Fuzzy Keyword Search Over Encrypted Data in Cloud Computing

S. Muthuselvi[1] S. Porkodi[2] S. Lavanya[3] J. Jeejo Vetharaj[4]
[1,2,3,4]Department of Computer Science & Engineering
[1,2,3,4]Veltech Multitech Dr RR & Dr SR Engineering College

Abstract— Cloud Computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud Computing allows consumers and business to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more efficient computing, centralizing storage, memory, processing and bandwidth. Perhaps, the biggest concerns about cloud computing are security and privacy. As cloud computing becomes more prevalent with the increased rate of growth and adaptation of cloud. Daily, more and more sensitive information is being centralized into the cloud. For the protection of valuable proprietary information, the data must be encrypted before outsourcing. The existing search allows the user to search over encrypted data using keywords but in this technique accounts only for exact keyword search. In this paper, we formalize the problem and solve it by effective fuzzy keyword search over encrypted cloud data while maintain keyword privacy. We are proposing the mechanism called the Wildcard Based technique which returns the matching files when user searching inputs exactly match the predefined keywords, or the closest possible matching files based on similarity keyword semantics, when exact match fails.

Key words: Encrypted Data, Fuzzy Keyword Search

I. INTRODUCTION

Cloud computing is emerging a key computing platform for sharing resources .The sharing resources is based on three models: Infrastructure-as-a-Service, Platform-as-a-Service and Software-as-a-Service. These services are customized as per demand .Cloud computing are more popularly referred as just "the cloud". Cloud resources are usually not shared by the multiple users but it dynamically relocated as per demand. The storage of data in to the cloud reduces the burden of storage and maintenance of data on the user. The focus of our project is on enabling effective privacy-preserving fuzzy keyword search for information stored in cloud environments. Fuzzy Keyword Search augments system is mainly used by returning the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. Edit distance is used to quantify keywords similarity and for the development of a novel technique for constructing fuzzy keyword sets. This technique eliminates the need for counting all the fuzzy keywords and the total size of the fuzzy keyword sets is significantly decreases.

II. EXISTING SYSTEM

This approach provides fuzzy keyword search over the encrypted files while achieving search privacy using the technique of secure trapdoors. However, it provides serious efficiency disadvantages. The simple method in constructing fuzzy-keyword sets would introduce a large storage complexity, which greatly affect the usability. But existing system have many disadvantages:

- Approach used by existing system have efficiency problem.
- Fuzzy keyword set requires large storage capacity.

For Example:

The first character of keyword after a substitution operation of the listing variants

MASTER: {ASTER, BASTER, DASTER, YASTER, ZASTER}

A. *Design Goals:*

The goal is to solve the problem of effective fuzzy keyword search over encrypted cloud data while maintain keyword privacy. The goals are:

- To exploit edit distance to quantify keywords similarity.
- To design efficient and effective fuzzy search scheme.
- To validate the security of proposed solutions.

III. PROPOSED SYSTEM

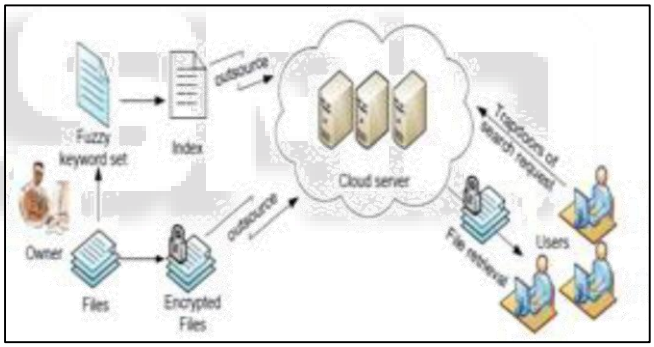

Fig. 1: Architecture of the fuzzy-keyword search

The fuzzy keyword search scheme returns the search results according to the following rules:

- If the user's searching input exactly matches the pre-defined set of keywords, the server will return the files containing the keywords.
- If there exist some error in spelling or some forma t inconsistencies in the searching input, the server will return the closest possible results based on pre-specified similarity semantics.

IV. CONSTRUCTION

The construction of effective fuzzy keyword search, is the main idea behind the secure search and it consists of two concepts:

1) Generate fuzzy keyword set that include exact keyword along with keyword that differ from exact keyword due to minor types or due to inconsistency in formatting.
2) T design mechanism to securely retrieve files based upon the keyword entered.

A. Advanced Technique used in Construction:

In this paper, we use the proposed system i.e., the technique used is ”edit distance”. The edit distance is defined as the smallest number of insertion, deletion and substitution required for changing one string to another. The edit distance from one string to another is calculated by number of operations that need to be carried out to transform one string to another. Edit distance technique have been applied to virtually all spelling correction tasks, including text editing and natural language interfaces.

B. Main Modules:

- Wildcard-based technique
- Gram-Based technique

1) Wildcard-Based Technology:

A special computer keyboard character or sequences of character, used to represents one or more other character. Usage in the computer and internet worlds is similar to a joker in a deck of playing cards that can be made a ”wildcard” to act as any other card in the deck. Wildcard based technique allows user to search for all files either names that contain similar qualities. For example, in Microsoft word files begin with letter by searching for s* doc, the asterisk is used as the wildcards to represent all other character sequences following the initial s letter. The technique can be useful if a user cannot remember a specific file name or would like to see an entire grouping of files that were created to share some part of their name.

In the above approach, all the variants have to be listed even if an operation is performed at the same position. Based on the denote set operations at the same position .The wildcard based fuzzy set edits distance to solve the problems.

a) For Example:

The keyword for MASTER with the pre-set edit distance 1, its wildcard based fuzzy keyword set can be constructed as**:**

SMASTER, 1={ MASTER, *MASTER, *ASTER, M*ASTER, M*STER, MASTE*R, MASTE*, MASTER*}

b) Edit Distance:

- Insertion: Inserting a single character into a word.
- Deletion: Deleting a character from a word.
- Substitution: Changing one character to another in a word.

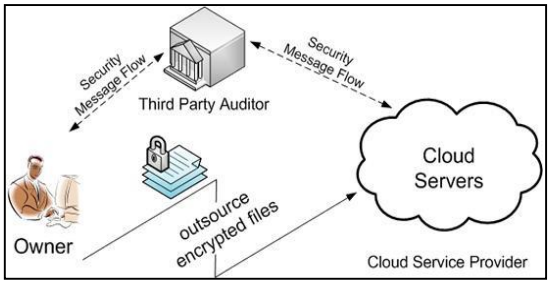

Fig. 2: Wildcard-Based Fuzzy set

c) Algorithm:

For Wildcard-Based Fuzzy set construction:

```
1: procedure CreateWildcardFuzzySet(wi, d)
2:     if d > 1 then
3:         Call CreateWildcardFuzzySet(wi, d − 1);
4:     end if
5:     if d = 0 then
6:         Set S'wi,d = {wi};
7:     else
```

2) Gram-Based Technique:

Another efficient and effective technique for constructing fuzzy set is based on grams is, for example: the sequence of character “to build” will have 5 grams as: to bu’’, ’’o buil”, and “build”. Such gram can be used as signature for approximate search. But here we used gram for matching purpose only. In gram based technique, edit distance operation can have its effect at one position only. It means after primitive operation the order of the remaining character after primitive operation the order of the remaining character remain same as it is present before the operation.

a) For example:

The gram-based fuzzy set SMASTER, 1 for keyword, MASTER can be constructed as

{MASTER,MSTER,MATER,MASER,MASTR,MASTE,MASTER}

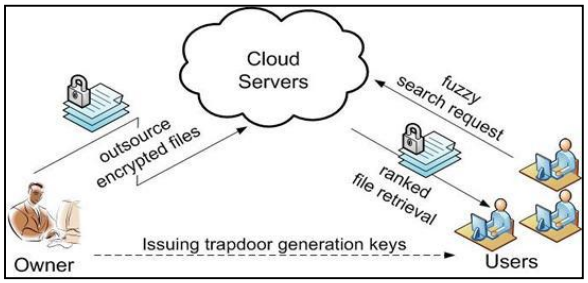

Fig. 3: Gram-based technique

Generally, given a keyword wi with ℓ single characters, the size of S′wi,τ is Cℓ−τ

ℓ, and the size of Swi,d is Cℓ

ℓ +Cℓ−1

ℓ +· · ·+Cℓ−d

ℓ. compared to wildcard –based construction, gram-based construction can further reduce the storage of the index from 40 MB to approximately 10MB under the same setting as in the wildcard –based approach. The procedure for gram-based fuzzy set construction is shown in the following algorithm.

Algorithm 2 Gram-based Fuzzy Set Construction

```
1: procedure CreateGramFuzzySet(wi, d)
2:     if d > 1 then
3:         Call CreateGramFuzzySet(wi, d − 1);
4:     end if
5:     if d = 0 then
6:         S'wi,d = {wi};
7:     else
8:         for (k ← 1 to |S'wi,d−1|) do
9:             for j ← 1 to 2 ∗ |S'wi,d−1[k]| + 1 do
10:                Set fuzzyword as S'wi,d−1[k];
11:                Delete the j-th character;
2:                 if fuzzyword is not in S'wi,d−1 then
13:                    Set S'wi,d = S'wi,d ∪ {fuzzyword}
14:                end if
15:            end for
16:        end for
17:    end if
18: end procedure
19: end procedure
```

V. RELATED WORK

A. Plaintext Fuzzy Keyword Search:

In recent day, the importance of fuzzy search has received attention in the context of plaintext searching in information retrieval community. They addressed this problem in the traditional information-access paradigm by allowing user to

search without any try-and-see approach for finding relevant information based n approximate string matching. At the first glance, it seems possible from one to directly apply these matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within a alphabet. However, this trivial construction suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

B. Investigation of fuzzy keyword:

In this paper, consider a cloud data system consisting of cloud server, data owner and data user. The fuzzy keyword set can be defined by using edit distance. If a collection is given of an encrypted data files, C=(F1,F2,…..FN) stored in the cloud server, a predefined set of distinct keywords, W={w1,w2……wp), the cloud server provides the search service for the authorized users over the encrypted data C. An authorized user types in a request to selectively retrieve data files. The cloud server maps the search request to a set of files, where each one of them is indexed based on a file ID and linked to a set of keywords. For example: the following is the listing variants after a substitution operation on the first character of the keyword MASTER: {AASTER,BASTER,CASTER,DASTER,……YASTER, ZASTER}.

C. Search Result:

A list of all the files containing the keywords that the user had searched is displayed on the screen .The user then downloads the file that the user needs by click on the download option.

Search Text File Compairing Keywords | Reset

File Name	Type	Containing Keywords	Uploaded By	Uploaded Date	Download Link
README.txt	txt	Install, .nstall, I.stall, In.tall, Ins.all, Inst.ll, Insta.l, Instal., Compatibility, .ompatibility, C.mpatibility, Co.patibility, Com.atibility, Comp.tibility, Compa.ibility, Compat.bility, Compati.ility, Compatib.lity, Compatibi.ity, Compatibil.ty, Compatibili.y, Compatibilit., oneye, .neye, o.eye, on.ye, one.e, oney. eyeos, .yeos, e.eos, ey.os, eye.s, eyeo. server, .erver, s.rver, se.ver, ser.er, serv.r, serve., uninstall, .ninstall, u.install, un.nstall, uni.stall, unin.tall, unins.all, uninst.ll, uninsta.l, uninstal., . ..	CLOUDMASTER000000006	Thu Apr 10 00:30:47 IST 2014	Download
README.txt	txt	Install, .nstall, I.stall, In.tall, Ins.all, Inst.ll, Insta.l, Instal., oneye, .neye, o.eye, on.ye, one.e, oney.	CLOUDMASTER000000006	Thu Apr 10 00:28:50 IST 2014	Download

Fig. 4: Search Result

D. Others

Private matching, as another related notion, has been studied mostly in the context of secure multiparty computation to let different parties compute some function of their own data collaboratively without revealing their data to the others. These function could be intersection or approximate private matching of two sets, etc. The private information retrieval is an often-used technique to retrieve the matching items secretly, which has been widely applied in information retrieval from database and usually incurs unexpectedly computation complexity.

VI. FUTURE SCOPE

In this paper, the future scope of the system is to do the indexing of the mapped words and fuzzy set so as to increase the functionality of the search procedure .Encryption of file formats can be done and it also decryption of images file can also done.

VII. CONCLUSION

In this paper ,the ultimate aim is to make a privacy-preserving fuzzy search for achieving effective usage of remotely stored encrypted data in cloud computing .By designing an advanced search mechanism for constructing the storage efficient fuzzy-keyword sets based on the similarity metric edit distance. Through rigorous security analysis, we show that our proposed system is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search. Extensive experimental results demonstrate the efficiency of our solution.

REFERENCES

[1] E. -J. Goh, ”Secure indexes,” Cryptology e Print Archive, Report 2003/216,2003 , http://eprint.acr.org/.
[2] D. Boneh, G.D. Crescenzo, R. Ostovsky, and G. Persiano ,”Public key encryption with keyword search,” in Proc .of EUCROCRYP’04,2004.
[3] S. J i,G. Li, C. Li and J. Feng, ”Efficient interactive fuzzy keyword search,” in proc.of WWW’09,2009.
[4] Google, “Britney spears spelling correction, “Referenced online http://www.google.com/jobs/britney.html,june.
[5] B.waters, D,Balfanz, G. Durfe, and D.Smeters, “Building an encrypted and searchable audit log,” in proc. of 11 th Annual Network and Distributed System, 2004
[6] Y.C. Chang and Mitzenmacher, “Privacy preserving Keyword searches on remote encrypted data,” in Proc. Of CAN’S05, 2005.

