ISSN No:-2456 –2165

Accurate Information Retrieval Using Semantic Searching from Encrypted Database

Sandhya Mohite
Department of Computer Engineering,
MIT Academy of Engineering,
University of Pune

Abstract—In cloud computing scenario security and privacy are important part. To reduce costing and for fast searching speed many data owners can outsource the data to public cloud storage.

This proposed system can outsourced data files in encrypted form which protect data files and preserve privacy. Random utilization of data can be reduced due to encrypted file format. This paper include concept of conceptual graphs for semantic searching of documents. Every data user is authorized by data owner, thats why authorized user can download encrypted documents by decrypting it.

Keywords—Encryption Technique, Key Generation, Security, Encrypted key

I. INTRODUCTION

There are many cryptographic techniques are developed in previous year but many of them are time consuming and more complex. Due to encryption methods data on cloud storage remains more secure.

Hash value is calculated for generation of keys. For downloading documents from cloud, each user should be added to the storage by admin of cloud. At the time of document download, verification code or hash value or encryption key should be send to the mail id of user. After enetering encrypted key, user can view encrypted documents In triple DES technique key size is too small, small key provide less security. As key size increases, privacy of system increases. Key sizes also have impact on privacy of system. So, to overcome this key size limit here AES algorithm (Advanced Encryption Technique) can be used.

Advanced encryption standard (AES) technique is six-times faster than previous technique Triple DES technique. In AES, there is no key size limit.AES algorithm uses 128-bits,192-bits and 256-bits keys for encryption purpose. As if Fiestel cipher considered, AES technique is faster than Feistel cipher.

AES technique is designed by the Belgian cryptographer Joan Daeman and Unicent Rijmen.AES is the symmetric key block cipher which uses same key for both encryption and decryption. Encryption and decryption generated by using AES algorithm. Same should be used for encrypt the document and same key will be used for decrypt the document. There is no different key should be generated for encryption and decryption of the document.

Document should be searched by using the keyword which is entered by user at the time of upload. Every time that keyword should be used for searching documents. Files may be .pdf file,.doc file or .jpg or .png file.

Ranking mechanism should be used for increasing searching time of document retrieval. The count of successfully searched document from total number document can be calculated. Maximum rank should be 1.

II. REVIEW OF LITERATURE SURVEY

W.K. Wong focuses k^{th} nearest neighbour (KNN) query technique.Encryption technique can be developed by W.K.Wong by using KNN query algorithm.

In KNN query algorithm, it searches k^{th} nearest point of particular query in the dataset. If nearest co-ordinate is matches with the query then it becomes very easy to find out appropriate database. But this KNN encryption technique not much secure, because anyone can entered query,if that query matches with nearest poit of dataset then co-ordinates get match and database can get hacked.

A swaminathan introduces cryptographic techniques. Also relevance score between query and document can be calculated help for protecting data.

Many cryptographic encryption techniques are used to protect data files from hackers. Encrypted data should be placed on cloud server .i.e.third party.But the drawback of using traditional encryption technique is user can search encrypted files by only entering the keyword and user can read the data without decrypting the file. Due to this privacy of encrypted file may get lost. Any one can get encrypted file easily without decrypting it.Because there should not be any decrypted key provided to user for downloading the documents.

It consists mainly two problems, first is user should go through all encrypted files stored on cloud server without knowledge of cloud. Each and every file should be retrieved by user for finding matching document to query. Second problem is it causes unnecessary network congestion. Drawback for this paper is, admin cannot given verification for every client/user, because of that any client allowed to be freely search files on cloud.

Sun et al. Introduces PKC and SKC searching techniques.Keys are generated in PKC algorithm but any documents are accessed by users on the cloud.

Boneh et al. Gives idea about public key encryption technique.PKC algorithm is single keyword searching technique.The disadvantage of this is user having public key can easily download document or edit document.

Dynamic addition of files and deletion of files can be introduced by Karmana et al.

Data files should be stored on cloud server in the encrypted file for preserving the privacy of documents on the cloud. But in the case of large data files stored on cloud, it iss difficult for user to download all files and decrypt every files from large data files stored on to the cloud. So, it is essential to provide ranked keyword search technique to retrieve documents efficiently and securely.

III. PROPOSED SYSTEM

In proposed system, SSCG technique [Semantic search scheme based on conceptual graphs over encrypted outsourced data] is applied. If we observe MRSE technique it focuses only on the semantic searching of documents related to the requested query.

SSCG scheme has been used here which can focus on semantic searching of document and also security of documents because in SSCG scheme only keyword of respected file can be known to the cloud server and document is unknown for cloud server.

Server performs ranked analysis by focusing only on keyword /ciphertext here, modified OPSE is applied for ranking results.

Our proposed system contain admin of the system which having username and password of the database. Admin having authority to add users, who require documents.

Documents are stored in encrypted format on the cloud to improve the privacy of documents. If only plain documents are stored on the cloud, there should be chances of hacking, means if in the case hacker get username and password of cloud, it is login successfully and hacker can upload plain document easily, so here we are storing documents in the encrypted format.

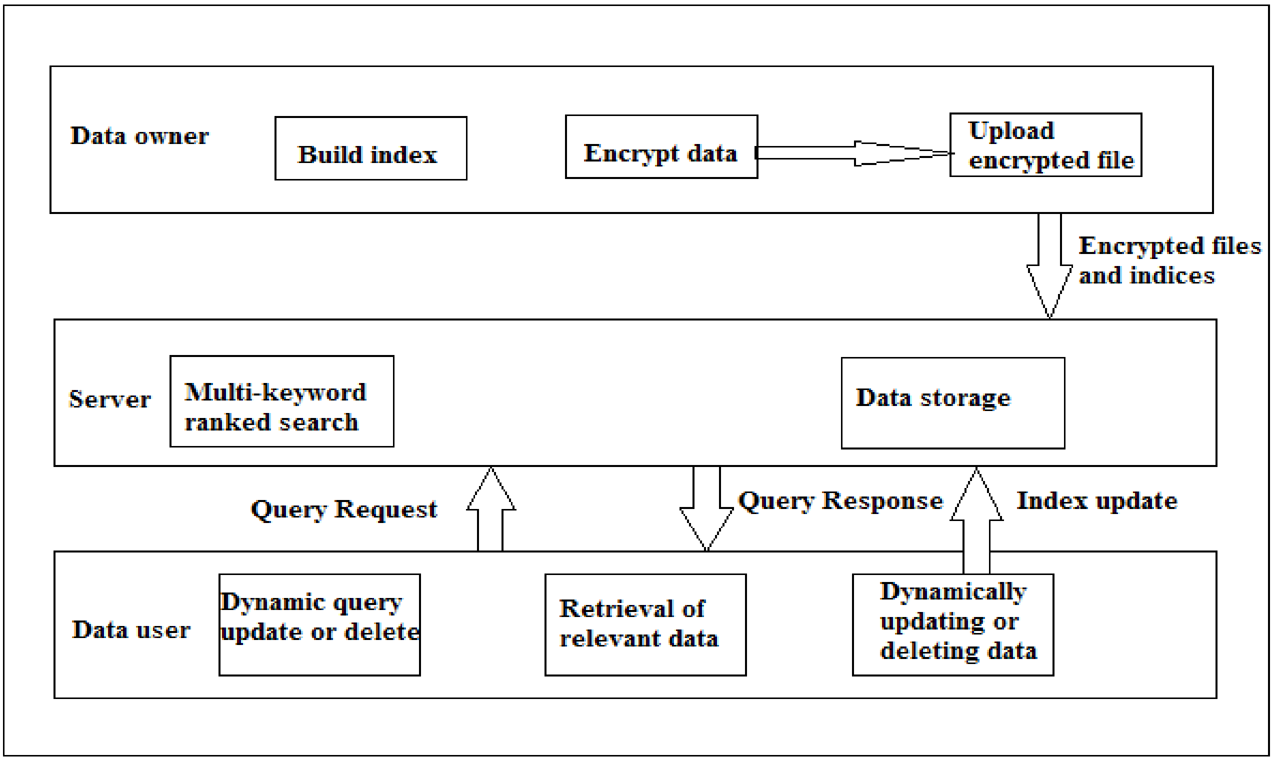
Data owner
Build index
Encrypt data
Upload encrypted file
Encrypted files and indices
Server
Multi-keyword ranked search
Data storage
Query Request
Query Response
Index update
Data user
Dynamic query update or delete
Retrieval of relevant data
Dynamically updating or deleting data

Fig.1 System Architecture

Second entity is the users of the system which is added by the admin of the cloud.Admin can upload the documents at the server,which is in encrypted format.At the time of upload,it require particular keyword of document.

If user going to search the document by entering the keyword, next step is activation code generation. If user is successfully added to the server by the admin, then by entering the respective mail id of that user,activation code or encrypted key can be send to the mail id of valid user. Then by entering that key user can get desired encrypted document which is searched and document can be visible to only that user.

IV. Ranked Analysis

Cloud server displays set of documents to the data server such as F(Q) which file matches with query can be selected.But some times there exists two or more files which are same and cloud server cannot be able to differentiate these documents,thatswhy,cloud server perform ranking function.

Ranking function simply calculate relevance measure between requested query and the document.In information retrieval, relevance score can be calculated by given formula,

TF × IDF

Where,

TF= Term Frequency
IDF= Inverse Document Frequency

Term frequency means number of terms or keyword appeared in respected document. Inverse document frequency means dividing number of searched files by total number of files.

V. CONCLUSION

In this paper,we focus on the semantic document search with queries and also security of cloud server. This paper tries to improve semantic searching of files as well as security of files on cloud server. We designed SSCG scheme which satisfy our objectives such information retrieval, semantic search, minimum information leackage and more security level. This paper also focuses on relevance score between retrieved documents and query.

REFERENCES

[1]. C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, A hierarchical clustering method For big data oriented ciphertext search, in Proc. IEEE, INFOCOM, Workshop on Security and Privacy.

[2]. D. X. D. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data, in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp.4455.

[3]. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, Publi key encryption with keyword search, in Proc. EUROCRYPT, Interlaken, SWITZERLAND,2004, pp. 506522.

[4]. A. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He,M. Wu, and D. Oard, Condentiality-preserving rank-ordered search, in Proc. ACM ACM Workshop Storage Security Survivability, Alexandria, VA, 2007, pp.7-12.

[5]. C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, Secure ranked keyword search over encrypted cloud data, in Proc. IEEE 30th Int.Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 2532.

[6]. Pang, J. Shen, and R. Krishnan, Privacy-preserving similaritybased text retrieval, ACM Trans. Internet Technol., vol. 10, no. 1,pp. 39, Feb. 2010.

[7]. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li,Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, in Proc. 8th ACM SIGSAC Symp. Inform., Comput. Commun. Security, Hangzhou, China, 2013,pp. 7182.

[8]. Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data, IEEE Transactions on parallel and distribute system,vol. 27, NO. 2, february 2016.

[9]. Wei Zhang, Student Member, IEEE, Yaping Lin,Member, IEEE, Sheng Xiao, Member,Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing,journal of latex class files, VOL. 6, NO. 1, January 2015.

[10]. Cao N, Wang C, Li M, Ren K, Lou W. Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE Trans Parallel Distrib Syst 2014 system,vol. 27, NO. 2, february 2016.