International Journal of Computer Trends and Technology- volume4Issue3- 2013

Unassailable Fuzzy Keyword Searching in Cloud Storages

Tritty Mamachan[#1], Roshni. M. Thanka[#2]

[#]*Department of Computer Science and Engineering, Karunya University*
Coimbatore, India

***Abstract*— Cloud Computing allows users to dispense as well as access data remotely. The ongoing environment in which the data is stored in cloud, has a security dent, which is still not yet resolved. A notable number of techniques have been developed to retrieve data using keyword searches through encrypted data. When the user enters a keyword incorrectly or half of the expected word, is the technique efficient enough to search through the data. Fuzzy keyword technique has been proposed to solve this problem, but is it unassailable? In this paper, fuzzy keyword technique has been merged with security measures to provide an efficient mechanism for keyword searching.**

***Keywords*— Cloud Service Provider, Cloud Storage, Partial Decipherment, Security, Fuzzy Keyword.**

I. INTRODUCTION

The increase in the usage of computers has depleted the security layers that could be used to protect the data within. The storage space has also become a major issue, in the case of large companies. Cloud computing can be stated as shift of work and resources from personal computers or individual enterprise applications to set a cloud of computers. Cloud Computing offers solutions to many problems like hardware, machine failures etc. The great advantage of cloud computing is "elasticity": the ability to add capacity or applications almost at a moment's notice. Larger companies find it easier to manage collaborations in the cloud.

Though cloud computing is highly flexible and cost effective, it introduces and lot of security issues. Cloud computing can bring paradigm shift and benefits, to the industry [7].

As cloud computing is becoming more accrual, more information is being centralized into the cloud. Data owners are relieved from the burden of data storage and maintenance, to enjoy the on demand high quality data service. The reason for the risk can be because the data owners and cloud servers are not in the same trusted domain.

Searchable encryption techniques have been developed over the years to allow an efficient and secure search with keywords. These techniques do not suit the cloud computing scenarios as they only support exact keyword search.

II. EXISTING METHODS

The existing methods that have been used for the keyword searching is explained in the section below.

SPKS: Secure and Privacy Preserving Keyword Search

The SPKS scheme enables the CPS's to participate in the partial decipherment, this will reduce the computational overhead on users, without leaking any information about the plain text [9]. It also supports keyword searching on encrypted data. This scheme will enable the CSP to determine whether the keyword specified by the user is in the email, but it will not be aware of the information contained in the email, nor the keyword that was searched. It is proven to be semantically secure under the Bilinear Diffie Hellman assumption and the random oracle model [2]. The figure 2 shows the framework of the SPKS scheme that is described in this section.

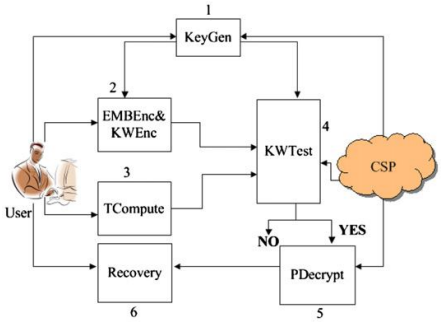

Fig.2 Framework of SPKS

This scheme consists of seven randomized polynomial time algorithms, which are the similar to that of in Efficient Privacy Preserving Keyword Searching Scheme. They are as follows:

a) Key Generation
b) Email Encryption
c) Keyword Encryption
d) Trapdoor Computation
e) Keyword Testing
f) Partial Decryption
g) Recovery

The main difference here was that, only a partial decryption is done here.

The user U and the CSP runs the KeyGen algorithm to generate their public or private key pairs. When U wants to store an email m containing keywords $W_1....W_k$ on cloud servers, U first runs the EMBEnc algorithm to encrypt the email, and then runs KWEnc to encrypt all the keywords, and finally sends both the ciphertext of the email and keywords to the CSP.

When U wants to retrieve emails containing keyword, he runs the TCompute algorithm to generate W_j's trapdoor T_{wj} and sends it to the CSP. On receiving the trapdoor the CSP runs the KWTest algorithm to determine whether a given email contains keyword W_j specified by U.

Before returning the results to U, the CSP runs PDecrypt to calculate an intermediate result for the decipherment. After that it returns along with the encrypted emails. When a ciphertext and is given, U runs the Recovery algorithm to recover the plain text.

III. PROPOSED SYSTEM

Fuzzy searching addresses the problem of finding appropriate information, by using the appropriate string matching technique.

In order to provide an effective and practical fuzzy searching with a better storage and search efficiency, the edit distance concept can be implemented into the keyword searching technique [7].

Wild Card Based Technique: All the variants of the keyword are edited based on distances. For example, if we take the word **LIGHT**, with edit distance 1 the wild card based fuzzy set can be created as follows [6].

SLIGHT,1 ={ LIGHT, *LIGHT, *IGHT, L*IGHT, L*GHT……LIG*T, LIGH*, LIGHT*}

The wildcard-based fuzzy set of *wi* with edit distance *d* is denoted as *Swi,d={S_wi,0, S_wi,1, · · · , S_wi,d}*, where *S_wi,τ* denotes the set of words *w_i* with *τ* wildcards.

Editing distance consists of three steps, namely:

a) **Substitution:** changing one character to another in a word.

b) **Insertion:** Inserting a single character into a word.

c) **Deletion:** deleting one character from a word.

The architecture can be depicted as in the following diagram.

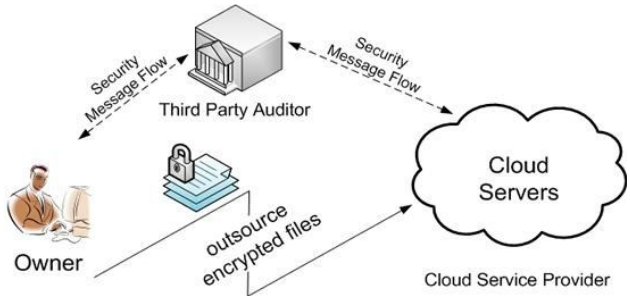

Fig.3. Architecture of Wild card Based Technique

While implementing the fuzzy keyword search a searchable encryption needs to be done. To achieve more efficient search, index approach is used where a single encrypted hash table index is built for the entire file collection. In the index table, each entry consists of the trapdoor of a keyword and an encrypted set of file identifiers, whose corresponding data files contain the keyword.

Two steps are needed to implement fuzzy keyword search.

1) Building up fuzzy keyword sets that incorporate not only the exact keywords but also the ones differing slightly due to minor typing errors, format inconsistencies, etc.

2) Designing an efficient and secure searching approach for file retrieval based on the resulted fuzzy keyword sets.

These steps can be further described as the following:

1) To build an index for *wi* with edit distance *d*, the data owner first constructs a fuzzy keyword set *Swi,d* using the wildcard based technique. Then he computes trapdoor set *{Tw_i }* for each *w_i ∈ Swi,d* with a secret key *sk* shared

between data owner and authorized users. The data owner encrypts FID*wi* as Enc(*sk*, FID*wi_wi*). The index table *{({Tw_i }w_i ∈Swi,d* , Enc(*sk*, FID*wi_wi*))*}wi∈W* and encrypted data files are outsourced to the cloud server for storage;

2) To search with (*w*, *k*), the authorized user computes the trapdoor set *{Tw_}w_∈Sw,k* , where *Sw,k* is also derived from the wildcard-based fuzzy set construction. He then sends *{Tw_}w_∈Sw,k* to the server;

3) Upon receiving the search request *{Tw_}w_∈Sw,k* , the server compares them with the index table and returns all the possible encrypted file identifiers *{*Enc(*sk*, FID*wi_wi*)*}*.The user decrypts the returned results and retrieves relevant files of interest[8].

Providing security is an important factor in the keyword searching, especially when cloud service provider is included. CSP's themselves can be threat to the data stored in the servers.

A trapdoor is computed for the service provider to gain access to the emails or files, without knowing much information about the contents that are stored by the user. The computation of the trapdoor is said to be the most tedious and important task, because it is providing an access to the stored information. Now the cloud service provider checks the keyword with those that are in the emails or files, to make sure the files with the correct keywords are given to the user.

If the keyword matches, it is returned to the user else file not found message is given to the user. Using the private key, public key and the cipher text obtained from the previous step, the partial decryption of the file is done by the service provider, which will lead to the decrease in the cost of computation and communication during encryption and decryption. This partially decrypted data is given to the user, from which they can decrypt the file they need. At the end is the recovery, where the obtained file is decrypted fully by the user.

IV. COMPARISON

A study about how SPKS scheme is more efficient that PEKS was done [9], by comparing the computation and communication cost during encryption as well as decryption, but it is highly costly in the case if computation.

TABLE I
COMPUTATIONAL COST OF ENCRYPTION

Operation	PEKS	SPKS
map	0	1
mul	2	2
exp	0	1
mod	M	0
has	0	3
xor	0	M+1

TABLE III
COMPUTATIONAL COST OF DECRYPTION

Operation	PEKS	SPKS
map	0	0
mul	1	0
exp	0	1
inv	1	0
mod	M	0
has	0	1
xor	0	M

Here n is a relatively small number in comparison with the average length of most emails. In this case, we need to split a longer email into several segments and pad the last segment to make each segment to have the equal length n. Due to these shortcomings, we go fuzzy keyword searching.

V. CONCLUSION

The keyword searching techniques improve the security of the user querying privacy, but cause an increase in the cost associated with computational and communication. Through rigorous security analysis, the proposed solution is secure and privacy- preserving.

ACKNOWLEDGMENTS

I sincerely thank all my Staffs and friends who give me support an encouragement in doing this survey. I also thank for all the research scholars who did the previous study on this topic through which I got insight about my topic.

REFERENCES

[1] Belare, A.Boldyreva, and A.O'Neil. "Deterministic and efficiently searchable encryption," International Proceedings of Rypto, of Lecture notes in Coomputer Science. Springer-Verlag, vol. 4622, 2007.

[2] Boneh D, Crescenzo G, Ostrovsky R, Persiano G. "Public key encryption with keyword search". International Proceedings of Eurocrypt, Lecture notes in computer science, vol. 3027;. p. 506–22, 2004.

[3] Cao N., C. Wang, M. Li, K. Ren, and W. Lou. "Privacy-preserving multikeyword ranked search over encrypted cloud data," in IEEE INFOCOM, Shanghai, China, April, 2011.

[4] David Q.Liu, Shilpashree Srinivasamurthy. "Survey on cloud computing security" In Computers and Electrical engineering, Elsevier, 2010

[5] Dawei Sun, Guiran Chang, Lina Sun and Xingwei Wang. "Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments", 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS], Science Direct, pp. 2852 – 2856, 2011.

[6] Julien Bringer and Hervé Chabanne. "Embedding edit distance to allow private keyword search in cloud computing". In Bringer and Chabanne Human-centric Computing and Information Sciences ,2012.

[7] Jan Li, Qian Wang, Cong Wang. Fuzzy Keyword search over encrypted data in cloud computing. In Proceedings of INFOCM, IEEE ,2010

[8] T. Balamuralikrishna, C. Anuradha. Fuzzy keyword search over encrypted data in cloud. In Asian Journal of Computer Science and Information Technology; vol.3;.p.86-88, 2011

[9] Qin Liu,Guojun Wang, Jie Wu. Secure and privacy preserving keyword searching for cloud storage services. In: Journal of Network and Computer Applications; vol.35,.p.927-933, 2012

[10] Wang Jianfeng, Ma Hua, Tang Quang. "A new efficient verifiable fuzzy keyword search scheme," In Journal of wireless mobile networks, ubiquitous computing and dependable applications," vol.3 No.4; p.61-71, 2011