American-Eurasian Journal of Scientific Research 11 (5): 398-404, 2016
ISSN 1818-6785

DOI: 10.5829/idosi.aejsr.2016.11.5.22959

Privacy Preserving Similarity Based Text Retrieval Through Blind Storage

Pinki Kumari and S. Jancy

Department-MCA, Faculty of Computing, Sathyabama University, Chennai-600119, Tamilnadu, India

Abstract: Now days cloud computing is improving rapidly due to their more advantage and more data owners give interest to outsource their data into cloud storage for centralize their data. As huge files stored in the cloud storage, there is need to implement the keyword based search process to data user. At the same time to protect the privacy of data, encryption techniques are used for sensitive data, that encryption is done before outsourcing data to cloud server. But it is critical to search results in encryption data. In this system we propose similarity text retrieval from the blind storage blocks with encryption format. This system provides more security because of blind storage system. In blind storage system data is stored randomly on cloud storage. In Existing Data Owner cannot encrypt the document data as it was done only at server end. Everyone can access the data as there was no private key concept applied to maintained privacy of the data. But In our proposed system, Data Owner can encrypt the data himself using RSA algorithm. RSA is a public key-cryptosystem and it is widely used for sensitive data storage over Internet. In our system we use Text mining process for identifying the index files of user documents. Before encryption we also use NLP (Nature Language Processing) technique to identify the keyword synonyms of data owner document. Here text mining process examines text word by word and collect literal meaning beyond the words group that composes the sentence. Those words are examined in API of word net so that only equivalent words can be identified for index file use. Our proposed system provides more secure and authorized way of recover the text in cloud storage with access control. Finally, our experimental result shows that our system is better than existing.

Key words: Sensitive data · Multi-keyword ranked search · Nlp technique · Wordnet

INTRODUCTION

Over the last few years, improving advances in the network based computing field and application of mobile cloud computing (MCC) has been launched as possible technology in mobile services. The combination of cloud computing, mobile computing and wireless networks provides more quality resources of computation to mobile users, cloud computing providers and mobile users. Mobile Cloud Computing (MCC) is new platform which combine the cloud computing and mobile devices to develop new infrastructure. In this infrastructure both data processing and data storage is achieved outside of mobile devices. In the technology of mobile computing normally mobile users outsource their records or data to outside cloud servers, For eg., iCloud, it provide low-cost, scalable and stable way for data access and storage. However, every outsourced data may contain some sensitive privacy records, such as emails, personal photos, etc., which cause some chances to severe data confidentiality and privacy issues, if without protections. So encryption processes have to done before outsourcing the data to cloud. Here RSA algorithm is used for encrypt the data owner data. In modern computers RSA is used for encrypt and decrypt user information or messages. Basically RSA is symmetric cryptographic based algorithm. It means two keys are used for encryption process. Here encryption is achieved by public key and decryption is achieved by other key this key must be kept private. Using the private key only any user can decrypt messages on cloud. In our propose system the private send by data owner to user by cloud server at the requesting time. Cloud has huge amount storage in which number data are stored it. From this storage retrieving exact data for user who wants access particular data is a critical task. One basic way for utilization of data is search operation. I.e. to rapidly finds the information from large data amount. To provide better search result search

Corresponding Author: Pinki Kumari, Department-MCA, Faculty of Computing, Sathyabama University, Chennai-600119, Tamilnadu, India.

techniques needs to be used in cloud system. Searching the information on plain text is easy one and it is critical problem for cipher text information or data. Retrieving the data from huge volume of data can be achieved keyword based search operation. Normally single keyword search method will result the large more no of records from storage space and this method not satisfies the end user. So improve its searching performance ranking technique has been suggested. Ranked searching process will reduce the unwanted network traffic by resulting most applicable data to end user.

Related Work: Some existing author has proposed some techniques based on data security and keyword based searching mechanism. Some of the work has been discussed here. Author Boneh *et al*. [1] suggested a encryption technique based on the public key search, which helps to search the encrypted data by keyword. The context of application as follows: (1) Bob forward an email to Alice with encryption format by Public key of Alice. (2) Email gateway of Alice's want to check whether the forwarded email hold the urgent keyword so that it can route email accordingly; (2) But Alice not like to decrypt her message in email gateway. Bonech *et al*. [1] describe and build a method that allows a gateway can test whether the urgent word is keyword in the email using trapdoor provided by Alice, but absorb nothing about email. This technique can be acceptable to cloud computing environment with some of improvements. The technique of searchable encryption that provides search operation over on encrypted data in cloud. It is arranged as searchable symmetric encryption (SPE) and searchable public key encryption (SSE). The scheme of SSE has been suggested by Song *et al.* [2] that constructs the searchable process but it supports only single keyword search on encrypted data. Author Naveed *et al*. [3] builds system for blind storage to reach a searching on encryption data on cloud. But it only supports single keyword search. Author Cong Want overcomes [4] the issue of Boolean search that is searching technique traditional method which will encounter the data utilization effectively. This paper satisfy "as strong as possible" security promise. Furthermore it traverse relevant score from retrieval information to construct a searchable index with secure way and create a one to many order mapping technique to securely protect sensitive records. Ning cao [5] suggested a co-ordinating matching similarity measure with efficiently that results more feasible data which is relevant to multiple keywords that given by the data owner. Author Cong Wang [6] attentive on fetching the matching records in ranked order depending to their relevance basis and it could be achieved by indexing. This helps to enable the fast search from documents that hold a given keyword.

Proposed Work

Overview: In our proposed system we implement the concept for retrieving the encrypted text data from mobile cloud storage based on the multi keyword search. Here we store the data by blind storage method. In blind storage method data's are normally stored in fixed size of blocks and also stored in random manner. So, cloud cannot access the data without permission of data owner. In our proposed system whenever data owner upload the data into cloud storage, data content should be extracted by Natural Language Processing (NLP) method and extract related synonyms of keywords then data will be encrypted and upload to cloud. After the encryption process has finished, the data user wants to search over the encrypted data by the keyword. Before he/she request to access the data they must get the secret key from Data owner.

Architecture: The Figure 1 describes the overall architecture of our system, in this data owner uploads the data into cloud as blind storage. Before uploading the data it should be encrypted by using RSA algorithm. The mining of text process is a NLP (Nature Language Processing) and word net tools is used to extract the contents. Here NLP process is used to extract the literal meaning in that content and word net tool is used to provide the related synonyms of literal word in that content. Then data and contents will be stored on cloud. If data user will try to search a data on cloud server, cloud server will search the related data based on the data user given keyword. The cloud server provides some related filename to user. If user wants to view the content, the user should click the filename, at that time cloud send the user details and file name to the owner. Then data owner encrypt their private key by public key of data user. Data owner sends the encrypted key to cloud server, server sends key to data user, then user decrypt the key using the private key. After that data user receives the private key from data owner then access the data in blind storage [7-10].

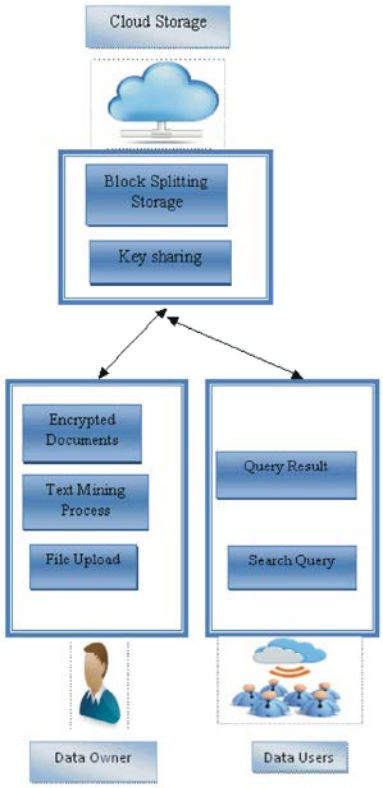

Fig. 1: Overall Architecture

Design Goals

Group Creation: Before uploading the data to cloud storage data owner should be register in this environment and create a user group. Data users also register and give request to group owner to add into group. Once data owner accept the request from user, he/she add the requested user into group. Data user only can access the respective documents of data owner.

Text Mining Process: In text mining process, NLP and word net tools is used to extract the relevant content from data owner data before encrypting the data. NLP process is used extract the literal meaning in file contents. Word net tool is like a dictionary and it is used here for provide the related synonyms to literal word in that content.

Data stored in Blind Storage System: In our proposed system data should be encrypted before upload into cloud. Once data owner encrypts the data it uploads to cloud storage through blind storage mechanism. A blind storage system is constructed on cloud server side to help updating, adding and deleting the documents and hides the access pattern of search user from cloud server. In the system of blind storage, all data are divided into some fixed-size block. Based on document associated seed random integer sequence generated and it will assigned to each of these blocks. In cloud server view, it notice the upload of encrypted documents blocks and downloaded. Therefore, all the index and documents can be stored in the system of blind storage to achieve a encryption scheme by searchable way.

Query Processing: Data user will try to search a query in cloud server. The cloud servers map the keywords and search the associated files. The cloud server provides the related filename to user. To view the content the user should click the filename; at that time user request to cloud server and server send the user details and filename to the data owner. Then data owner knows all public key of user so he encrypt the private key using data user public key and the encrypted key send to server and server send that key details to user, then user decrypt the key using our private key. After that the data user get private key of data owner and then access the data through blind storage.

Algorithms Used

RSA Algorithm: This RSA algorithm is used to decrypt and encrypt the file contents. It is an asymmetric type algorithm. It contain three steps; Generating key, encryption and decryption.

Key Generation: RSA require private and public key. The public key is used to encrypt messages and it will be shown to everyone. Private Key is used for decrypting messages, but it won't show to everyone.

RESULT AND DISCUSSION

Search on encryption data have to help the given three function. First, the searching encryption schemes have to support multi keyword operation, Second, to recognize relevant results more quickly. Third, search efficiency should support if database size could be larger. The above all are done in our proposed system. In our proposed system we use RSA algorithm for data encryption. When RSA algorithm is applied on the data then we get encrypted data and then encrypted data is stored on the cloud storage based on the blind storage technique. User can access the data after downloading and decrypting file. For encryption and decryption keys are provided Here we show our experimental with screenshots and graph.

Fig. 2: Data owner and user Registration form

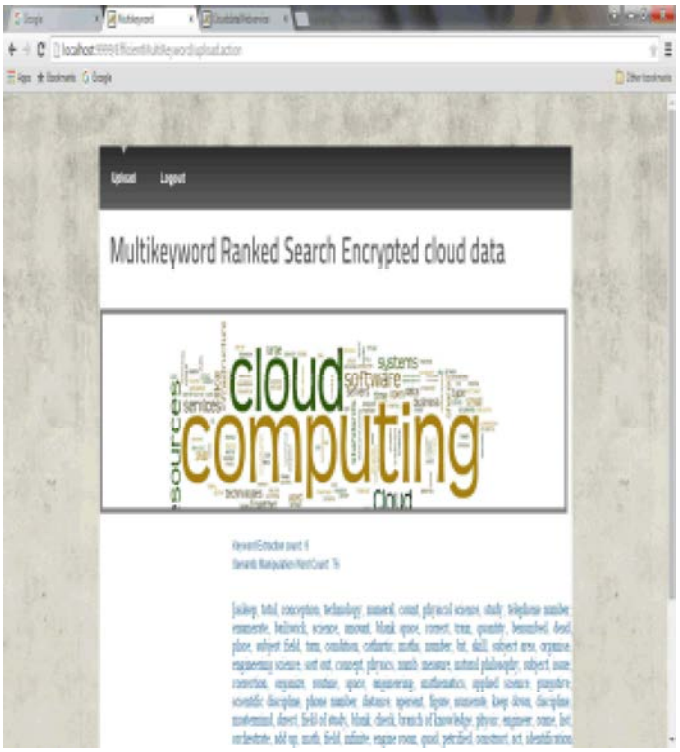

Fig. 3: Keyword extraction form data

Fig. 4: User enter keyword for searching

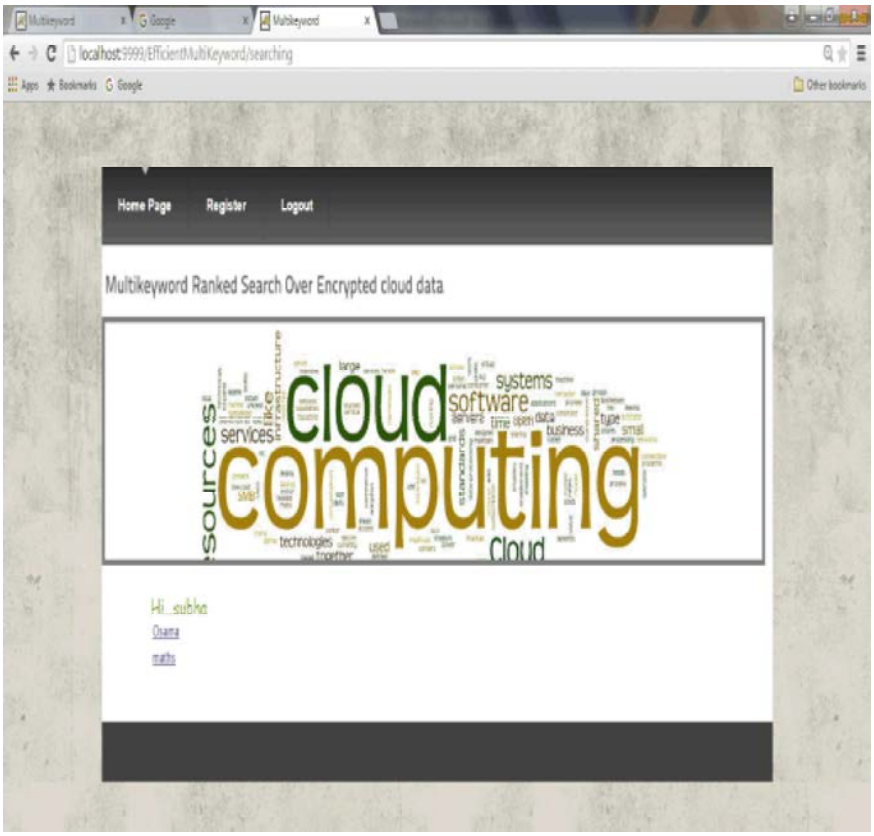

Fig. 5: Related content based on keyword

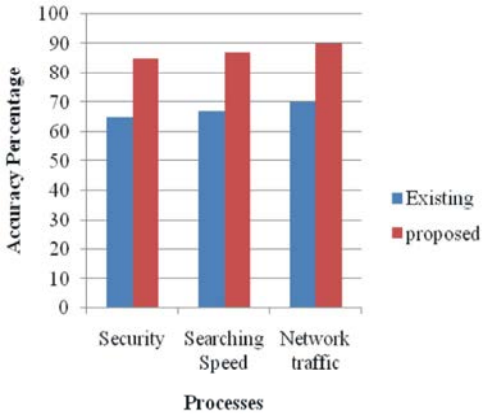

Fig. 6: Processes Accuracy of Proposed System

The above figure shows the data owner and user registration form. Before uploading the data to cloud storage Data owner need to crate group with some user who wants to access the data.

The above output shows keyword extraction phase. In that before uploading the data, data owner have to extract some keywords from that data. These keywords are used to identify the related contents when searching on cloud storage [11-14].

The above screen shows searching operation based on the keywords. Once user got the keyword from data owner, they can search the data on cloud by these keywords.

The above screen shows the outputs of keyword. When user enters the keyword on cloud storage it will show some related contents based on the keywords. If user wants to access these contents, they need click on contents for access. Once user click on the contents user detail will send to data owner and data owner provide the permission for whether that user can access the data or not.

The above Figure 6 explains process accuracy with existing system. Compare to Existing system our proposed system performance is better in all above mentioned process. In our system we RSA for data encryption and NLP for identify the file content. This system also reduces the network traffic by only sending back the most relevant results from cloud to search users.

CONCLUSION

In this proposed system we use multi-keyword related search scheme to allow efficient, accurate and secure on encrypted MCC (Mobile Cloud Data). Security surveys have showed that our proposed scheme can effectively done documents confidentiality and index, data security and keyword privacy. By this system, it also used to reduce the unwanted network traffic by sending the only relevant results from storage of cloud to search user. Since the more number of the documents contained in database could cause the searchable results more delay. The searchable encryption system provides efficient way to fast respond to search request with little delays. Our proposed system performance results shows better efficiency in terms of the computation overhead and functionality compared with existing ones.

REFERENCES

1. Boneh, D., G. Crescenzo, R. Ostrovsky and G. Persiano, 2004. Public Key Encryption with Keyword Search. Proceedings of Eurocrypt, Lecture Notes in Computer Science, 3027: 506-522.
2. Xsong, D., D. Wagner and A. Perrig, 2000. Practical techniques for searches on encrypted data, in proc. IEEE symp. Secur. Privacy, pp: 44-55.
3. Naveed, M., M. Prabhakaran and C.A. Gunter, 2014. Dynamic searchable encryption via blind storage, in proc. IEEE symp.secur.privacy, pp: 639-654.
4. Wang, C., N. Cao, J. Li, K. Ren and W. Lou, 2010. secure ranked keyword search over encrypted cloud data, in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), pp: 253-262.
5. Cao, N., C. Wang, M. Li, K. Ren and W. Lou, 2014. Privacy preserving keyword ranked search over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst., 25(1): 222-223.
6. Wang, C., N. Cao, K. Ren and W. Lou, 2012. Enabling secure and efficient ranked keyword search over outsourced cloud data,IEEE Trans.Parallel Distrib. Syst., 23(8): 1467-1479.
7. Abolfazli, S., Z. Sanaei, E. Ahmed, A. Gani and R. Buyya, 2014. Cloud-based augmentation for mobile devices: motivation, taxonomies and open challenges. Communications Surveys & Tutorials, IEEE, 16(1): 337-368.
8. Liu, F., P. Shu, H. Jin, L. Ding, J.Yu, D. Niu and B. Li, 2013. Gearing resource-poor mobile devices with powerful clouds: architectures, challenges and applications. Wireless Communications, IEEE, 20(3): 14-22.
9. Sarah Perez, 2009. Why cloud computing is the future of mobile, http://www.readwriteweb.com/archives/why_cloud_computing_is_the_future_of_mobile.php, Retrieved on February 2015.

10. Li, H., Y. Dai, L. Tian and H. Yang, 2009. Identity-based authentication for cloud computing, in Cloud Computing. Berlin, Germany: Springer-Verlag, pp: 157-166.
11. Weiss, A., 2007. Computing in the Clouds. net Worker, 11(4): 16-25.
12. Gartner Identifies the Top 10 Strategic Technologies for 2009,http://www.gartner.com/it /page.jsp?id=777212/ [14 Oct 2008].
13. Amazon Elastic Compute Cloud (EC2), http://www.amazon.com/ec2/ [18 Jul 2008].
14. Wang, B., S. Yu, W. Lou and Y.T. Hou, 2014. Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud, in Proc. IEEE.INFOCOM, pp: 2112-2120.