ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Secured Data Transfer Cloud Encrypted Outsourced Data with Accurateness Enhancement

Misha Mathew [1], Ms.S.Brindha[2]

P.G. Student, Department of Computer Science and Engineering, Easa College of Engineering and Technology, Coimbatore, India[1]

Assistant Professor, Department of Computer Science and Engineering, Easa College of Engineering and Technology, Coimbatore, India[2]

ABSTRACT: In the current cloud computing scenario keyword-based search over encrypted outsourced data has become an important tool. The majority of the existing techniques are focusing on multi-keyword exact match or single keyword fuzzy search. However, those existing techniques find less practical significance in realworld applications compared with the multikeyword fuzzy search technique over encrypted data. The first attempt to construct such a multikeyword fuzzy search scheme was reported by Wang et al., who used locality-sensitive hashing functions and Bloom filtering to meet the goal of multi-keyword fuzzy search. Nevertheless, Wang's scheme was only effective for a one letter mistake in keyword but was not effective for other common spelling mistakes. Moreover, Wang's scheme was vulnerable to server out-of-order problems during the ranking process and did not consider the keyword weight. First, I develop a new method of keyword transformation based on the uni-gram, which will simultaneously improve the accuracy and creates the ability to handle other spelling mistakes. In addition, keywords with the same root can be queried using the stemming algorithm. I consider the keyword weight when selecting an adequate matching file set. Experiments using real-world data show that our scheme is practically efficient and achieve high accuracy.

KEY WORDS: Data Mining, Multy-keyword fuzzy ranked Search, Stemming Algorithm, Encoding/Decoding, Data Encryption

I. INTRODUCTION

In order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to users. A series of searchable symmetric encryption schemes have been proposed to enable search on cipher text. Traditional SSE schemes enable users to securely retrieve the cipher text, but these schemes support only Boolean keyword search. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage.

This scheme solved the problems of multi-keyword fuzzy search with high efficiency and accuracy. The most important result was that their scheme does not require the predefined fuzzy set. However, some other problems arose in this scheme. First, converting the keyword into a bi-gram set will increases the Euclidean distance.

I develop a novel method of keyword transformation based on the uni-gram. For misspelling of one letter, this method reduce the Euclidean distance between the misspelled keyword and the correct keyword.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

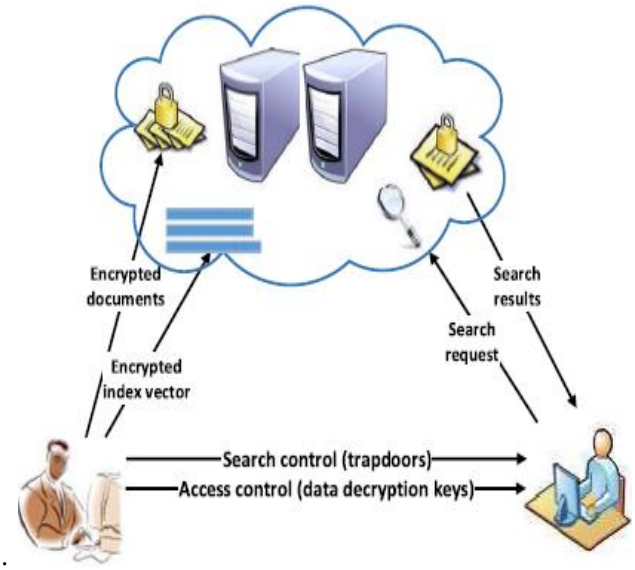

Fig. 1. The system model of our scheme.

II LITERATURE SURVEY

1. Ensuring Security And Privacy Preservation For Cloud Data Services

With the rapid development of cloud computing, more and more enterprises/individuals are starting to outsource local data to the cloud servers. However, under open networks and not fully trusted cloud environments, they face enormous security and privacy risks (e.g., data leakage or disclosure, data corruption or loss, and user privacy breach) when outsourcing their data to a public cloud or using their outsourced data.. I first present security threats and requirements of an outsourcing data service to a cloud, and follow that with a high-level overview of the corresponding security technologies. I then dwell on existing protection solutions to achieve secure, dependable, and privacy-assured cloud data services including data search, data computation, data sharing, data storage, and data access

Searchable Symmetric Encryption: Improved Definitions And Efficient Constructions

Private-key storage outsourcing allows clients with either limited resources or limited expertise to store and distribute large amounts of symmetrically encrypted data at low cost. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. To address this, several techniques have been proposed for provisioning symmetric encryption with search capabilities the resulting construct is typically called searchable encryption. The area of searchable encryption has been identified by DARPA as one of the technical advances that can be used to balance the need for both privacy and national security in information aggregation systems

III. PROPOSED METHODOLOGY

A. System Model

In this paper, I consider a cloud system consisting of data owner, data user and cloud server, see Fig. 1. In our system model, data owner has a collection of n data files F = (F1, F2, F3, , Fn) and outsources them to the cloud server in the encrypted form C. To enable efficient search operation on these encrypted files, data owner will build a secure searchable index I on the keyword set W extracted from F. Both the index I and the encrypted data files C, are outsourced to the cloud server. To search the encrypted data files for t given keywords, an authorized user computes a corresponding trapdoor T and sends it to cloud server. Upon receiving the trapdoor, the cloud server is responsible to search the index I and return the corresponding set of the encrypted documents. To improve the file retrieval accuracy

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

and save the communication cost, the search result should be ranked by the cloud server and return the top-K relevant files to the user as the search results.

B. *Threat Model*

In our threat model, both data owners and data users are trusted. However, the cloud server is honest-but-curious. Even though data files are encrypted, the cloud server may try to obtain other sensitive information from user search requests while performing keyword-based search over C. So the search should be performed in a secure manner that allows data files to be securely retrieved while revealing as little information as possible to the cloud. I consider the threat models as follow:

1) Know Ciphertext Model

2) Known Background Model

C. Design Goals

Our multi-keyword fuzzy search scheme should support more spelling mistakes. For example, "network security" related files should be found for a misspelled query "netward security"," network security","netrwork security" and "netwrk security".

D. Preliminariess

Three important techniques are used in our design: Stemming algorithm, Bloom Filter, Locality-Sensitive Hashing (LSH).

1) Stemming Algorithm:

A stemming algorithm is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form. A stemmer for English, for example, should identify the string "cats" (and possibly "catlike", "catty" etc.) as based on the root "cat", and "stems", "stemmer", "stemming", "stemmer" as based on "stem. It is widely adopted in Information Retrieval systems to improve performance.

2) Bloom Filter:

Bloom filter is a kind of data structure with very high space efficiency. It makes use of the m-bit array to represent a collection, and can determine whether an element belongs to the collection. It is initially set to 0 in all positions and for a given set $S = \{a1, a2, \cdots, an\}$, use l independent hash functions from $H = \{hi \mid hi : S \rightarrow m, 1 \leq i \leq l\}$ to insert an element $a \in S$ into the Bloom filter by setting the positions to be 1.

3) Locality-Sensitive Hashing (LSH):

LSH is an algorithm for solving the approximate or exact Near Neighbor Search in high dimensional spaces. LSH hashes input items so that similar items are mapped to the same buckets with high probability.

A. Multi-Keyword Fuzzy Search:

One of the first works to address the problem of multi-keyword fuzzy search over encrypted data problem and did not require a predefined fuzzy set, referred to as MFSE. In MFSE, a keyword was first transformed into a bi-gram set and the Euclidean distance was used to capture keywords similarity. The MFSE subsequently used LSH functions from the same hash family to generate the Bloom filter based index and query.

1) Keyword Search Algorithm:

In computer science, a search algorithm is an algorithm that retrieves information stored within some data structure. Data structures can include linked lists, arrays, search trees, hash tables, orvarious other storage methods. The appropriate search algorithm often depends on the data structure being searched. Searching also encompasses algorithms that query the data structure, such as the Sql Select command.

2) Permutation Search:

Permutation search is a useful tool that searches for terms that contain the set of keywords ordered in different sequences. This is a more restricted and targeted search, and is particularly useful when you are trying to target a specific group of keywords. To use this feature, the keywords must be entered separated by commas. "n/a" in the search column indicates that the keyword has no searches registered in the Keyword Discovery database.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

B. Scoring And Ranking:

Some of the multi-keyword searchable symmetric encryption schemes support only Boolean queries.Scoring is a natural way to weight the relevance. Based on the relevance score, files can then be ranked in either ascendingly or descendingly. Several models have been proposed to score and rank files in information retrieval (IR) community

1) Computing Vector Score:

In a typical setting I have a collection of documents each represented by a vector, a free text query represented by a vector, and a positive integer k. I seek the k documents of the collection with the highest vector space scores on the given query.

A faster algorithm for vector space scores :

FASTCOSINESCORE(q)
1 float $Scores[N] = 0$
2 **for each** d
3 **do** Initialize $Length[d]$ to the length of doc d
4 **for each** query term t
5 **do** calculate $w_{t,q}$ and fetch postings list for t
6 **for each** pair$(d, tf_{t,d})$ in postings list
7 **do** add $wf_{t,d}$ to $Scores[d]$
8 Read the array $Length[d]$
9 **for each** d
10 **do** Divide $Scores[d]$ by $Length[d]$
11 **return** Top K components of $Scores[]$

2) Efficient Scoring and Ranking:

The unit vector has only two non-zero components. In the absence of any weighting for query terms, these non-zero components are equal - in this case, both equal 0.707. For the purpose of ranking the documents matching this query, I am really interested in the relative scores of the documents in the collection. For any two documents d1,d2. For any document d, the cosine similarity is the weighted sum, over all terms in the query q, of the weights of those terms in d. This in turn can be computed by a postings intersection exactly as in the algorithm , with line 8 altered since I take wt, q to be 1 so that the multiply-add in that step becomes just an additionGiven these scores, the final step before presenting results to a user is to pick out the K-highest-scoring documents.

C. TRSE Design

Existing SSE schemes employ server-side ranking based on order preserving encryption to improve the efficiency of retrieval over encrypted cloud data. However, server-side ranking based on order preserving encryption violates the privacy of sensitive information, which is considered uncompromisable in the security-oriented third-party cloud computing scenario.To achieve data privacy, ranking has to be left to the user side.

1) Data Mining Algorithms:

An algorithm in data mining is a set of heuristics and calculations that creates a model from data. The algorithm uses the results of this analysis over many iterations to find the optimal parameters for creating the mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics.

D. Security Intend Computational Encryption

To alleviate the computational burden on user side, computing work should be at the server side, so I need an encryption scheme to guarantee the operability and security at the same time on server side. Homomorphic encryption allows specific types of computations to be carried out on the corresponding cipher text. The result is the cipher text of the result of the same operations performed on the plaintext.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

1) Encoding/Decoding Algorithms

The Rijndael algorithm is a new generation symmetric block cipher that supports key sizes of 128, 192 and 256 bits, with data handled in 128-bit blocks - however, in excess of AES design criteria, the block sizes can mirror those of the keys. Rijndael uses a variable number of rounds, depending on key/block sizes, as follows:
9 rounds if the key/block size is 128 bits
11 rounds if the key/block size is 192 bits
13 rounds if the key/block size is 256 bits

IV. RESULT

Search

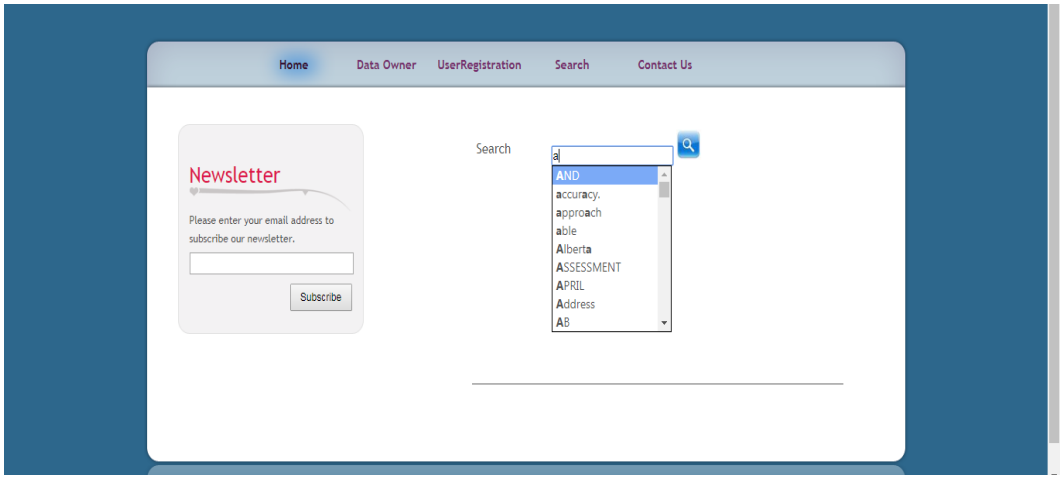

Fig.1. User Search File

Data ownr has a collection of data files F=(F1,F2,F3,Fn) and outsourced them to the cloud server in the encrypted form C.Searching the specific file is shows as above.

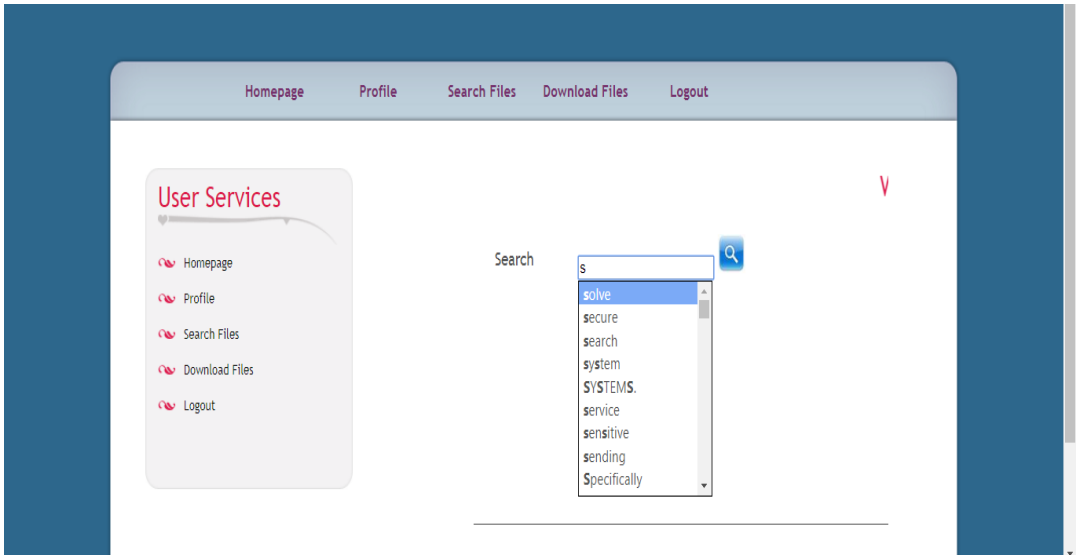

Fig.2. Autofill Implementation

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Multy-keyword fuzzy search scheme should support more spelling mistakes. Inorder to solve this autofill is included as above.

Download File

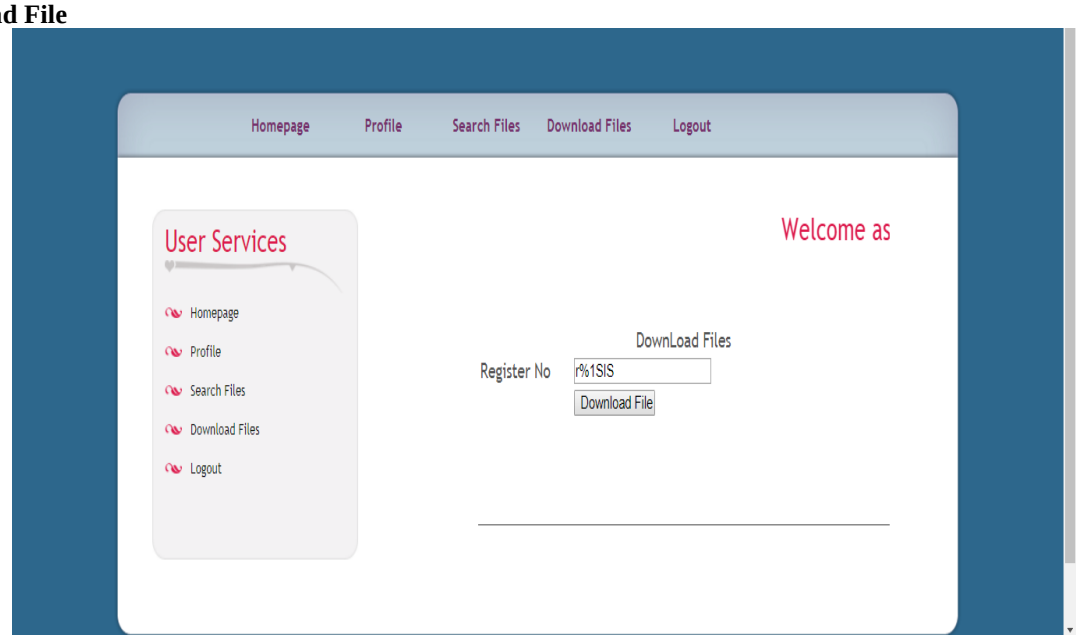

Fig.3 downloading file

File effectively searched and got downloaded by implementing the keyword search algorithm as above.

V. CONCLUSION FUTURE WORK

In this paper, I would investigate the problem of multikeyword fuzzy ranked search over encrypted cloud data. I propose a multi-keyword fuzzy ranked search scheme based on Wang et al.'s scheme. Concretely, I develop a novel method of keyword transformation and introduce the stemming algorithm. With these two techniques, the proposed scheme is able to efficiently handle more misspelling mistake. Moreover, my proposed scheme takes the keyword weight into consideration during ranking.I also give thorough security analyses and conduct experiments on real world data set, which indicates the proposed
scheme's potential of practical usage. My future works can be summarized as follows: When the user's query is a sentence, one can extract the attributes of a sentence, and then express the relationship between attributes and search though the attributes, it is called as semantic search. I failed to achieve the ideal state because of the keyword weight. I will develop a way to reflect the keyword weight and enable update. I will design a verifiable search scheme over encrypted cloud data. Nowadays, many works were mainly focusing on the cases of single data owner and hence not effective for multidataowner. Note that multidataowner scheme has more realistic significance.

REFERENCES

1. J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, "Ensuring security and privacy preservation for cloud data services," ACM Computing Surveys, 2016.
2. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.
3. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

4. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in CryptologyEurocrypt 2004. Springer, 2004, pp. 506–522.
5. Z. Ying, H. Li, J. Ma, J. Zhang, and J. Cui, "Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating," Sci China InfSci, vol. 59, no. 4, pp. 042 701:1–16, 2016.
6. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.
7. E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.
8. Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 442–455.
9. Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Pairing-Based Cryptography–Pairing. Springer, 2007, pp. 2–22.
10. P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.