Journal of Engineering and Applied Sciences 13 (1): 32-37, 2018
ISSN: 1816-949X

Secure Ranked Keyword Search Algorithm for Encrypted Cloud Data using Double Indexing Technique

Payal V. Kale and Rashmi Welekar
Department of Computer Science Engineering,
Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

Abstract: With the emergence of cloud computing, most of the data owners are outsourcing their corpus data from local sites to the commercial public cloud for great flexibility and economic savings. But tactful data has to be encrypted before outsourcing for preserving data privacy and also encrypted data gives an effective data utilization services which are a challenging task. However, traditional searchable encryption techniques use Boolean search to search data via. keywords but this technique not satisfied the effective data utilization for multi-user data need and also to retrieved massive data files from the cloud. This problem is solving by using ranked keyword search over the encrypted cloud data which improves system usability and file retrieval accuracy. Further to protect the score information one-to-many order-preserving mapping techniques has developed and explored the relevance score from information retrieval to build a secure searchable index. In this study, we used double indexing concept to improve retrieval speed of search request. Also, score dynamics concept is explored for to make an update in encrypted cloud files. In this study to improve the efficiency and security, we proposed a system where we will be adding a physical layer security component which will help the system to be more secure and less vulnerable to attacks.

Key words: Ranked keyword based search, score dynamics, order preserving mapping, cloud computing, concept based search, IMEI based security

INTRODUCTION

Cloud computing term refer to performing computing tasks using services delivered completely over the internet. Cloud computing is a movement away from applications needing to be installed on an individual's computer towards the applications being hosted online. Cloud computing provides incredible flexibility. Professionals can work from any computing device anywhere as long as they have access to the Web.In this model a wide pool of systems is linked in private or public networks to provide dynamic extendable infrastructure for application, data and file storage that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud computing financially empowers the worldview of information administration outsourcing. In any case, to ensure information security, thoughtful cloud information must be encoded before outsourced to the business open cloud which makes compelling information use benefit an exceptionally difficult assignment. For positioned keyword based search and idea based pursuit in scrambled cloud information, Order Preserving Encryption (OPE) is an effective device to encode relevance scores of the upset record. In this way, delicate information must be encoded before being outsourced to a business public cloud. By utilizing an updated score, we can recover the document from cloud information to an interested outsider.

Along these lines, tactful information must be encoded before being outsourced to a business open cloud. Encryption of tactful information presents snags to the preparing of the information. Data recovery winds up noticeably troublesome in the scrambled area on the grounds that various outsourced documents can be expensive and conventional inquiry designs can't be sent to figure content recovery specifically. Users need to download every one of the information, decode it all and after that search term like plaintext recovery. To defeat this, Searchable Encryption (SE) was proposed to make a question in the encoded space conceivable while as yet protecting client's security.

First proposed a search scheme only supporting single Boolean keyword search. After that plenty of searchable encryption methods arose to improve efficiency and reduce communication overhead. Ranked

Corresponding Author: Payal V. Kale, Department of Computer Science Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

Keyword based search method is not as efficient method while searching the keyword other than stored keyword, so in this study, we introduced the concept of concept based search algorithm with the help of this algorithm we can search the relevant search query result out of the search indexed keyword. In the proposed study, we used a double indexing method to efficiently retrieve the relevant file and also for improving the speed of retrieval.

Literature reivew: Wang *et al*. (2012) and Ren *et al*. (2012) define and solve the problem of secure ranked keyword search over encrypted cloud data and also security related problem. Ranked search greatly enhances system usability by enabling search. They motivated and solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored encrypted data in cloud computing (Subashini and Kavitha, 2011), a survey of the different security risks that pose a threat to the cloud is presented. They specify more specific to the different security issues that has emanated due to the nature of the service delivery models of a cloud computing system.

Boldyreva *et al*. (2009) they propose a simple and efficient transformation that can be applied to any OPE scheme. Our analysis shows that, the transformation yields a scheme with improved security (Xiao and Yen, 2012), they analyze the security of the OPE encryption scheme SE m, n and give the upper bound on the probability for the adversary to recover the plain text encrypted by SE m, n under chosen plain text attacks. Agrawal *et al*. (2004) describes the order preserving encryption algorithm for numeric data (Goh, 2003) they build secure index for documents. This secure index allows an authorized user to search for an encrypted file which contains the keyword without decrypting that file.

Most concept-based methods automatically derive user's topical interests by exploring the contents of the user's browsed documents and search histories, Liu *et al*. (2002) proposed a user profiling method based on user's search history and the Open Directory Project (ODP). The user profile is represented as a set of categories and for each category, a set of keywords with weights. The categories stored in the user profiles serve as a context to disambiguate user queries. If a profile shows that a user is interested in certain categories, the search can be narrowed down by providing suggested results according to the user's preferred categories. Speretta and Gauch, (2005) and Gauch *et al*. (2003) proposed a method to create user profiles from user browsed documents. Privacy of such keywords are given by Xu *et al*. (2007).

Swetha and Narasinga (2016) also explored with the storing of encrypted data on distributed cloud storage. Ellaji and Shobha (2015) proposed the security relaed issues in keyword search over encrypted cloud data. Preservving privacy of keywords are explained by Chang and Mitzenmacher (2005).

MATERIALS AND METHODS

System Architecture of the proposed researcher is divided in three parts such as data owner, cloud server, data user (Fig. 1).

Data owner deals with collection of n data files those file data owner wants to be outsource on the cloud server. Before outsourcing that file on the cloud server, initially he builds a secure, searchable index from a set of m distinct, unique words which has been extracted from the collection of data files and store both the index and the encrypted file collection on the cloud server.

These files then outsource on the cloud server for keeping the capability to search through them for effective data utilization. We consider a server which is direct and anomalous in our model is persistent with most of the earlier searchable encryption techniques. We assume the cloud server work in a direct fashion and precisely follow the depute protocol specification, but are anomalous to conclude and analyze the sequence of processing step received during the protocol so as to grasp extra information.

So, as the cloud server has no intention to presently change the sequence of processing step or disrupt any other kind of services. However, in some unexpected events, the cloud server may behave beyond the direct and anomalous model. To search a given keyword w from the file collection an user which has authorization establish and give in a search request in a secret form, i.e., a trapdoor Tw of the keyword w to the cloud server. On

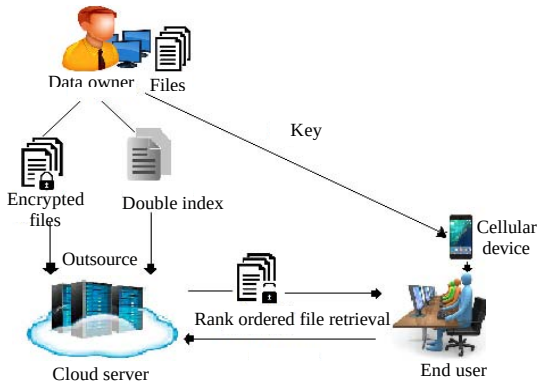

Fig. 1: System architecture

Table 1: Posting list of inverted index

Terms	w				
File ID	F1	F2	F3	F4	Fn
Relevance score	0.027	0.027	0.019	0.019	0.015

receiving that search request Tw, the cloud server will search the index I and return the corresponding set of files from the file collection to the user. To authenticate a ranked search result (or Top-k retrieval), one needs to ensure that the retrieved results are the most relevant ones and the relevance sequence among the results is not disrupted. In this model authorized user received key via. the cellular device which has been send by the data owner by taking the IMEI number (Table 1).

Efficient data privacy: To protect the tactful cloud data, we used one to many order preserving symmetric algorithm to encrypt/decrypt the documents. However, the security definition and the constructions of OPE in and are based on the assumption that OPE is a deterministic encryption scheme which means that a given plaintext will always be encrypted as a fixed cipher text. However, deterministic encryption leaks the distribution of the plaintexts, so it cannot ensure data privacy in most applications.

For instance, in privacy-preserving keywords search, OPE is used to encrypt relevance scores in the inverted index. When using a deterministic OPES, the resulting cipher text shares exactly the same distribution as the relevance score by which the server can specify the keywords. Therefore, improved the OPE in and proposed a "One-to-Many OPE" in their secure keyword search scheme where they tried to construct a probabilistic encryption scheme and conceal the distribution of the plaintexts. The overall working of OPE is shown in Fig. 2.

Inverted index: Indexing structure stored the key word or term which has the unique keyword the keyword stored in index are extracted from file and then indexed for to maintain the security as strong as possible. Inverted index is used for mapping search request to the set of files as shown in Fig. 3. And the numerical score of set of mapped files are estimated to give relevance search query result. And the posting list of inverted index is shown in Fig. 4.

Ranked keyword based search technique: In this study ranking function is used to evaluate the relevance score of the matching search request files. We uses the term frequency and inverse document frequency rule to evaluate the statistical measurement of relevance score where term frequency is simply the number of times a

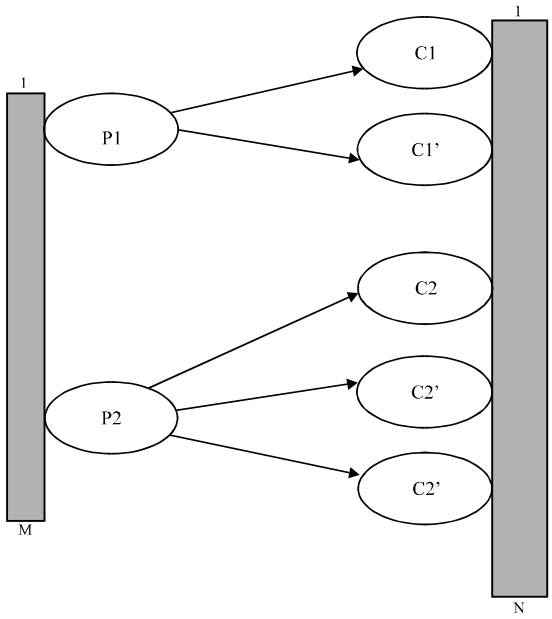

Fig. 2: One-to-many order preserving mapping technique

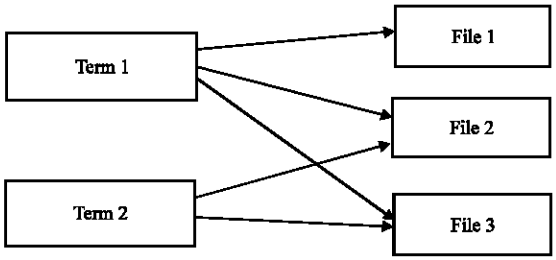

Fig. 3 : Mapping between term and set of files

given term or present within a file and inverse document frequency is obtained by dividing the number of files in the whole collection by the number of files containing the term. Using this function we can evaluate relevant file for the search request by the end users. The ranking function is define as:

$$\text{Socre}(Q,F_d)=\sum_{t\in Q}\frac{1}{F_d}\times(1+\ln f_{d,t}).\ln\left(1+\frac{N}{f_t}\right)$$

Here:

- Q = The searched keywords
- $f_{d,t}$ = The TF of term t in file F_d
- ft = The number of files that contain term
- N = The total number of files in the collection
- F_d = The length of file F_d, obtained by counting the number of indexed terms, functioning as the normalization factor

Concept based search technique: Concept-based information retrieval is an alternative IR approach that

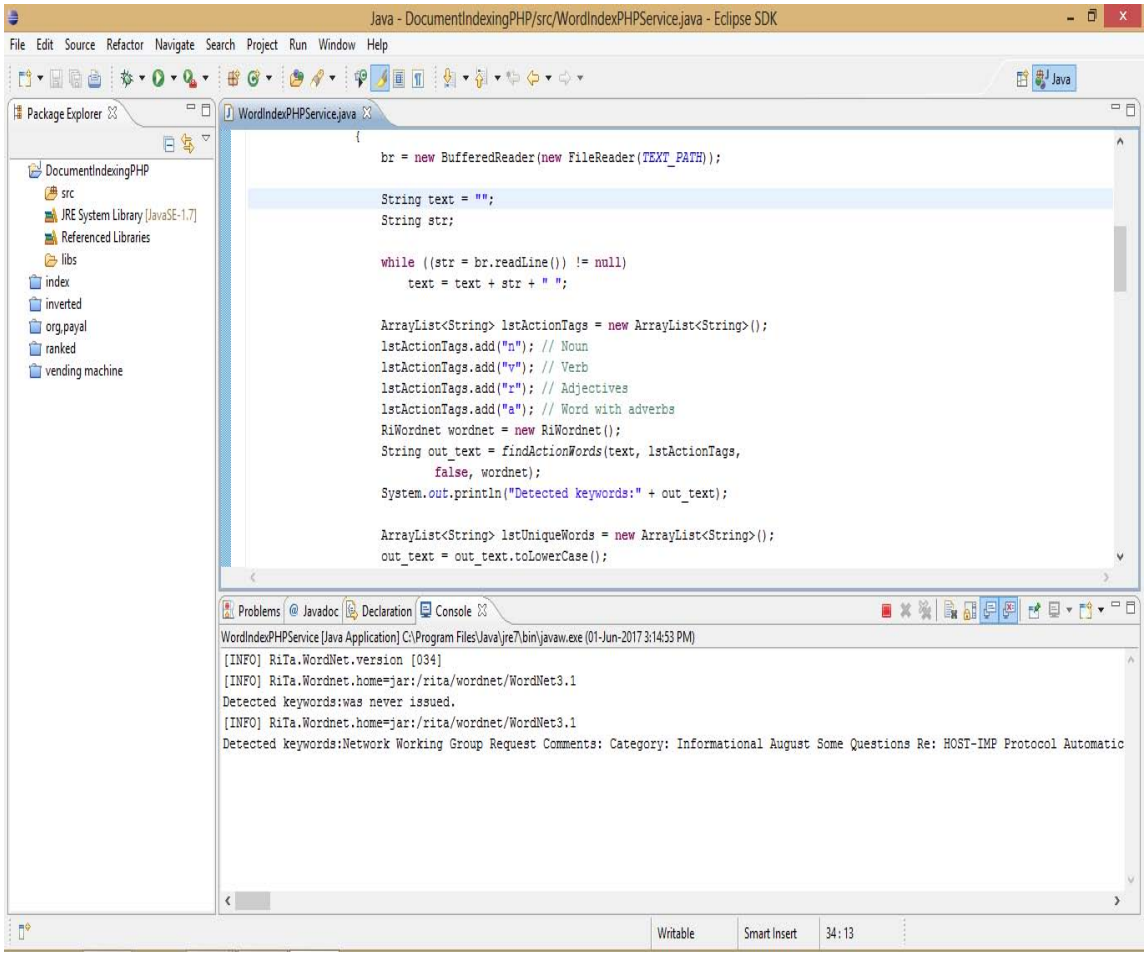

Fig. 4: Keyword extraction

aims to solve the problem of keyword based search algorithm. In concept based search. concept-based search aims to improve retrieval effectiveness by organizing search results based on their meaning.

RESTULTS AND DISCUSSION

In this study data owner outsourced his encrypted and index file on a cloud. Ranked keyword search and concept search algorithm are implemented for retrieving the file when the user sends the request to the cloud. And based on relevance score user will retrieve the search result. Score dynamic concept is implemented for updating the index records. And order preserving technique is implemented for maintaining the order of the encrypted file on a cloud. IMEI number security is implemented for the allowing only authorized technique. And the double indexing technique is implemented for improving efficiency.

Keyword extraction: Figure 5 shows the extraction of keywords from a file using the WordNet dictionary. In this keyword is extracted from file are unique.

User search file: Figure 6 shows that the input given by user using the search engine is a keyword that keyword matched with the indexed keywords and on the relevancy basis search file is retrieved. Retrieved file is relevant or not are checked by calculating the relevance score of file using the ranking function.

User request for file to admin: After retrieving a set of relevance file user requests the admin for the approval on file which user wants to download.

Admin approved file: The file which is send by user to admin is approved or reject is done by admin. The Requested file is approved and user can download that file is shown in Fig. 7.

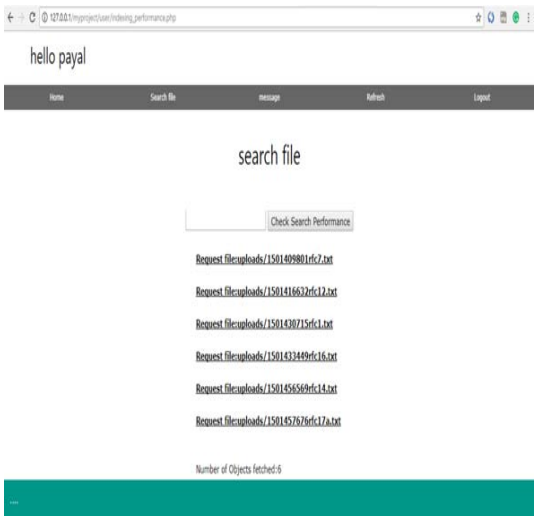

Fig. 5: User search file screen

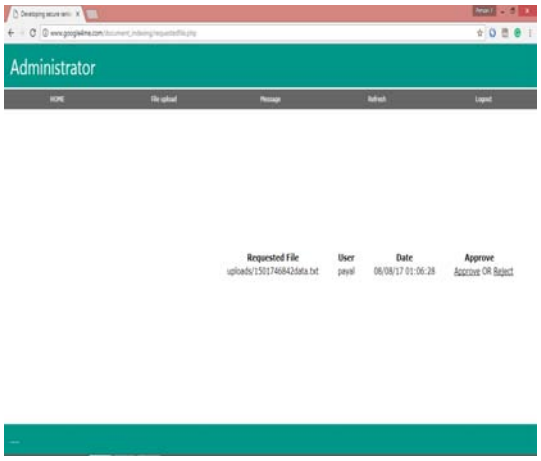

Fig. 6: User request for file to admin

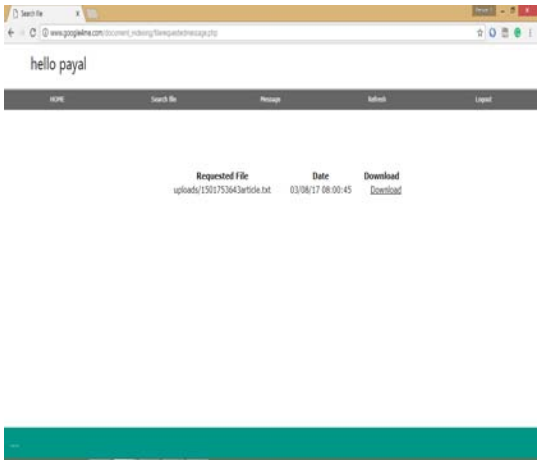

Fig. 7: Admin approved file

CONCLUSION

The existing searchable encryption framework, it is very inefficient to achieve ranked search data and appropriately weaken the security guarantee. The proposed system uses ranked keyword search technique and concept search technique for searching data on cloud server and also uses a double indexing technique for providing security. Using IMEI number based Authentication allows only authorized users to access the cloud storage data. It follows in improving security. The ranked keyword search and concept search gives a more efficient searchable technique for searching the more relevant files.

REFERENCES

Agrawal, R., J. Kiernan, R. Srikant and Y. Xu, 2004. Order preserving encryption for numeric data. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, June 13-18, 2004, ACM, Paris, France, ISBN:1-58113-859-8, pp: 563-574.

Boldyreva, A., N. Chenette, Y. Lee and A. O'neill, 2009. Order-preserving symmetric encryption. Proceedings of the 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques Vol. 5479, April 26-30, 2009, Springer, Cologne, Germany, pp: 224-241.

Chang, Y.C. and M. Mitzenmacher, 2005. Privacy preserving keyword searches on remote encrypted data. Proceedings of the 3rd International Conference on Applied Cryptography and Network Security (ACNS'05) Vol. 5, June 7-10, 2005, Springer, New York, USA.,-pp: 442.

Ellaji, C.H. and R.D. Shobha, 2015. Security in multi keyword search over encrypted cloud data. Intl. J. Emerging Trends Eng. Res., 3: 113-116.

Gauch, S., J. Chaffee and A. Pretschner, 2003. Ontology-based personalized search and browsing. Web Intell. Agent Syst. Intl. J., 1: 219-234.

Goh, E., 2003. Building secure indexes for searching efficiently on encrypted compressed data. http://eprint.iacr.org/2003/216/.

Liu, F., C. Yu and W. Meng, 2002. Personalized web search by mapping user queries to categories. Proceedings of the 11th International Conference on Information and Knowledge Management, November 4-9, 2002, ACM, McLean, Virginia, USA., ISBN:1-58113-492-4, pp: 558-565.

Ren, K., C. Wang and Q. Wang, 2012. Security challenges for the public cloud. IEEE Internet Comput., 16: 69-73.

Speretta, M. and S. Gauch, 2005. Personalized search based on user search histories. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Sept. 19-22, IEEE Computer Society Washington, DC, USA., pp: 622-628.

Subashini, S. and V. Kavitha, 2011. A survey on security issues in service delivery models of cloud computing. J. Network Comput. Appl., 34: 1-11.

Swetha, K. and R.M.R. Narasinga, 2016. Dynamic searchable encryption over distributed cloud storage. Assian J. Inf. Technol., 15: 4763-4769.

Wang, C., N. Cao, K. Ren and W. Lou, 2012. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans. Parallel Distib. Syst., 23: 1467-1479.

Xiao, L. and I.L. Yen, 2012. Security analysis for order preserving encryption schemes. Proceedings of the 2012 46th Annual Conference on Information Sciences and Systems (CISS), March 21-23, 2012, IEEE, Princeton, New Jersey, USA., ISBN:978-1-4673-3139-5, pp: 1-6.

Xu, Y., K. Wang, B. Zhang and Z. Chen, 2007. Privacy-enhancing personalized web search. Proceedings of the 16th International Conference on World Wide Web, May 8-12, 2007, ACM, Banff, Alberta, Canada, ISBN:978-1-59593-654-7, pp: 591-600.