Mr. Mahesh Lanjewar, Swapnali Ghadge, Sneha Mane, Priti Dalvi / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 2, Mar-Apr 2012, pp.870-874

Fuzzy Keyword Search Over Encrypted Data Using Cloud Computing

Mr. Mahesh Lanjewar , Swapnali Ghadge, Sneha Mane, Priti Dalvi
(PVPP College Of Engineering Department Of Information Technology, Mumbai University)

ABSTRACT
Cloud computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth. Perhaps the biggest concerns about cloud computing are security and privacy. If a client can log in from any location to access data and applications, it's possible the client's privacy could be compromised. In existing technique we retrieve the files from the cloud, by searching the keywords on the encrypted data. There are many searching technique which were implemented in the cloud these technique supports only exact keyword search. Typical users searching behaviours are happen very frequently these are the drawbacks with the existing system which are not suitable for cloud computing environment and which effects system usability. Using fuzzy search the exact keywords are displayed along with similarity keywords, which solve the problems faced by the cloud users. This paper concentrates on solving the problems of the user who search the data with the help of fuzzy keyword on cloud.

Keywords– Edit distance, fuzzy search, trapdoor.

1. INTRODUCTION
Cloud computing, the new term for the long dreamed vision of computing as a utility, enables convenient, on-demand network access to a centralized pool of configurable computing resource that can be rapidly deployed with great efficiency and minimal management overhead .
The amazing advantages of Cloud Computing include: On-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing, and transference of risk. Thus, Cloud Computing could easily benefit its users in avoiding large capital outlays in the deployment and management of both software and hardware. Undoubtedly, Cloud Computing brings unprecedented paradigm shifting and benefits in the history of IT.

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud. However, the fact that data owners and cloud server are not in the same trusted domain may put our sourced data at risk, as the cloud server may no longer be fully trusted. The individual users might want to only retrieve certain specific data files during a given session. One of the most popular ways is to selectively retrieve files through keyword-based search. Unfortunately, data encryption restricts user's ability to perform keyword search and thus makes the traditional plaintext search methods unsuitable for Cloud Computing. To securely search over encrypted data, searchable encryption techniques have been developed in recent years. Although allowing for performing searches securely and effectively, the existing searchable encryption techniques do not suit for cloud computing scenario since they support only exact keyword search. It is quite common that users' searching input might not exactly match those pre-set keywords due to the possible typos, the naive way to support fuzzy keyword search is through simple spell check mechanisms. However, this approach does not completely solve the problem and sometimes can be ineffective. Thus, the drawback of existing schemes signifies the important need for new techniques that support searching flexibility. In this paper, we study a new computing paradigm, called, fuzzy search. Fuzzy search means when searching for relevant records, the system also tries to find those records that include words similar to the keywords in the query, even if they do not match exactly. Moreover, in Cloud Computing, data owners may share their outsourced data with a large number of users. The individual users might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways is to selectively retrieve files through keyword-based search instead of retrieving all the encrypted files back which is completely impractical in cloud computing scenarios. Such keyword-based search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios, such as Google Search. Unfortunately, data encryption restricts user's ability to perform keyword search and thus makes the traditional Plaintext search methods unsuitable for cloud computing. Besides this, data Encryption also demands the protection of keyword privacy since keyword usually contain important information related to the data files.
In this paper ,we focus on enabling effective yet privacy preserving fuzzy keyword search in cloud computing .to the best of our knowledge we formalize for the first time the problem of effective fuzzy keyword search over encrypted cloud data.
While maintaining keyword privacy. Fuzzy keyword search greatly enhances system usability by returning the matching files when user's searching inputs exactly match. The

predefined keyword or the closest possible matching files based on keyword similarity semantics, when exact match fails.
More specifically, we use edit distance to quantify keywords similarity and develop a novel technique, i.e., a wildcard-based technique, for the construction of fuzzy keyword sets.. Based on the constructed fuzzy keyword sets, we propose an efficient fuzzy keyword search scheme. Through rigorous security analysis, we show that the proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search.

2. RELATED WORK

2.1. Fuzzy Set Theory:

Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets were introduced simultaneously by Lotfi A. Zadeh and Dieter Klaua[1] in 1965 as an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition — an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1. In fuzzy set theory, classical bivalent sets are usually called crisp sets. The fuzzy set theory can be used in a wide range of domains in which information is incomplete or imprecise, such as bioinformatics. Fuzzy sets can be applied, for example, to the field of genealogical research. When an individual is searching in vital records such as birth records for possible ancestors, the researcher must contend with a number of issues that could be encapsulated in a membership function. Looking for an ancestor named John Henry Pittman, who you *think* was born in (probably eastern) Tennessee circa 1853 (based on statements of his age in later censuses, and a marriage record in Knoxville), what is the likelihood that a particular birth record for "John Pittman" is *your* John Pittman? What about a record in a different part of Tennessee for "J.H. Pittman" in 1851? (It has been suggested by Thayer Watkins that Zadeh's ethnicity is an example of a fuzzy set)

2.2. Plaintext fuzzy keyword search.

Recently, 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 using try-and-see approach for finding relevant information based on approximate string matching. At the first glance, it seems possible for one to directly apply these string matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within an alphabet. However, this trivial construction suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

2.3. Searchable encryption.

Searchable encryption schemes provide an important mechanism to cryptographically protect data while keeping it available to be searched and accessed. In a common approach for their construction, the encrypting
Entity chooses one or several keywords that describe the content of each encrypted record of data. To perform a search, a user obtains a trapdoor for a keyword of his/her interests and uses this trapdoor to find all the described by this keyword.
We present a searchable encryption scheme that allows users to privately search by keywords on encrypted data in public key setting and decrypt the search result. To this end , we define and implements two primitives: public key encryption(PEOKS) and oblivious keyword search and committed blind anonymous identity-based encryption. PEOKS is an extension of public key encryption with keyword search in which users can obtain trapdoor from the secret key holder without revealing the keywords. PEOKS scheme is used to build public key encrypted database that permits private searches i.e.; neither the keyword nor the search result are revealed.

2.4. Fuzzy Keyword Investigation:

The fuzzy keyword set can be defined by using edit distance as follows: Given a collection of n encrypted data files C = (F1, F2. . . FN) stored in the cloud server, a set of distinct keywords W = {w1, w2, …,wp} with predefined edit distance d, and a searching input (w, k) with edit distance k ($k \leq d$),the execution of fuzzy keyword search returns a set of file IDs whose corresponding data files possibly contain the word w, denoted as FID_w: if w = wi belongs to W, then return FID_{wi} ; otherwise, if w does not belong to W,then return $\{FID_{wi}\}$, where $ed(w,wi) \leq k$. Note that the above definition is based on the assumption that $k \leq d$. Infect, d can be different for distinct keywords and the system will return $\{FID_{wi}\}$ satisfying $ed(w,wi) \leq \min\{k, d\}$ if exact match fails. For example, the following is the listing variants after a substitution operation on the first character of keyword CASTLE: {AASTLE, BASTLE, DASTLE, · · · YASTLE, ZASTLE}.

2.5.The Significant Fuzzy Search Procedure:

Based on the storage-efficient fuzzy keyword sets, we show how to construct an efficient and effective fuzzy keyword search procedure. The scheme of the fuzzy keyword search goes as follows:

1) To build an index for w_i with edit distance d, the data owner first constructs a fuzzy keyword set $S_{wi,d}$ using the wildcard based technique. Then he computes trapdoor set $\{T_{w'i}\}$ for each w′i € $S_{wi,d}$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 {({$T_{w'i}$} w′i € $S_{wi,d}$ 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 € W , where $S_{w,k}$ is also derived from the wildcard-based fuzzy set construction. He then sends {Tw′}w′ € $S_{w,k}$ to the server;

3) Upon receiving the search request {Tw′}w′ € $S_{w,k}$, the server compares them with the index table and returns all the possible encrypted file identifiers {Enc(sk, FID $_{wi}$||wi)} according to the fuzzy keyword definition. The user decrypts the returned results and retrieves relevant files of interest. In this construction, the technique of constructing search request for w is the same as the construction of index for a keyword. As a result, the search request is a trapdoor set based on $S_{w,k}$, instead of a single trapdoor as in the straightforward approach. In this way, the searching result correctness can be ensured.

3. EXISTING SYSTEM AND PROPOSED SYSTEM

3.1. Existing System

Using the technique of secure trapdoor the existing system allows user to perform fuzzy keyword search over encrypted data. It also achieve search privacy. But existing system have many disadvantages:

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

For example, after substitution operation on the first character of keyword, results we get are given below. CASTLE: {AASTLE, BASTLE, DASTLE, YASTLE, ZASTLE}.

3.2. Design Goals

In this paper, we try to solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. So our goals are:

1) to exploit edit distance to quantify keywords similarity
2) to design efficient and effective fuzzy search scheme.
3) To validate the security of proposed solution.

3.3 Proposed System

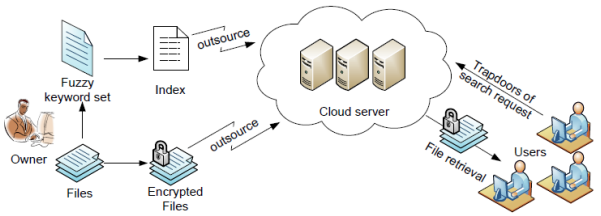

Fig. 1: Architecture of the fuzzy keyword search

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 keyword;

If there exist some error in spelling or some format inconsistencies in the searching input, the server will return the closest possible results based on pre-specified similarity semantics. Architecture of fuzzy keyword search is shown in the Fig. 1. Let us consider a semi-trusted server. Even though data files are encrypted, the cloud server may try to derive other sensitive information from user's search requests while performing the keyword-based search over Cloud. Thus, the search should be done in a well secured way that permits data files to be securely retrieved during exhibiting as little information as possible to the cloud server. In this paper, when designing fuzzy keyword search scheme, we will follow the security definition deployed in the traditional searchable encryption .More specifically, it is required that nothing should be leaked from the remotely stored files and index beyond the outcome and the pattern of search queries

3.4. Construction of effective fuzzy keyword search

The main idea behind performing the secure fuzzy keyword search consists of two concepts:

1) Generate fuzzy keyword set that include exact keyword along with keyword that that differ from exact keyword due to minor typos or due to inconsistency in formatting

2) To design mechanism to securely retrieve files based upon the keyword entered.

3.5. Advanced Technique for Constructing Fuzzy Keyword Sets

For the construction for effective fuzzy keyword set we propose technique i.e. edit distance technique.

The edit distance is defined as the smallest number of insertion, deletions and substitution required for changing one string to another (Lavenshtein 1965). The edit distance from one string to another is calculated by number of operations (substitution, insertion, deletion) 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.

4. MAIN MODULES:

- **Wildcard – Based Technique**
- **Gram - Based Technique**
- **Symbol – Based Trie – traverse Search Scheme**

4.2.1 Wildcard – Based Technique

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 "wild card" to act as any other card in the deck. Wildcard based technique allows users to search for all files either names that contain similar qualities.foe example , in Microsoft word files begin with the letter s by searching for s*.doc. the asterisk is used as the wildcard 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.

The wildcard-based fuzzy set edits distance to solve the problems of performing edit distance at same position . For example, for the keyword CASTLE with the preset edit distance 1, its wildcard based fuzzy keyword set can be constructed as
SCASTLE,1={CASTLE,*CASTLE,*ASTLE,C*ASTLE,C*STLE, CASTL*E, CASTL*,CASTLE*}.

Edit Distance operation can be perform in 3 ways:

a. Substitution
b. Deletion
c. Insertion

a) Substitution: changing one character to another in a word;
b) Deletion: deleting one character from a word;
c) Insertion: inserting a single character into a word.

4.2.2 Gram – Based Technique

Another efficient and effective technique for constructing fuzzy set is based on grams. Gram is ,for example the sequence of character " to build" will have 5 grams as :to bu","o bui","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 operation can have its effect at one position only .It means after primitive operation the order of the remaining character (other than one to which edit distance operation affect) remain same as it is present before the operation. For example, the gram-based fuzzy set SCASTLE, 1 for keyword CASTLE can be constructed as **{CASTLE, CSTLE, CATLE, CASLE, CASTE, CASTL, ASTLE**}

4.2.3. Symbol – Based Trie – traverse Search Scheme

We propose symbol based trie search scheme in order to increase searching efficiency. In symbol-based trie-traverse search scheme" multi-way tree is constructed for storing the fuzzy keyword set and finally retrieving the data. This greatly reduces the storage and representation overheads. Cloud database can have many files and administrator who is the owner and user.Adminitrator have authority to add , delete and view any users data. When user upload file to database it needs to encrypted. DES algorithm is used for encryption which makes data secure and protect it from unauthorized access. In order to access file user must have authorization. Authorization of user provided to the user by unique key which he gets when he perform login for the first time. All the fuzzy search words in trie can be found by depth first search approach. User can get search result by entering keywords along with conjunction of single words i.e. AND,OR,BOTH.
For example, if file has attribute,
Illeness: diabetes, Fever, hospital: A, B, C, D**.**
then user can access that file with dummy attribute (AND,OR,BOTH) as illness: fever AND hospital :A etc.

5. CONCLUSION & FUTURE WORK

In this paper, for the first time we formalize and solve the problem of supporting efficient yet privacy preserving fuzzy search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. We used an advanced technique (i.e., wild card-based technique) to construct the storage efficient
fuzzy keyword sets by using edit distance technique. With the help of symbol based tri search scheme we enhance searching efficiency. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search.
Future work is on security mechanisms that support search semantics that takes into consideration conjunction of keywords, sequence of keywords, and even the complex natural language semantics to produce highly relevant search results and search ranking that sorts the searching results according to the relevance criteria.

6. REFERENCES

1) Google, "Britney spears spelling correction, "Referenced Online http://www.google.com/jobs/britney.html,June 2009.
2) M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Proceedings of Crypto 2007, volume 4622 of LNCS. Springer-Verlag, 2007.
3) D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000.
4) E.-J. Goh, "Secure indexes," Cryptology ePrint Archive,Report 003/216,2003, http://eprint.iacr.org/
5) D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. Of EUROCRYP'04, 2004.
6) B. Waters, D. Balfanz, G. Durfee, and D. Smetters, "Building an encrypted and searchable audit log," in Proc. of 11th Annual Network and Distributed System, 2004.
7) Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.

8) R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS'06, 2006.
9) D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proc. of TCC'07, 2007, pp. 535–554.
10) F. Bao, R. Deng, X. Ding, and Y. Yang, "Private query on encrypted data in multi-user settings," in Proc. of ISPEC'08, 2008
11) C. Li, J. Lu, and Y. Lu, "Efficient merging and filtering algorithms for approximate string searches," in Proc. of ICDE'08, 2008
12) J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. N.Wright "Secure multiparty computation of approximations," in Proc. of ICALP'01
13) R. Ostrovsky, "Software protection and simulations on oblivious rams," Ph.D dissertation, Massachusetts Institute of Technology,1992.
14) V. Levenshtein, "Binary codes capable of correcting spurious insertions and deletions of ones," Problems of Information Transmission, vol. 1,no.1, pp. 8–17, 1965
15) Tech terms: what every telecommunications and digital media person should know By Jeff Rutenbeck, Jeffrey Blaine