrypted Phrase Search in the Cloud Storage

G. Gautham Raj[1], K. Santhosh Kumar[2], A. Mruthyunjayam[3], P. Akhil Goud[4], G. Shivani[5]
[3]Associate Professor, [1,2,4,5] B. Tech, Department of Computer Science and Engineering, St. Martin's Engineering College, Hyderabad, Telangana, India

***Abstract:** This study has been undertaken to investigate solutions to search over encrypted documents stored on cloud servers. There are a number of searchable encryption schemes that allow secure conjunctive keyword searches over encrypted data but not able to achieve much improved storage and computational cost. In this paper, we present a phrase search technique based on Bloom filters that is faster than existing solutions, with better storage and computational cost. Our scheme can be summarized as the use of multiple n-gram Bloom filters, to provide conjunctive keyword search and phrase search.*
***Keywords:** Conjunctive keyword search, Phrase search, Encryption*.

I. INTRODUCTION

As cloud storage is being used on wide range, many have serious issues regarding security and privacy of getting confidential information over the Internet. In recent years the continuous data breaches draws attention towards the need for more secure cloud storage systems. Generally cloud providers perform the encryption but store the private keys with them instead of providing it to the data owners. In such case the cloud can take any data it required, without providing any privacy to its users. More over in case of data breaches, storage of private keys and encrypted data by the cloud provider is vulnerable. Hence, experts have been exploring techniques for secure storage on public and private clouds where a data owner maintains the private keys.

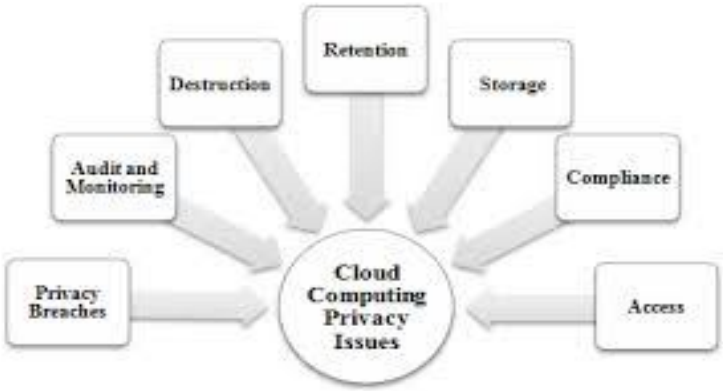

Fig: 1 Privacy issues in Cloud computing

In this paper, we present an expression search scheme which provides much faster responses than the previous solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. We also specified the updations to the scheme to reduce the storage cost and also the response time. Our approach is also the first to effectively allow Expression search to run independently without performing a conjunctive keyword search to identify candidate documents. We begin by presenting the literature survey in section 2 and system overview in section 3. We describe the expression search technique in section 4.

II. LITERATURE SURVEY

A. Encrypted Phrase Searching in the Cloud

As cloud computing is increasing in popularity, it is difficult to both maintain privacy in datasets while still providing adequate retrieval and searching procedures. This paper introduces a novel approach in the field of encrypted searching that allows both encrypted phrase searches on untrusted cloud. By storing encrypted keyword-location data along with specially truncated encrypted keyword indexes in a relational database, we are able to allow for a full range of search features in our encrypted searches, something that has never been accomplished before. Furthermore, our approach permits the encrypted corpus and index to both be stored on cloud data servers. We modify currently available open-source search engine software to complete a prototype and provide results from experiments on a large scale real-world dataset that has more than half a million documents.

We propose a unique and novel encrypted keyword searching design and phrase searches within the cloud. It allows the efficient searching and ranking of results that is a function of the number of unique keywords. We have developed a fully functional prototype that can work within the confines real-world response times on a large-scale dataset. In practice we show that the size overhead of the encrypted indexed database is a linear function of a non-encrypted index. Finally, our model is dependent on the number of documents with the query keyword set.

III. OVERVIEW

A. *Existing system*

As firms and individuals use cloud technologies, many are aware regarding security and privacy of accessing private information over the cloud. Due to the continuous data breaches the need for more secure cloud storage systems has became essential. While it is generally agreed that encryption is necessary, after performing the encryption cloud providers maintain the private keys instead of handing over it to the data owners. Which in turn leads to the insecurity to its users? The storage of private keys and encrypted data by the cloud provider is also problematic in case of data breach.

B. *Disadvantages of Existing System*

1) The ranking of search results and searching with keywords that might contain errors
2) Potential statistical attacks on the indexes are indentified

C. *Proposed system*

In this paper, we present a phrase search scheme and it is scalable, where documents can easily be removed and added to the corpus. We also describe modifications to the scheme to lower storage cost at a small cost in response time. Our framework differs from previous works, where keywords generally consist of meta-data rather than content of the files.

D. *Advantages of Proposed System*

1. The proposed system achieves a much faster response time than existing solutions
2. The proposed algorithms can easily be adapted to the scenario of an organization wishing to setup a cloud server for its employees by implementing a proxy server in place of the data owner and having the employees/users authenticate to the proxy server.

IV. SYSTEM ARCHITECTURE

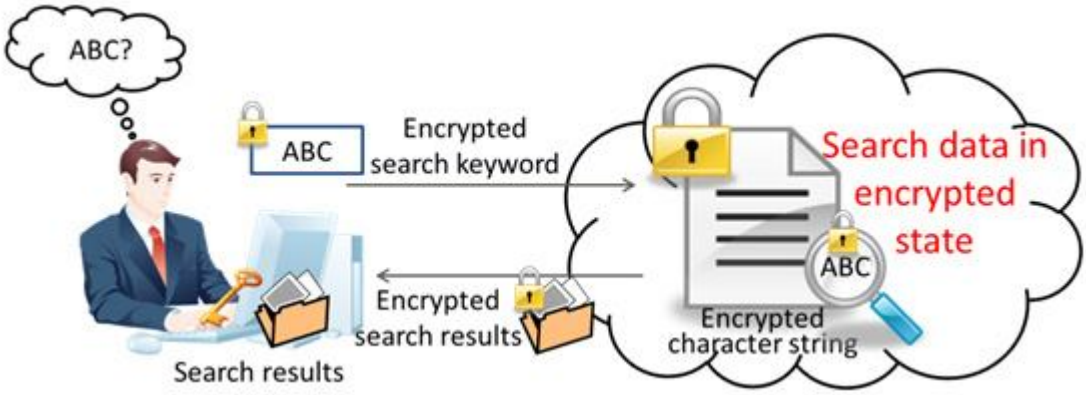

Fig: 2 Communication framework for keyword search over encrypted data

To perform a search, the data owner computes and sends a trapdoor encryption of the queried keywords to the cloud to initiate a protocol to search for the requested keywords in the corpus. Finally, the cloud responds to the data owner with the identifiers to the requested documents.

Module_Description

A. *Data Owner*

The data owner generates the encryption keys for hashing and encryption function. Then, all the files in the corpus are parsed for keywords. Bloom filters are attached to the hashed keywords and then n-grams are appended. The documents are then uniformly

encrypted and uploaded to the cloud server. To add files to the database, the data owner parses the files and uploads them along with the Bloom filters to the cloud server. To delete a file from the database, the data owner request to the cloud server, who removes the file. To perform a search, the data owner computes and sends a trapdoor encryption of the queried keywords to the cloud to initiate a protocol to search for the requested keywords in the corpus.

B. Cloud Server

The encrypted files sent by the data owner are received by the cloud server along with file trapdoors. As a result the cloud sends the identifiers of the required documents to the data owner.

C. Phrase search protocol

In phrase search technique, each document is parsed for lists of keyword pairs and triples. For example, 'Good Morning, Good Night' would yield the pairs, 'Good Morning', 'Morning Good' and 'Good Night', and the triples, 'Good Morning Good' and 'Morning Good Night'. A keyed hash is computed, $Hk_p(kw_j/kw_{j+1})$ for each keyword pair, and the result is used to set k bits in the Bloom filter, $B^2_{D\,i}$. Similarly the triples are hashed to generate the bloom filter, $B^3_{D\,i}$. These filters are arranged into matrices. For the first document these matrices contain filters $B^x_{D\,1}$as the first row. Then these matrices are transposed in order to generate bloom filters indexes for pairs I_{BF^2} and triples I_{BF^3}. To carry out the phrase search technique the data owner must first perform the Bloom filter hash computation of the pair,$Hk_p(kw_1 \backslash kw_2)$,for the keyword sequence,$kw' = \{kw_1, kw_2 \dots kw_q\}$, this is done to set the bits in the query filter if the phrase contains two keywords. If the phrase contains three or more than three keywords, it evaluates $Hk_p(kw_j \backslash kw_{j+1} \backslash kw_{j+2})$. Then the server after receiving the set bit locations, computes$T = I_{BF^2,q1} \& I_{BF^2,q2} \dots \& I_{BF^2,qx}$, where $I_{BF^2,qi}$ is the q_i^{th} row in I_{BF^2} if the phrase contains two keywords, and similarly using I_{BF^3} for longer phrases.
for each set bit index, i, in T, the following is true:

$\{Hk_p(kw_1 \backslash kw_2)\} \in B^2_{D\,i}$
for pairs and

$$\{Hk_p(kw_j \backslash kw_{j+1} \backslash kw_{j+2})\} \in B^3_{D\,i},\ where\ j = 1\ to\ q-2,$$

for triples
If they match the respective encrypted documents are retrieved by the cloud server.

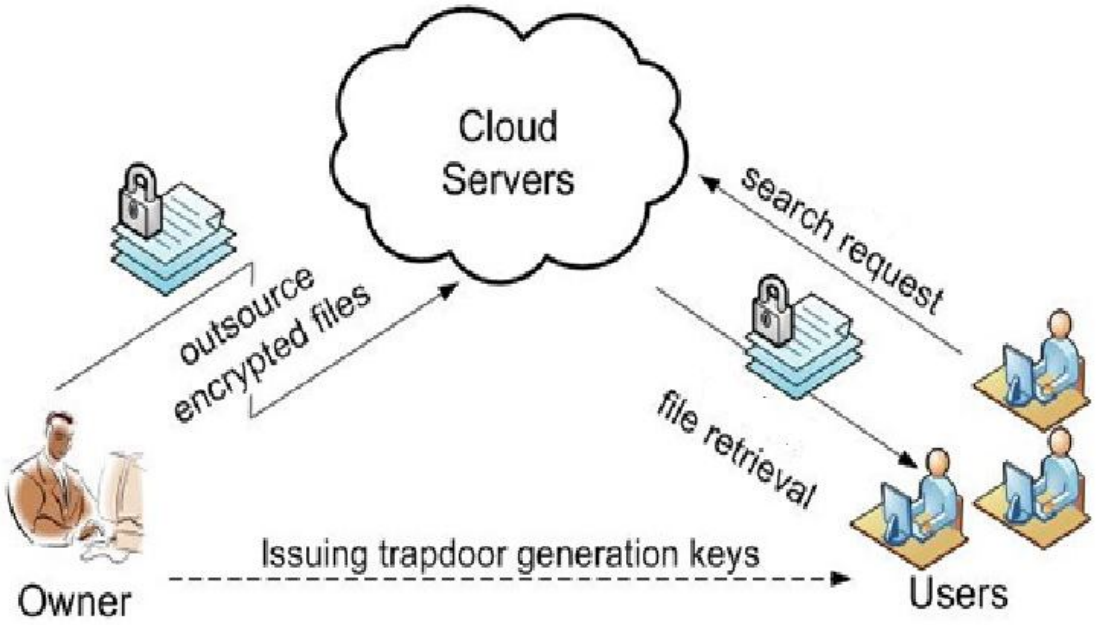

Fig: 3 Secure file retrieval

V. CONCLUSION

In this paper, our scheme achieves faster response time that needs only a single round of communication and bloom filter verification. Our solution is used to regulate the high computational cost by performing phrase search as n-gram verification rather than sequential chain verification and a location based search. In order to identify the candidate documents our approach independently runs the phrase search technique rather than performing the conjunctive keyword search. The proposed solution can also be adjusted to achieve maximum speed or high speed with a reasonable storage cost.

REFERENCES

[1] Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," In proceedings of Eurocrypt, 2004, pp. 506–522.
[2] M. Ding, F. Gao, Z. Jin, and H. Zhang, "An efficient public key encryption with conjunctive keyword search scheme based on pairings," in IEEE International Conference on Network Infrastructure and Digital Content, 2012, pp. 526–530.
[3] C. Hu and P. Liu, "Public key encryption with ranked multikeyword search," in International Conference on Intelligent Networking and Collaborative Systems, 2013, pp. 109–113.
[4] Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query," IEEE Transactions on Consumer Electronics, vol. 60, pp. 164–172, 2014.
[5] S. Zittrower and C. C. Zou, "Encrypted phrase searching in the cloud," in IEEE Global Communications Conference, 2012, pp. 764–770.
[6] Y. Tang, D. Gu, N. Ding, and H. Lu, "Phrase search over encrypted data with symmetric encryption scheme," in International Conference on Distributed Computing Systems Workshops, 2012, pp. 471–480.
[7] H. Poon and A. Miri, "An efficient conjunctive keyword and phrase search scheme for encrypted cloud storage systems," in IEEE International Conference on Cloud Computing, 2015.
[8] "A low storage phrase search scheme based on bloom filters for encrypted cloud services," to appear in IEEE International Conference on Cyber Security and Cloud Computing, 2015.
[9] Y. Yang, H. Lu, and J. Weng, "Multi-user private keyword search for cloud computing," in IEEE Third International Conference on Cloud Computing Technology and Science, 2011, pp. 264–271.
[10] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proceedings of the 2000 IEEE Symposium on Security and Privacy, 2000, pp. 44–55.
[11] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, "Class-based n-gram models of natural language," Computational Linguistics, vol. 18, no. 4, pp. 467–479, 1992.