www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 3 March,2014 Page No. 5102-5104

Secure Fuzzy keyword Search using an Advanced Technique over Encrypted Cloud Data

Deeptha Hegde, Saritha

Department of Computer Science and Engineering
Sahyadri College of Engineering and Management
Mangalore-575007
deeptha.kadambar@gmail.com

Department of Information Science and Engineering
Sahyadri College of Engineering and Management
Mangalore-575007
Saritha.ise@sahyadri.edu.in

Abstract -**Cloud computing is a technology that uses the internet and central remote servers to maintain data and applications. As the data produced by individuals and enterprises that need to be stored and utilized are increasing rapidly, data owners are motivated to outsource their local complex data management systems into the cloud for its great flexibility and economic savings. To ensure that the data can be stored in the cloud securely, data owners encrypt their data before outsourcing to the cloud, which makes searching scheme on a large amount of encrypted data a demanding task. The traditional searchable encryption schemes provide a number of approaches to search on encrypted data, but they all support only exact keyword search. Exact keyword search is unsuitable for cloud storage systems, because it doesn't allow users to make any spelling errors or format inconsistencies, and thereby reduces the system usability. The proposed system uses a wild-card based technique with edit distance as the similarity metric to obtain a fuzzy keyword sets.This solves the problems of the cloud user who search the encrypted cloud data with the help of fuzzy keyword, thereby maintaining keyword privacy.**

Keywords: edit distance, fuzzy keyword, index table, trapdoor

I. INTRODUCTION

Cloud computing is a Web-based processing model which enables users to outsource their data to the cloud servers over internet. By storing their data into the cloud, the data owners can be freed from the burden of data storage and maintenance so as to enjoy the on-demand high quality data storage service. To protect data privacy and combat unsolicited accesses, important data has to be encrypted before outsourcing [4] so as to provide end-to-end data confidentiality assurance in the cloud. In Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to retrieve only certain specific data files they are interested in during a given session. To provide this facility, one of the best ways is to selectively retrieve the files through keyword based search instead of providing all the information irresptive of user's interest. Such keyword based search techniques provides only the data which the users wants such as google search [2]. The actual traditional encryption method to support keyword search is simple spell check mechanism. This mechanism does not work perfectly for all types of keywords. It is ineffective because it needs more user interaction when the spell check algorithm works, which unnecessarily gives burden to the user. So that user effort is more in this mechanism compare to other. Another reason is sometimes spell check algorithm does not work when the user enters wrong keyword such as toy and boy. i.e., it only works for exact keyword match which restricts the users to perform keyword search which is not suitable for cloud computing Thus, the drawbacks of existing schemes signifies the important need for new techniques that supports flexible search, by tolerating both minor typos and format inconsistencies.

Fuzzy keyword search greatly enhances system usability 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 using an advanced technique, i.e., an wildcard-based technique to construct f fuzzy keyword sets. This technique eliminates the need for listing all the fuzzy keywords and the resulted size of the fuzzy keyword sets is reduced significantly. Based on the fuzzy keyword sets, an efficient fuzzy keyword search scheme is proposed.

II. RELATED WORK

The notion of searches on encrypted data was first proposed by Song [4]. It deals with search problems between a user and an untrusted server. They proposed a scheme , where each word in the file is encrypted independently under a special two-layered encryption construction. Thus, it creates an overhead since searching is linear to the whole file collection length. Goh [5] proposed to use Bloom filters to construct the indexes for the data files, reducing the work load for each search request proportional to the number of files in the collection. Chang [8] also developed a similar per-file index scheme . To further enhance search efficiency, Curtmola [9] proposed a per-keyword based approach, wherein a single encrypted hash table index is built for the entire file collections. In the index table, each entry consists of the trapdoor of a keyword and an encrypted collection of related file identifiers whose corresponding data files contains the keyword. As a complementary approach, Boneh [6] presented the first public-key based searchable encryption scheme, with an analogous scenario to that of [4]. All these existing schemes support search on only exact keyword, and hence are not suitable for Cloud Computing.

The importance of fuzzy search has received attention in the context of plaintext searching in information retrieval community [13-15]. 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 using approximate string matching. Even though 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, it suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

Private matching [16],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 anyone. These functions could be the intersection or approximate private matching of two sets. The private information retrieval [17] is a technique to retrieve the matching items securely, which has been widely applied in information retrieval process from database and usually incurs unexpectedly computation complexity.

III. DESIGN METHODOLOGY

The key idea behind secure fuzzy keyword search is two-fold: 1) building up fuzzy keyword sets that incorporate not only the exact keywords but also the ones differing slightly due to minor typos, format inconsistencies, etc.; 2) designing a storage-efficient and secure searching approach for file retrieval based on the resulted fuzzy keyword sets.

To provide more practical and effective fuzzy keyword search constructions with regard to storage and search efficiency, an advanced technique, i.e., a wildcard based technique is used to denote edit operations.

The edit distance ed (w1, w2) between two words w1 and w2 is the number of operations required to transform one of them into the other. The three primitive operations of edit distance are,

1) Substitution: substituting one character with another in a word.

2) Deletion: deleting one character from a word.

3) Insertion: inserting a character into a word

The wildcard-based fuzzy set of w_i with edit distance d is denoted as $S_{wi,d}$={$S'_{wi,0}$, $S'_{wi,1}$, • • • , $S'_{wi,d}$ }, where $S'_{wi,\tau}$ denotes the set of words w'_i with τ wildcards. Each wildcard represents an edit operation on w_i. For example, consider the keyword CASTLE with the pre-set edit distance 1, then, its wildcard-based fuzzy keyword set can be constructed as $S_{CASTLE,1}$={CASTLE,*CASTLE,*ASTLE, *ASTLE,C*STLE, • • • , CASTL*E, CASTL*, CASTLE*}. The total number of variants on CASTLE constructed in this way is only 13 + 1, instead of 13 × 26 + 1 when the edit distance is set to be 1 . In other words, for a given keyword w_i with length l, the size of $S_{wi,1}$ will be only 2l+1 +1,and not (2l+ 1) × 26 + 1 as compared to other. The larger the pre-set edit distance, the more storage burden can be reduced and hence the proposed technique can help reduce the storage of the index from 30GB to approximately 40MB. The edit distance can be set to 2 and 3 and so on. In other words, the. number is only $O(l^d)$ for the keyword with length l and edit distance d.

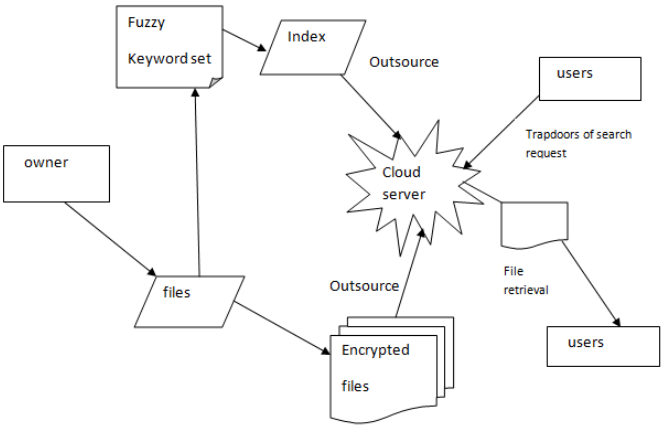

Fig.1 Architecture of fuzzy keyword search

Based on the storage-efficient fuzzy keyword sets, an efficient and effective fuzzy keyword search scheme as shown in Fig .1 is constructed as follows:

Based on the storage-efficient fuzzy keyword sets, an efficient and effective fuzzy keyword search scheme as shown in Fig .1 is constructed as follows:

1) The data owner first constructs a fuzzy keyword set $S_{wi,d}$ using the wildcard based technique with edit distance d ,to build an index for wi. Then he computes trapdoor set {Tw_i’} for each wi’∈ Swi,d with a secret key sk shared between data owner and authorized users. The data owner encrypts $FIDw_i$ as Enc(sk,$FIDw_i$||w_i). The index table {({Tw_i’} w_i’∈$S_{wi,d}$, Enc(sk, $FIDw_i$||w_i))} w_i∈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 {$T_{w'}$}w’∈S*w,k*, where S*w,k* is also derived from the wildcard-based fuzzy set construction. He then sends { $T_{w'}$}w’∈S*w,k* to the server;

3) Upon receiving the search request { $T_{w'}$}w’∈S*w,k* , the server compares them with the index table and returns all the possible encrypted file identifiers {Enc(sk, FIDw_i||w_i)}.

The user then decrypts the returned results and obtains relevant files of interest, thereby maintaining the data privacy, index privacy, trapdoor privacy and non-impersonation [5].

IV. RELIABILITY ANALYSIS

Reliability in web services is the most important factor. The users should be able to access the service of provided by the web more consistently without failure. User should be able to search and download the required data more efficiently.

The factors which affect the reliability of the fuzzy keyword search in encrypted data are [6]:

- Availability
- Security
- Performance

Availability: Availability is ensured by providing the application as a service in the web so that users can access the application at anytime and anywhere through internet.

Security: Security to the data in the web server is ensured by encrypting the data before uploading by the data administrator.Admin security and user security is ensured by strong admin and user authentication respectively.

Performance: Because of the usage of searchable encryption scheme and the construction of fuzzy keyword set, the time required to search the files and the space required to store the fuzzy keyword set is largely reduced, thereby increasing the performance.

V. CONCLUSION

As we know that cloud computing is the latest innovative technology, a user can store his personal and private encrypted files in a cloud and can retrieve them whenever he wants. The objective of the proposed system is to solve the problem of supporting efficient yet privacy-preserving fuzzy search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. Advanced technique namely wildcard-based technique is implemented to construct storage-efficient fuzzy keyword set using the edit distance concept. Based on the constructed fuzzy keyword sets, an efficient fuzzy keyword search scheme is proposed. Thus the proposed solution is secure and privacy-preserving, thereby realizing the goal of fuzzy keyword search.

REFERENCES

[1] Jin L , Qian Wang , Cong Wan, Ning Cao , Kui Ren , and Wenjing Lo "Fuzzy Keyword Search over Encrypted Data inCloud Computing",2010.

[2] Google, "Britney spears spelling correction," [Online].Available: http://www.google.com/jobs/britney.html, June 2009.

[3] 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.

[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," IEEE Symposium on Security and Privacy'00, 2000.

[5] E.J. Goh, "Secure indexes," Cryptology ePrint Archive, Report 2003/216, 2003, http://eprint.iacr.org/.

[6] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search,", 2004.

[7] 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.

[8] Y.C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data,", 2005.

[9] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions,", 2006.

[10] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data,", 2007, pp. 535–554.

[11] F. Bao, R. Deng, X. Ding, and Y. Yang, "Private query on encrypted data in multi-user Settings," ISPEC'08, 2008.

[12] C. Li, J. Lu, and Y. Lu, "Efficient merging and filtering algorithms for approximate string searches," , 2008.

[13] A. Behm, S. Ji, C. Li, and J. Lu, "Space-constrained gram-based indexing for efficient approximate string search," ICDE'09.

[14] S. Ji, G. Li, C. Li, and J. Feng, "Efficient interactive fuzzy keyword search,", 2009.

[15] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. N. Wright, "Secure multiparty computation of approximations,"2001.

[16] R. Ostrovsky, "Software protection and simulations on oblivious rams," Ph.D dissertation, Massachusetts Institute of Technology, 1992.

[17] 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.