Multi-Keyword Ranked Search over Encrypted Cloud Data with Multiple Data Owners

Kalyani Sonawane
Department of Computer Engineering
Marathwada Mitra Mandal’s College of Engineering, Pune

***Abstract*— Clients usually outsource their data to the cloud storage and enjoy on-demand services provided by cloud server. Cloud data may contain sensitive personal information. To provide data security and services to data user in multi-owner model is a challenging task. It is essential to perform secure search without knowing actual data and trapdoors and provide ranked results to data user. This paper proposes a dynamic secret key generation protocol and user authentication protocol to prevent attackers for performing illegal search. We develop a new method of keyword transformation based on the uni-gram which is used to handle spelling mistakes.**

***IndexTerms*— Cloud computing, data outsourcing, cloud storage, data utilization.**

I. INTRODUCTION

In recent years, Cloud computing is gaining much momentum in the IT industry which can be used to organize various resources of computing, storage and applications. Many IT enterprises and individuals are outsourcing their databases to cloud server. Variety of users can access and share information stored in the cloud independent of locations. The outsourced data may contain very sensitive information such as e-mails, company financial data, government documents, Personal Health Care records, facebook photos and business documents.

Cloud service providers (CSPs) can access user’s sensitive data without any authorization. General approach of CSPs is to protect the data confidentiality in which data is encrypting before outsourcing it to cloud server and this will affect a huge cost of data usability. In secure search over encrypted data, data owners outsourced their data to cloud server in encrypted form to preserve their privacy. When data user wants to search any file, data user send keyword request to cloud server. Cloud server then generate top relevant results to data user. Secure search over encrypted data is shown in following figure 1.

Secure search over encrypted data not only reduce computation cost and storage cost for secure keyword search but also support multi-keyword ranked search, fuzzy keyword search and similarity search. All these schemes are limited to single-owner model. Earlier work support single-owner model, where data owner has to stay online to generate trapdoors for data user. Therefore, this paper proposes a multi-owner model to overcome the limitations of the earlier methods, where encrypted data are stored by multiple data owners and simultaneously data owners stay online to generate trapdoors. Different data owners share different secret keys to encrypt their secret data with different secret keys.

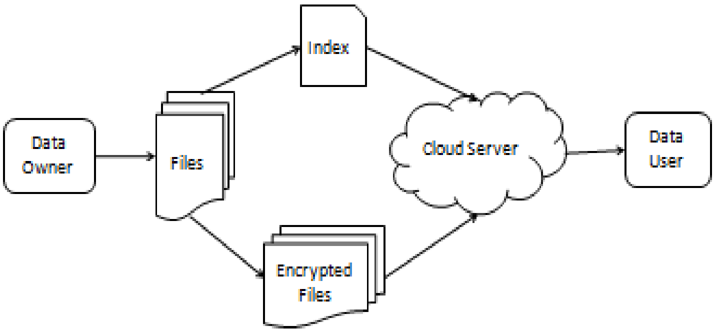


Fig.1. Secure Search over Encrypted Data

In this paper, secure search protocol is propose in which cloud server can perform secure search without knowing the actual value of keywords and trapdoors. Different data owners use different secret keys to encrypt files and keywords which are outsource to cloud server. Authenticated data users can send encrypted query without knowing secret keys of data owners. A novel dynamic secret key generation protocol and a new data user authentication protocol use to prevent the unauthorized users from accessing cloud data.

II. LITERATURE SURVEY

Secure search over encrypted data have been previously applied to cloud server. Wang et al. [20] proposed secure search scheme over encrypted cloud data. In searchable encryption, clients store data into encrypted form to the cloud server and keyword searching can be perform on ciphertext. Searchable encryption (SE) techniques [5], [6], [7], [8], [14], [15], [16], [17], [18] can partially fulfill the need for secure outsourced data search.

Secure search over encrypted cloud data reduces the computation and storage cost. Secure ranked multi-keyword search, fuzzy keyword search, similarity search all these searches are also performed on encrypted cloud data. Data user authentication technique, Different-key encrypted keywords matching and privacy preserving ranked search of files methods are used to solve the problem of secure multi-keyword search for multiple data owners and multiple data users in cloud computing environment. When huge amount of data owners [3], [9] are involved then they generate trapdoors simultaneously which affect the flexibility and usability of search system.

A. Data User Authentication Technique:

Data user authentication technique is used to prevent system from attackers who pretending to be legal data users performing searches. Ming Li [3] proposed fine-grained authorization framework in which user obtains search capabilities under local trusted authorities (LTAs). Third party auditors (TPA) used to authenticate data user before performing any searching on cloud server [4]. Another technique to provide security against attackers is user revocation [3], [9], [11] where data user cannot perform any searches once he is revoked.

B. Matching Different-Key Encrypted Keywords:

Early works mostly only support single keyword search
[17]. Later, several multi-keyword search schemes were proposed [5], [6], [7], [8], [12], [13], [15]. Data owner store data in encrypted form and data user generate trapdoors [3], [4], [19] to send query request in encrypted form. Re-encryption of keyword index and trapdoors [9], [11] used to increase more security from attackers.

Wenhai Sun [5], proposed tree-based index structure so that practical search efficiency is much better than linear search. Ning Cao [6], proposed coordinate matching which provides as many matches as possible which capture the relevance of data documents to the search query and inner product similarity to quantitatively evaluate such similarity measure. Zhihua Xia [11] proposed a scheme which supports dynamic update operations like deletion of documents and insertion of documents and tree-based index structure and Greedy Depth first Search algorithm use to provide efficient multi-keyword ranked search. Hongwei Li [12] support complicated logic search by using the mixed AND, OR and NO operations of keywords for practical and very efficient multi-keyword search scheme. [22] proposed problem of personalized multi-keyword ranked search over encrypted cloud data. A user interest model is build for individual user with the help of semantic ontology WordNet by using user search history.

C. Privacy Preserving Ranked Search:

In searchable symmetric encryption schemes, due to large number of documents, search results should be retrieved in an order of the relevancy with the searched keywords. Scoring is the natural way to weight the relevancy of the documents. TFIDF [4], [6], [7], [8], [19] is well-known method to compute the relevance score. Wong et al. [13] proposed a secure k-nearest neighbor (kNN) scheme which can confidentially encrypt two vectors and compute Euclidean distance of them [6], [8], [11], [12], [20].

III. PROPOSED SYSTEM

A. Problem Statement

To design multi-keyword ranked search scheme for multiple data owners and multiple data users over encrypted mobile cloud data which provides accuracy, efficiency and secure search.

B. System Architecture

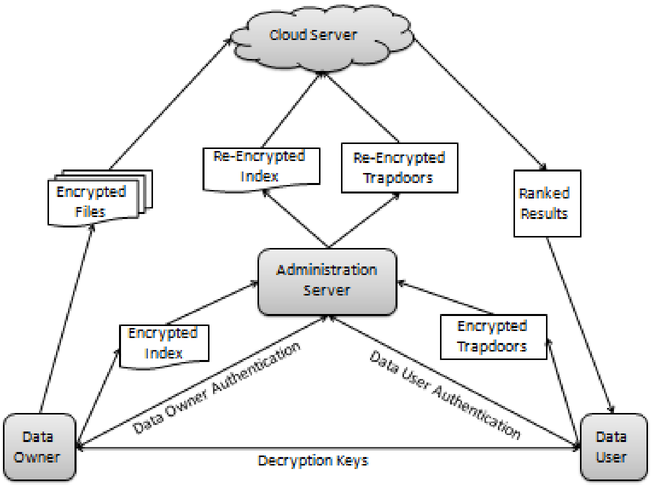


Fig.2. System Architecture

In multi-owner and multi-user cloud computing model, four entities are involved such as data owners, data users, cloud server and administration server. Data owners have collection of files. Data owners build secure searchable index of keyword set and keywords are extracted from files. Data owners submit keyword index to administration server. Data owners encrypt files and outsource encrypted files to cloud server. When administration server receives encrypted keyword index then administration server re-encrypt keyword index. Administration server then outsource re-encrypted keyword index to the cloud server.

When data user wants to search over files from cloud server, he first computes the corresponding trapdoors and submits them to the administration server. Administration server authenticates data user then re-encrypts trapdoors and submit them to cloud server. Cloud server searches encrypted index of data owner and returns top-k relevant encrypted files to the data user. When data user receives top-K files from cloud server, then data user download files and decrypts these files.

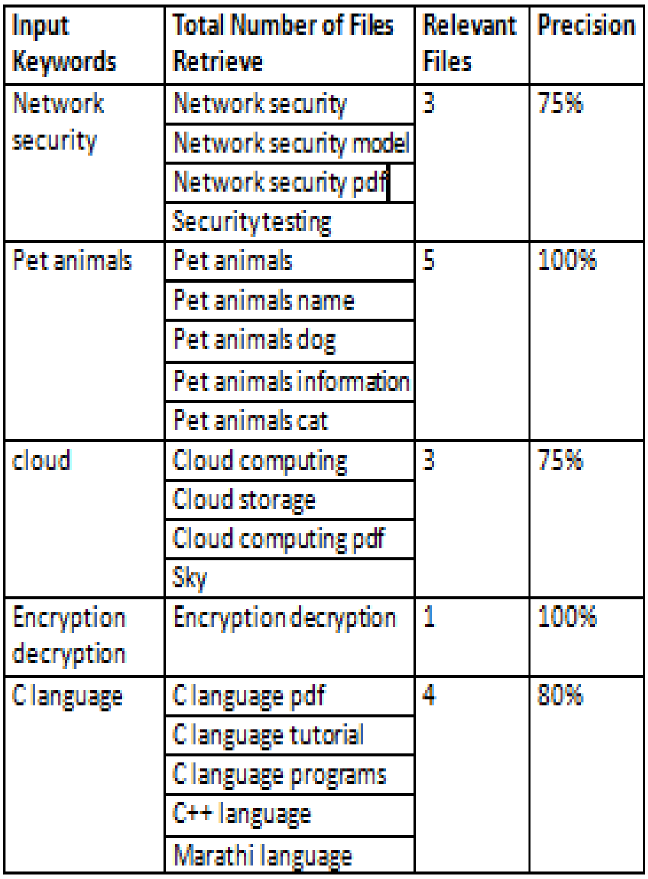
IV. INPUT, OUTPUT AND SYSTEM ACCURACY

INPUT: Data owners have a collection of files. Data owners encrypt these files and extract keywords from files. Encrypted files and re-encrypted keyword index are input to the cloud server.

OUTPUT: We conducted experiments on real dataset, and then compute keyword frequency in each file, the length of each file, the number of files containing specific keywords. Then calculate relevance score of a keyword to retrieve top-k encrypted files. Time cost of index construction, time cost of trapdoor generation, time cost of administration server and time cost of search can be calculated to compare performance with existing system. To provide multi-keyword ranked search for effective utilization of outsourced and encrypted cloud data, system should support following security and performance measures.

SYSTEM ACCURACY: We used precision to measure the result accuracy. To generate the fuzzy search, we randomly chose keywords and modified it into a fuzzy keyword. An important parameter in our scheme is the number of the keywords in the query. The main reasons for the improved the accuracy was the use of the new method for the keyword transformation and the threshold T.

The precision of the exact match slightly decreased as the number of the query keywords increased from 1 to 10. However, the precision slowly grew, from fig. 3 as the query keywords increase from 2 to 10.



Input Keywords	Total Number of Files Retrieve	Relevant Files	Precision
Network security	Network security	3	75%
	Network security model		
	Network security pdf		
	Security testing		
Pet animals	Pet animals	5	100%
	Pet animals name		
	Pet animals dog		
	Pet animals information		
	Pet animals cat		
cloud	Cloud computing	3	75%
	Cloud storage		
	Cloud computing pdf		
	Sky		
Encryption decryption	Encryption decryption	1	100%
C language	C language pdf	4	80%
	C language tutorial		
	C language programs		
	C++ language		
	Marathi language		

Table 1 Precision Table

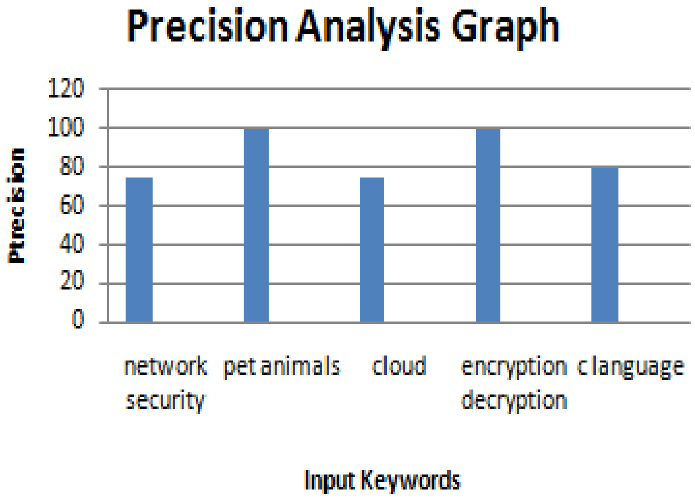


Fig.3. Precision Analysis Graph

V. CONCLUSION

This research presents a secure multi-keyword search for multiple data owners and multiple data users in the cloud computing environment. Dynamic secret key generation and a new data user authentication algorithms are use to authenticate data users and detect attackers who perform illegal searches. Secure search protocol is use to enable the cloud server to perform secure search among multiple owners data encrypted with different secret keys. We developed a novel method of keyword transformation and introduce the stemming algorithm. With these techniques, the proposed scheme is able to efficiently handle more misspelling mistake. Our proposed scheme takes the keyword weight into consideration during ranking.

REFERENCES

[1] Wei Zhang, Yaping Lin, Sheng Xiao, Jie Wu, 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] Kalyani Sonawane, Rahul Dagade, “A Survey on Multi-Keyword Ranked Search over Encrypted Cloud Data with Multiple Data Owners”, International Journal of Computer Applications(0975-8887), Volume 162 No 11, March 2017.

[3] Ming Li, Shucheng Yu, Ning Cao, Wenjing Lou, “Authorized Private Keyword Search over Encrypted Data in Cloud Computing”, 31st International Conference on Distributed Computing Systems, 2011.

[4] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, Wenjing Lou, “Privacy-Preserving Public Auditing for Secure Cloud Storage”, IEEE Transactions on Computers, Vol. 62, No. 2, February 2013.

[5] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, “Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking”, IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 11, November 2014.

[6] Ning Cao, Cong Wang, Ming Li, Kui Ren, Wenjing Lou, “Privacy Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 1, January 2014.

[7] Zhangjie Fu, Xingming Sun, Zhihua Xia, Lu Zhou, Jiangang Shu, “Multi-keyword Ranked Search Supporting Synonym Query over Encrypted Data in Cloud Computing”, 2013.

[8] Zhangjie Fu, Xingming Sun, 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.

[9] Wenhai Sun, Shucheng Yu, Wenjing Lou, Y. Thomas Hou, Hui Li, “Protecting Your Right: Verifiable Attribute-Based Keyword Search with Fine-Grained Owner-Enforced Search Authorization in the Cloud”, IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 4, April 2016.

[10] Zhangjie Fu, Jiangang Shu, Xingming Sun, Nigel Linge, “Smart Cloud Search Services: Verifiable Keyword-based Semantic Search over Encrypted Cloud Data”, IEEE Transactions on Consumer Electronics, Vol. 60, No. 4, November 2014.

[11] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang, “A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data”, IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 2, February 2016.

[12] Hongwei Li, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, Xuemin (Sherman) Shen, “Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data”, IEEE Transactions on Dependable and Secure Computing, Vol. 13, No. 3, May/June 2016.

[13] Bing Wang, Wei Song, Wenjing Lou, Y. Thomas Hou, “Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee”, 2015 IEEE Conference on Computer Communications (INFOCOM).

[14] Wenhai Sun, Xuefeng Liu, Wenjing Lou, Y. Thomas Hou, Hui Li, “Catch You If You Lie to Me: Efficient Verifiable Conjunctive Keyword Search over Large Dynamic Encrypted Cloud Data”, 2015 IEEE Conference on Computer Communications (INFOCOM).

[15] Hongwei Li, Dongxiao Liu, Kun Jia, Xiaodong Lin, “Achieving Authorized and Ranked Multi-keyword Search over Encrypted Cloud Data”, IEEE ICC 2015-Communication and Information Systems Security Symposium.

[16] Wei Zhang, Yaping Lin, “Catch You if You Misbehave: Ranked Keyword Search Results Verification in Cloud Computing”, VOL. 6, NO. 1, JANUARY 2015.

[17] R. Curtmola, J. A. Garay, S. Kamara, R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions”, In Proc. of ACM CCS 06, 2006.

[18] C. Wang, N. Cao, J. Li, K. Ren, W. Lou, “Secure ranked keyword search over encrypted cloud data”, in Proc. IEEE Distributed Computer System, Genoa, Italy, Jun. 2010, pp. 253262.

[19] Hongwei Li, Dongxiao Liu, Yuanshun Da11i, Tom H. Luan, Xuemin Shen, “Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage”, Transaction On Emerging Topics in Computing, 6 March, 2015.

[20] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, Secure kNN computation on encrypted databases”, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139-152.

[21] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data”, in Proc. IEEE Distrib. Comput. Syst., Genoa, Italy, Jun. 2010, pp. 253262.

[22] Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, Fengxiao Huang, “Enabling Personalized Search over Encrypted Outsourced Data With Efficiency Improvement”, IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 9, September 2016.