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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Secure Multi-Keyword Search Scheme for Cloud Data

Vaibhavi Kulkarni, Prof. Priya Pise

Student, Dept. of Computer Engineering, Indira college of Engineering and management, Pune, India

Professor, Dept. of Computer Engineering, Indira college of Engineering and management, Pune, India

ABSTRACT: Cloud computing is today's one of the significant issue. Utilizing cloud computing, one can store, process and manage their data on remote servers. The basic barriers to the storage are the privacy and security. Sensitive data usually has to be encrypted before outsourcing for the data privacy. Accordingly, providing strong security to cloud information admin is of principle significance. This excludes data utilization like keyword-based data retrieval. So, enabling a ciphered cloud documents search service is very useful. While the more cloud users and their huge data, it is very difficult for the search operation to allow multi-keyword search and provide ranked result to get the successful relevant of data which is needed. A multi-keyword search, in cloud computing, underpins dynamic operations like insertion and deletion of documents.

KEYWORDS: Cloud computing; Searchable Encryption; Cipher-text; Multi-keyword ranked search; Security.

I. INTRODUCTION

This is era of Cloud computing. Here the computing resources are shared by many users. The benefits of cloud is to everyone, from individual users to organizations. The data storage in cloud is one of important feature. The virtualization of software and hardware resources in cloud reduces investment for owning the data warehouse and its maintenance. In Cloud computing, data owners are motivated to outsource their huge data management systems from local sites to commercial public cloud for great flexibility and economic benefits. But for protecting privacy of the data, sensitive data has to be encrypted before storing, which obsoletes traditional data utilization based on plaintext keyword search. So, enabling an encrypted cloud data search service is very useful.

While taking into account, the number of data users and documents in cloud, it is very tedious for the search service to provide multi-keyword query and provide result similarity ranking to meet the data retrieval need. Related works on searchable encryption focus on single keyword search, and rarely differentiate the search results. Nonetheless, these ways are not very useful because of their high computational complexity for both the cloud sever and user. On the contrary, more functional precise intent solutions, such as searchable encryption (SE) schemes have made particular contributions in terms of efficiency, functionality and protection. Searchable encryption schemes permits the user to store the encrypted data to the cloud and execute keyword search over cipher-text. So some distance, considerable works have been proposed below exclusive chance items to achieve various search functionality, such as single keyword search, similarity search, multi- key phrase boolean search, ranked search, multi-keyword ranked search, etc.

Among them, multi-keyword ranked search achieves more and more awareness for its functional applicability. Some dynamic schemes are proposed to inserting and deleting operations on report collection. These works possibly that the data owners need to store their knowledge on the cloud server. However few of the dynamic schemes help effective and efficient multi-keyword ranked search.

II. RELATED WORK

In [3] a privacy-aware BedTree based approach to support fuzzy multi-keyword feature is proposed. Incremental updates can be easily done using this solution. The evaluation results show that this approach is more cost-effective in terms of storage size and construction time. The search time is usually better than the wildcard approach for multi-

DOI: 10.15680/IJIRCCE.2017. 0506182

keyword queries where many encrypted files are returned using single word queries for approaches that do not support multi-keyword queries. In paper [2], for the first time the problem of full-scale fuzzy keyword set construction according to the input keyword, and construct the feedback scheme to produce pointer vector including fuzzy keyword, edit distance and keywords dynamic score, which is feasible in hybrid cloud model is solved. Thus, different vectors form the character vector database within its data structure. It can go to the trapdoor construction procedure after access to the database with its edit distances to construct fuzzy keyword set, which makes the fullest use of the retrieval history and statistical misspelled keywords, precisely and quickly realizing the aim of ranked fuzzy keyword search over cloud encrypted data. Thorough rigorous security analyses realize privacy preservation, as well as improvement of the solution which can meet satisfaction needs of users. In [4] public key encryption with keyword search (PKES) enables senders to send encrypted data to a receiver like traditional public key encryption (PKE) schemes. The difference between PKES and PKE is that the receiver in PKES can search on the encrypted data which is stored on the third-party server (like a cloud storage server). Most of the existed PKES schemes are based on bilinear map, so they are costly in computation and hard to be used in practice. In [5] the Dictionary-based Fuzzy Set Construction is presented, in which each keyword is corresponding with much less fuzzy keywords. This improvement greatly reduces the index size, thereby reducing the storage and communication overheads. The experiment results show that when the number of keywords is 4 10, the index size ratio between ours and theirs is 1:3.4(d = 1) and 1:20.4(d = 2). In [6] as Cloud Computing gets to be predominant, more touchy data are being unified into the cloud. Albeit customary searchable encryption plans permit a client to safely seek over encoded through pivotal words and specifically recover documents of interest, these procedures bolster just correct catchphrase look. In this paper, interestingly the issue of viable fluffy pivotal word seeks over encoded cloud information while keeping up magic word security is formalized and taken care of. Fluffy essential word look significantly improves framework ease of use by giving back the coordinating documents when clients seeking inputs precisely coordinate the predefined watchwords or the nearest conceivable coordinating records in light of pivotal word similitude semantics, when careful match fizzles. The alter separation to measure essential words closeness and add to two propelled methods on building fluffy watchword sets, which accomplish advanced capacity and representation overheads. A fresh out of the box new image based tree navigate looking plan is proposed, where a multi-way tree structure is developed utilizing images changed from the came about fluffy catchphrase sets. The proposed arrangement is secure and protection saving, while effectively understanding the objective of fluffy catchphrase look. Broad trial results exhibit the effectiveness of the proposed arrangement. In paper [7] the issue of building a safe distributed storage administration on top of an open cloud infrastructure where the administration supplier is not totally trusted by the client is considered. At an abnormal state, a few architectures that join late and non-standard cryptographic primitives so as to accomplish our objective are depicted. Such a construction modeling would give to both clients and administration suppliers and give a diagram of late advances in cryptography, inspired particularly by distributed storage. In paper [8] the idea of Cloud Computing to accomplish a complete definition of what a Cloud is, is examined, utilizing the principle qualities regularly connected with this worldview in the writing. More than 20 definitions have been considered taking into account the extraction of an agreement definition and additionally a base definition containing the key qualities. This paper gives careful consideration to the Grid worldview, as it is regularly mistaken for Cloud innovations. It additionally depicts the connections and qualifications between the Grid and Cloud application.

III. PROPOSED ALGORITHM

A. *Design Considerations:*

To prevent the cloud server from learning additional information from the dataset, the index tree, and the queries. The specific search privacy requirements are summarized as follows,

1) Index Confidentiality and Query Confidentiality: the underlying plaintext information (including keywords in the index and query, keywords TF values stored in the index, and IDF values of query keywords) should be protected from cloud server;

2) Trapdoor Unlinkability: the cloud server should not be able to determine whether two encrypted queries (trapdoors) are generated from the same search request;

3) Keyword Privacy: the cloud server could not identify the specific keyword in query, index or dataset.

B. *Notations and Preliminaries*

F – Plaintext document collection, denoted as a set of n documents F = (f1 , f2 , ... , fn)
C– Encrypted document collection for F, denoted as C= (c1 ,c2 , ... ,cn).
W – Dictionary, i.e., the keyword set consisting of m keywords, denoted as W = (w1 ,w2 , ... ,wm).
I – searchable encrypted index tree.
T – Unencrypted form of index tree I, which is stored in data owner side.
Du – the index vector stored in node u of index tree.
Q – Query vector generated from search request.
I u – the encrypted form of Du .
T –Encrypted form of Q, which is always called the trapdoor for the search request.

C. *Description of the Proposed Algorithm:*

Vector space model is widely used in plaintext information retrieval, which efficiently represents documents with multi-dimensional vectors. And in this paper, we use the cosine measure together with TF×IDF rule to provide accurate ranking. Here, term frequency (TF) is simply the number of times a given term or keyword appears within a document, and inverse document frequency (IDF) is obtained through dividing the number of documents in the whole collection by the number of documents containing the term. For document vector, each element represents the TF value of corresponding keyword in this document, and for query vector, each element represents the IDF value of corresponding keyword in the dataset. To quantify the similarity of each document vector and the query vector, the deviation of angles (i.e., cosine values) between the two vectors is calculated.

$$\cos(d_i, q) = \frac{d_q \times q}{|d_i||q|}$$

IV. Pseudo Code

GDFS: Greedy Depth First Searches described as estimating the promise of nodenby a ”heuristic evaluation function which, in general, may depend on the description often, the description of the goal, the information gathered by the search up to that point, and most important, on any extra knowledge about the problem domain.”

```
Step 1: OPEN = [initial state]
Step 2: while OPEN is not empty or until a goal is found
        do
Step 3: A. Remove the best node from OPEN, call it n.
        B. If n is the goal state, backtrace path to n (through recorded parents) and return
        path.
        C. Create n's successors.
        D. Evaluate each successor, add it to OPEN, and record its parent. done
Step 4: OPEN = [initial state]
        CLOSED = []
Step 5: while OPEN is not empty
        do
        1. Remove the best node from OPEN, call it n, add it to CLOSED.

        2. If n is the goal state, backtrace path to n (through recorded parents) and return
           path.
        3. Create n's successors.
        4. For each successor do:
                a. If it is not in CLOSED and it is not in OPEN: evaluate it, add it to OPEN, and
                record its parent.
        b. Otherwise, if this new path is better than previous one, change its recorded parent.
                i. If it is not in OPEN add it to OPEN.
                ii. Otherwise, adjust its priority in OPEN using this new evaluation.
done.
```

V. Results

The primary execution measurements used to assess the proposed systems are question reaction time and encryption time. Fig.1 Show the reaction time measures the length of time from the time the question is issued until the outcomes are gotten at the customer. It gives the calculation time at the server and the customer, still on the grounds that the time required for exchange of last and transitional results in the middle of customer and server.

File size in KB	[query Process time using Previous technique] DES	query Process time using AES
6000	44	22
7500	54	31
8900	66	40
10000	77	53
12000	88	70

Fig.1. Behavior of DES and AES Algorithm

VI. Conclusion and Future Work

Paper solved the problem of multi-key word ranked search over encrypted cloud data, and set up a range of privacy requirements. From various multi-keyword semantics, the efficient similarity measure of coordinate matching, i.e., as many equivalent as possible is selected, to effectively capture the relevance of outsourced documents to the query Keywords, and utilize inner product similarity to quantitatively calculate such comparison measure. The further enhancements of our ranked search method, including supporting more search semantics, i.e., TF-IDF, and dynamic data process. Detailed analyses in investigating privacy and efficiency assurance of proposed schemes are mentioned, and testing on the real-world data set demonstrate our proposed schemes which introduces low transparency on both calculation and communication. The proposed algorithm gives energy efficient way for data transmission and enhancing the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. As number of nodes increases the complexity in the network. By increasing number of nodes and one can analyze the performance.

REFERENCES

1. Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang, Member, IEEE, “A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data ” , IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS 2016.
2. Wang Jie, Yu Xiao, Zhao Ming, Wang Yon, A Novel Dynamic Ranked Fuzzy Key-word Search Over Cloud Encrypted Data, 2014, IEEE 12th International Conference on Dependable, Autonomic and Secure Computing.
3. M. Chuah, W. Hu, Privacy-aware BedTree Based Solution for Fuzzy Multi-keyword Search over Encrypted Data, 2011 31st International Conference on Distributed Computing Systems Workshops.
4. Wenjun Luo, Jianming Tan, PublicKey Encryption withKeyword Search Based On Factoring, Proceedings of IEEE CCIS2012
5. Chang Liu, Liehuang Zhu, Longyijia Li, Yuan Tan, FuzzyKeyword Search On Encrypted Cloud Storage Data With Small Index, 2011, Proceedings of IEEE CCIS2011.
6. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, Fuzzy keyword search over encrypted data in cloud computing, in INFOCOM, 2010 Proceedings IEEE.
7. S . Kamara and K. Lauter, Cryptographic cloud storage, in RLCPS, January2010, LNCS. Springer, Heidelberg.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

8. L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, A break in the clouds: towards a cloud definition, ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 5055, 2009.

9. D. Boneh and B. Waters, “Conjunctive, subset, and range queries on encrypted data” , in Proceedings of the 4th conference on Theory of cryptography. Springer-Verlag, 2007, pp. 535554.

10. B. Zhang and F. Zhang, “An efficient public key encryption with conjunctive subset keywords search” , Journal of Network and Computer Applications, vol. 34, no. 1, pp. 262267, 2011.

11. J. Katz, A. Sahai, and B. Waters, “Predicate encryption supporting disjunctions, polynomial equations, and inner products”, in Advances in Cryptology EUROCRYPT 2008. Springer, 2008, pp. 146162.