Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 6, June -2017

Fuzzy Keyword Search for MultipleData Owners in Cloud Computing

[1]Sayali Sabale, [2]Mrs. Shanthi. K. Guru

[1,2]Department of Computer Engineering,D.Y. Patil College of Engineering, Akurdi, Pune -411044

Abstract- The huge amount of data outsourced every day by individuals or each enterprises. When user need to store their data in such a way that it can be accessed uninterruptedly, then the cloud comes into picture to store the data with better flexibility and cost saving. As the data might be confidential or sensitive. Considering the privacy searchable encryption can be used. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. That is, there is no tolerance of minor typos and format inconsistencies which, on the other hand, are typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in Cloud Computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficacy very low. 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. Calculate edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads.

Keyword:- *Cloud computing, searchable encryption, multi-keyword search, fuzzy search, user revocation, multiple users*

1. INTRODUCTION

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as emails, personal health records, government documents, etc. By storing their data into the cloud, the data owners can be relieved from the burden of data storage and maintenance so as to enjoy the on-demand high quality data storage service. However, the fact that data owners and cloud server are not in the same trusted domain may put the outsourced data at risk, as the cloud server may no longer be fully trusted. It follows that sensitive data usually should be encrypted prior to outsourcing for data privacy and combating unsolicited accesses. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. 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 keywords usually contain important information related to the data files [1].

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. 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. This technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzy keyword sets is significantly reduced. 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].

Stemming Algorithm

A stemming algorithm [4] 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". A stemming algorithm reduces the words "fishing", "fished", and "fisher" to the root word, "fish". On the otherhand, "argue", "argued", "argues", "arguing", and "argus" reduce to the stem "argu" (illustrating the case where the stem is not itself a word or root). It is widely adopted in Information Retrieval systems to improve performance.

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 $S = \{a1, a2,..., an\}$, 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. To check whether an element q is in S, feed it to each of the l hash functions to get l array positions. If the bit at any position is 0, q /∈S; otherwise, either $q \in S$ or q yields a false positive [5].

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 hash function family H is $(r1, r2, p1, p2)$- sensitive if any two points x, y and satisfy:
If $d(x, y) \leq r1$; $Pr\ [h(x) = h(y)] \geq p1$ (1)
If $d(x, y) \geq r2$; $Pr\ [h(x) = h(y)] \leq p2$ (2)
Where $d(x, y)$ is the distance between the point x and the point y [6].

System Model:

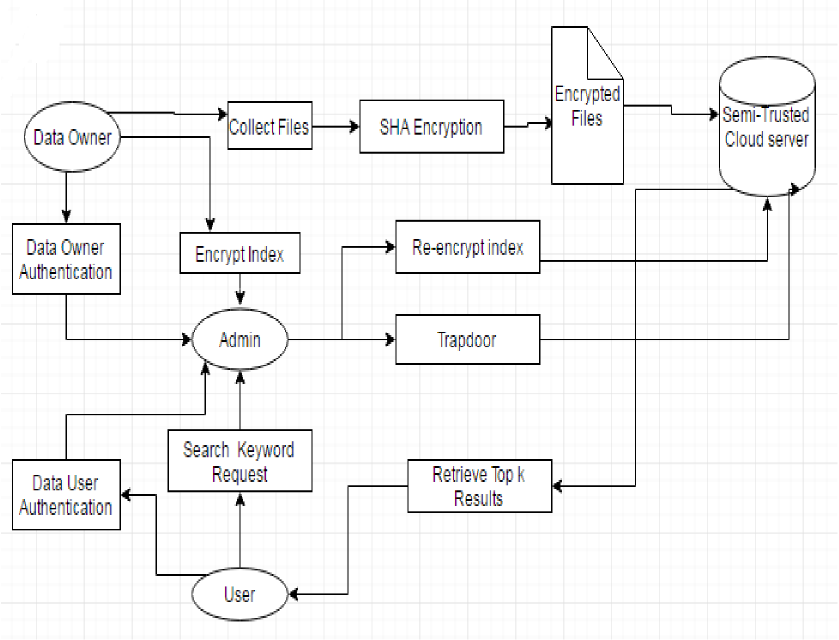


Fig. 1 System Architecture [1]

In multi-owner and multi-user cloud computing model, four entities are involved, as shown in system architecture they are data owners, the cloud server, administration server, and data users. In multi-owner and multi-user cloud computing model, four entities are involved, as shown in system architecture they are data owners, the cloud server, administration server, and data users. Data owners have a collection of files F. To enable efficient search operations on these files which will be encrypted, data owners first build a secure searchable index I on the keyword set W extracted from F, then they submit I to the administration server. Finally, data owners encrypt their files F and outsource the corresponding encrypted files C to the cloud server. Upon receiving I, the administration server re-encrypts I for the authenticated data owners and outsources the re-encrypted index to the cloud server. Once a data user wants to search keywords over these encrypted files stored on the cloud server, he first computes the corresponding trapdoors and submits them to the administration server. Once the data user is authenticated by the administration server, the administration server will further re-encrypt the trapdoors and submit them to the cloud server. Upon receiving the trapdoor T, the cloud server searches the encrypted index I of each data owner and returns the corresponding set of encrypted files.

Modules and Results
Data Owner Module
In this module, the data owners should be able to upload the files. The files are encrypted before the files are uploaded to the cloud. The data owners are provided an option to enter the keywords for the file that are uploaded to the server. These keywords are used for the indexing purpose which helps the search return values very quickly. The data owners will also be provided with a request approval screen so they are able to approve or reject the request that are received by the data

users. The data owners should be able to upload the files. The files are encrypted before the files are uploaded to the cloud. The file before upload will have to be encrypted with a key so that the data users cannot just download it without this key. This key will be requested by the data users through the trap-door. The encryption of these files uses SHA algorithm so that unauthorized users will not be able to download these files.

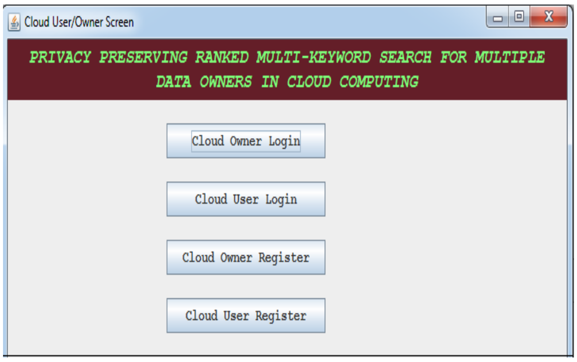


Fig 2. Home Screen

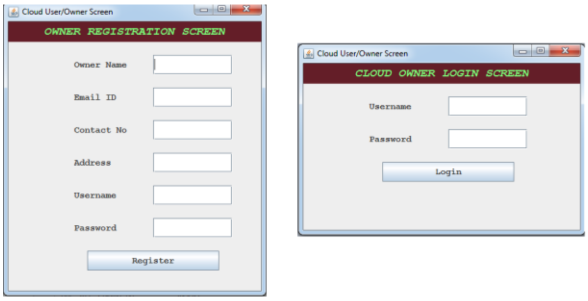
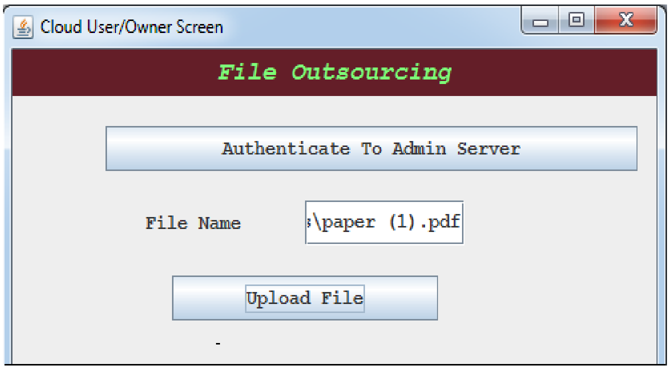


Fig. 3 Data Owner Registration and Login



:\paper (1).pdf

Fig. 4 File Upload

Data User Module

Data users are users on this system, who will be able to download files from the cloud server that are uploaded by the data owners. Since the files stored on the cloud server could be in huge numbers, there is a search facility provided to the user. Once, the result appears for the specific search, these users should be able to send a request to the respective data owners of the file through the system (also called trap-door request) for downloading these files. The data users will also be provided a request approval screen, where it will notify if the data owner has accepted or rejected the request. If the request has been approved, the users should be able to download the decrypted file.

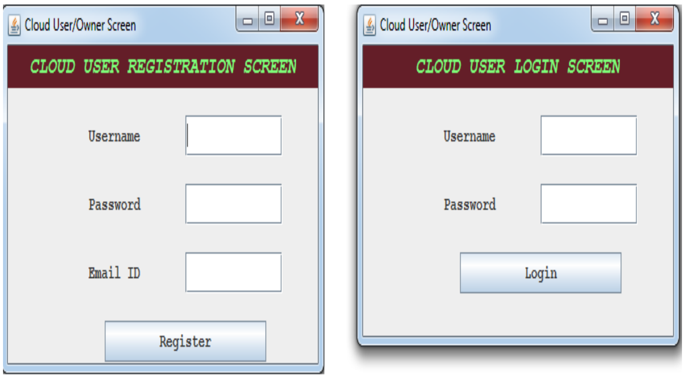
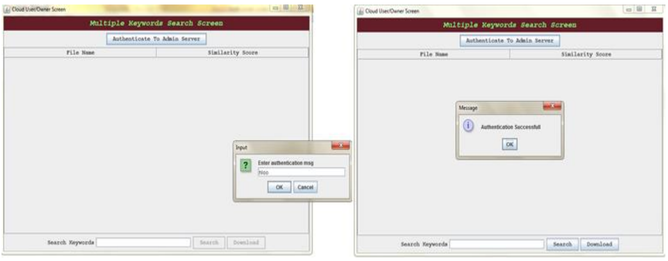


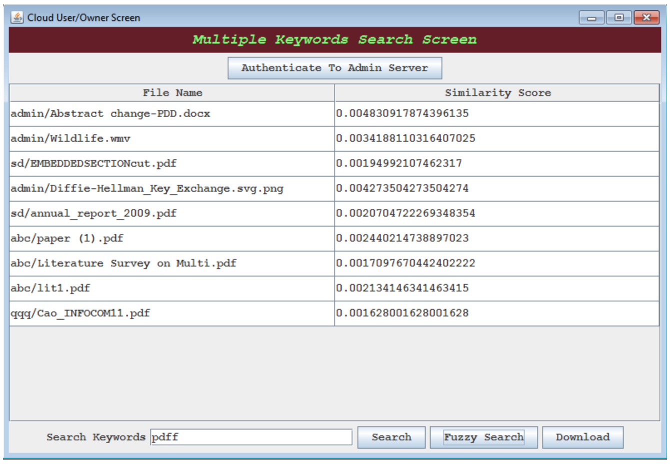
Fig. 5Data User Registration and Login



Enter authentication msg

Authentication Successful

Fig. 6 Search



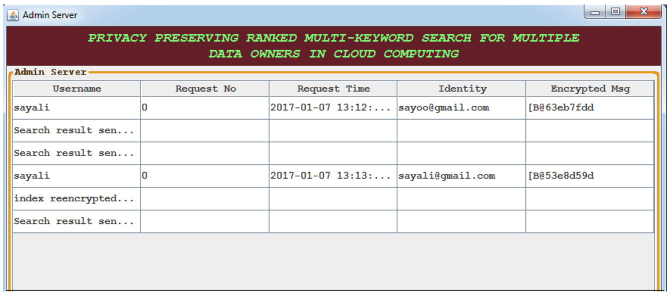
File Name	Similarity Score
admin/Abstract change-PDD.docx	0.004830917874396135
admin/Wildlife.wmv	0.0034188110316407025
sd/EMBEDDEDSECTIONcut.pdf	0.00194992107462317
admin/Diffie-Hellman_Key_Exchange.svg.png	0.004273504273504274
sd/annual_report_2009.pdf	0.0020704722269348354
abc/paper (1).pdf	0.002440214738897023
abc/Literature Survey on Multi.pdf	0.0017097670442402222
abc/lit1.pdf	0.002134146341463415
qqq/Cao_INFOCOM11.pdf	0.001628001628001628

Search Keywords pdff

Fig. 6 Result

Cloud Server

Cloud Server will stored encrypted documents and re-encrypted index. User will send search request to the admin server. Admin server will encrypt that keyword and trapdoor will get generate. Cloud server will search the documents using that trapdoor and display top k results to the user using ranking algorithm. This module also takes care of creating an index for faster search.


Admin Server
PRIVACY PRESERVING RANKED MULTI-KEYWORD SEARCH FOR MULTIPLE DATA OWNERS IN CLOUD COMPUTING
Admin Server

Username	Request No	Request Time	Identity	Encrypted Msg
sayali	0	2017-01-07 13:12:...	sayoo@gmail.com	[B@63eb7fdd
Search result sen...				
Search result sen...				
sayali	0	2017-01-07 13:13:...	sayali@gmail.com	[B@53e8d59d
index reencrypted...				
Search result sen...				

Fig. 5 Cloud Server

Conclusion:-

Fuzzy keyword search over encrypted cloud data for multiple data owner construct a variety of security requirements. From various multi-keyword concepts, we developed the efficient principle of similarity matching. Multiple-keyword ranked search schemes enable authenticated data users to achieve secure, convenient, and efficient searches over multiple data owners data. To efficiently authenticate data users and detect attackers who steal the secret key and perform illegal searches, and a new data user authentication protocol. To enable the cloud server to perform secure search among multiple owners data encrypted with different secret keys, systematically construct a novel secure search protocol. To rank the search result and preserve the privacy of relevance scores between keywords and files, a novel Additive Order and Privacy Preserving Function family. This approach is computationally efficient, even for large data and keyword sets.

References

[1] Wei Zhang, Student Member, Yaping Lin, Sheng Xiao, JieWu, Fellow, and Siwang Zhou,"Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing," IEEE TRANSACTIONS ON COMPUTERS, VOL. 65, NO. 5, MAY 2016

[2] Zhihua Xia, Member, Xinhui Wang, Xingming Sun and Qian Wang, Member, IEEE, “A Secure and Dynamic Multi-keyword Ranked SearchScheme over Encrypted Cloud Data”, IEEE transactions on Parallel and Distributed systems,2015

[3] Privacy Preserving String Pattern Matching on Outsourced Data, Bargav Jayaraman

[4] Qin Liuy, Guojun Wangyz, and Jie Wuz, “Secure and privacy preserving keyword searching for cloud storage services”, ELSEVIER Journal of Network and computer Applications, March 2011

[5] Cong Wang et al., “Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 23, no. 8, August 2012

[6] International Journal of Computer Applications (0975 – 8887) Volume 126 – No.14, September 2015

[7] Wenhai Sun et al., "Privacy-Preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking", the 8th ACM Symposium on Information, Computer and Communications Security, Hangzhou, China, May 2013.

[8] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Member, IEEE Computer Society, and Minglu Li,” Toward Secure Multi-keyword Top k Retrieval over Encrypted Cloud Data”, IEEE Journal of Theoretical and Applied Information Technology 10th August 2014. Vol. 66 No.1 © 2005 - 2014 JATIT & LLS. All rights reserved. ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195 64 Transactions on dependable and secure computing, vol. 10, no. 4, July/August 2013

[9] Ning Cao et al., “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, jan 2014

[10] Wenhai Sun et al., "Protecting Your Right: Attribute-based Keyword Search with Fine-grained Owner-enforced Search Authorization in the Cloud", IEEE INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014

[11] Secure Ranked Keyword Search over Encrypted Cloud Data, IEEE PAPER, 2010.

[12] Sudha et al., “A Survey on Encrypted Data Retrieval in Cloud Computing”, International Journal of Advanced Research in Computer Science and Software Engineering 5(1), January - 2015, pp. 895-899

[13] Zhangjie Fu, Member, IEEE, Xingming Sun, Senior Member, IEEE, Nigel Linge, Lu Zhou, Achieving Effective Cloud Search Services :Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014

[14] C. Wang, N. Cao, K. Ren, and W. Lou, Enabling Secure and Efficient Ranked Keyword Search Over Outsourced Cloud Data, IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8,pp. 1467-1479, Aug. 2012.

[15]H. S. Rhee, J. H. Park, W. Susilo, Trapdoor security in a searchable public-key encryption scheme with a designated tester," Journal of Systems and Software, vol. 83, no. 5, pp.763771, 2010.