Token Search towards Secure Storage and Retrieval of Data in Cloud Storage

S. Hemavathy[1], D.Kavitha[2]
P.G. Scholar[1] , AssistantProfessor[2]
Department of Computer Science and Engineering,
Valliammai Engineering College

Abstract: **Cloud computing enables the data service and it promises the capability of sharing encrypted data with different users through public cloud storage and it may have the security issues over the data confidentiality and authentication access control.. In this paper we develop a searchable encryption technique for multi-keyword ranked search over the data storage in the context of cloud through blind storage. The blind storage allows a client to store a set of files on a remote server that the server does not familiar with the files that are stored in it. The efficient multi-keyword ranked scheme can return the search resultant based on the ranked search with efficient accuracy by using enhanced K-nearest neighbor technique. To make the search efficient blind storage system is used to conceal access control issues in searchable encryption technique. The security analysis scheme demonstrates to achieve authentication and confidentiality of the document. The proposed scheme focuses on security attacks like man in the middle in the middle.**

Keywords: Cloud Computing, Blind storage, Multi-keyword search, Access control, Privacy search.

I .INTRODUCTION

Cloud computing is an emerging technology which has been widely used for storing and retrieving large amount of data over the internet. It helps the user to store their data at very low cost and provide to save their personal data such as photos, documents over the cloud environment. Cloud helps business people more storage space for the data. Since most sensitive data are stored in the cloud, it needs more confidentiality and privacy for the data owners. It is therefore necessary to encrypt the user data before outsourcing the data in the cloud environment. The data stored in the cloud can be shared among the data users; in turn the data users need to decrypt the data by getting the key from the data owner. The key factor of using cloud is to store the user data securely at very low cost. The data owner and cloud server are no longer in the same trusted domain may put the outsourced encrypted data at risk. The cloud server may leak data information to unauthorized entities or even be hacked; hence data in the cloud need to be protected well for the security purpose. Cloud mainly provides three major services Software as a service (SAAS), Platform as service (PAAS), Infrastructure as a service (IAAS).The cloud can be deployed as Public, Private and Hybrid clouds. Depending upon the business needs and organizational constraints, appropriate cloud deployment model can be selected although a general recommendation from a security standpoint would be adopted.

Cloud contains large amount of data stored in it, hence retrieving the data which is needed for the search user is difficult in cloud. One fundamental way of data utilization is the search operation.i.e.to quickly sort out information of interest from huge amount of data. To provide the search efficiency search techniques has to be used in the cloud. The data user wants to retrieve the file what they are interested in instead of getting undifferentiated results. The data retrieval is an efficient process for the plain text scenario and it turns difficult for the cipher text data. The solution for retrieving the data can be done efficiently by keyword based retrieval. The traditional method is the single keyword search which will retrieve large number of data and which will not satisfy the end user. Hence in order to improve its performance ranking method has been proposed. It is also necessary for ranking system to support multiple keyword search as single keyword search often yields far too coarse results. Ranked search will also eliminate the unnecessary network traffic by sending back only the most relevant data to the search user. The keyword given by the data user helps the server to narrow down the search result and retrieve the data based on the keyword. The keyword search helps the server and also the search user to retrieve the data what the search user want.

2. BLIND STORAGE

In order to overcome the above attacks, cloud need to provide more security for the data which is uploaded by the data owner. One way is to store the data more securely which can be adopted by using the Blind storage. It allows the data owner to save the data more securely and makes it visible only to the data owner. A blind storage allows the data owner to save the set of files on a remote server in such a way that the server does not know about the contents of the files what the owner is stored. Blind storage supports adding new files, updating or deleting existing files. The server will know only the file name of what the data owner is uploading. Hence the

blind storage leaks only little information to the server. By storing the data in the blind storage the data owner can prevent other data users to unaware about the content of the file. However data will be saved in the form of fixed blocks and each block will be indexed in order to know about the file. Moreover the server will know about the existence of the file (and its size not the name used by the data user to refer to the file or its contents) only when the data user retrieves it later.

2.1 DATA STORED IN BLIND STORAGE

The documents stored in blind storage are divided into fixed size blocks. These blocks are indexed by a sequence of random integers generated by a document related seed. In the view of cloud it can see the blocks of encrypted document uploaded and downloaded. Thus the blind storage gives only little information to the cloud server. Specifically the cloud server does not know which blocks are of the same document, even the total number of the document and size of each document. All the documents and index will be stored in the blind storage system to achieve a searchable encryption scheme. If the data user want to retrieve the data from the blind storage then the user need to get the key and the index of the file from the data owner.

2.2 SYNTAX OF BLIND STORAGE

A blind storage system consists of a client and the "dumb" storage server. The server will provide two operations, 1.download, 2.upload.The data representation in the blind storage will be in the form of array of blocks and the download operations is allowed to specify a list of indices of blocks to be downloaded, same way the upload operation is allowed to specify a list of data blocks and indices of those blocks. A blind storage system is constructed by three polynomial time algorithms on the client side: Bstore.keygen, Bstore.Build and Bstore. AccessAmong the three Bstore Access is an interactive protocol.

Bstore.keygen takes security parameter as an input andoutput as a k_{Bstore}. Note that K_{Bstore}, which the client is required to retain throughout the lifetime of the system, isrequired to be independent of the data to be stored.

Bstore.Build takes as input $(K_{Bstore}, d_0, \{id_i, data_i\}_i^t=1$ where k_{Bstore} is a key,d_0 is an upper bound on the total number of data blocks to be stored in the system to ,$(id_i, data_i)$ are the id and the data of the files that the system to be initialized with, it outputs an array of blocks D to be uploaded to the server.

Bstore.Access takes as input a key k_{Bstore}, a file id, an operation specifier op € {read, write, update, delete}, andOptionally data data (if op is write or update).Then it interacts with the server and returns a status message and optionally file data. For the update operation, Bstore.Access allows more flexibility first it requires only id as input, and outputs the current size of the file with that ID; then it accepts as input when the size of the file will be after update; then it outputs the current file data, only then requires the new data with which the file will be updated.

3. DRAWBACK EXISTING SYSTEM

The communication overhead is the major drawback of existing system. Data owner needs attribute verification for every data share among the set of users and hence the data owner needs more time for sharing every data with the data user. The existing paper uses AES algorithm which requires sharing of same key with the data owner and search user and hence there will be a leakage of data and this causes security issues.

4. ADVANVANTAGE OF PROPOSED SYSTEM

The proposed scheme uses RSA algorithm in order to provide more security for data owner. Access control is implemented in the proposed scheme so that the data user can use the data either to read or write. This will help the data owner to know about the data uploaded among the group of users.

5. PROPOSED ARCHITECTURE MODEL

The overall architecture of the proposed system is described in fig 1, in this the data owner will upload data in the cloud through Blind storage. Before uploading the data it is encrypted by using RSA algorithm. The text mining process can be done by Natural Language processing and Word net tools, it is used to extract the files and contents. The Natural Language Processing (NLP) process is used to extract the literal meaning words in file content. Word net tool is like dictionary. It is used to give the related synonyms to literal word in that content. Data user will try to search a query in the cloud server. The cloud server will map the keywords and search the related files. The cloud server gives 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 file name to the data owner. Then data owner knows all public key of user so it is encrypted by private key of data user public key and encrypted key is send to the server in turn the server will send that key details to user, then user decrypt the key using the private key. After that the data user gets the private key of data owner and then

access the data through blind storage. The file will be stored in the form of blocks and indexed by text mining process. Discretional access control is used where Read and Write access is given to the search user. Read access will allow the user to see the data what the data owner published, whereas write access allows the user to download and edit the data. Files which are published by in public cloud will be encrypted by Base64 algorithm which helps the receiver to decode the files easily.

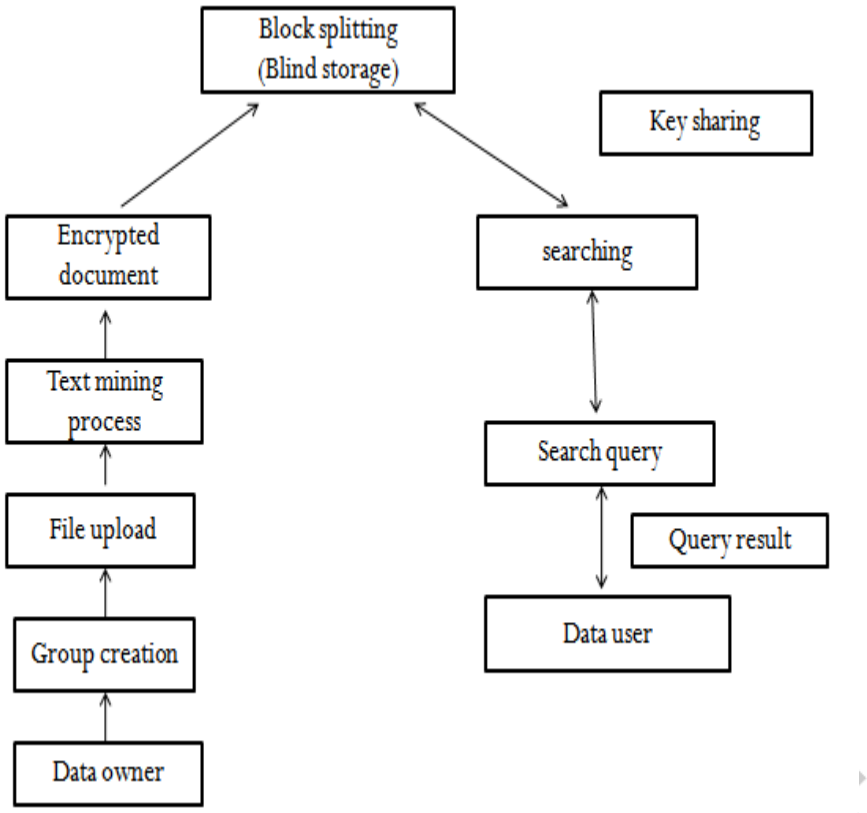

Fig 1: Architecture Diagram

5.1. MULTI-KEYWORD RANKED SEARCH EFFICIENCY

In order to provide the data what the user expert can be given with the help of keyword search. The traditional method used single keyword search for the retrieval of data which is needed for the search user. The single keyword search gives only some amount of data what the user need but this will not satisfy the search user. Therefore an efficient multi-keyword search has been proposed. To meet the requirement for practical uses and provide better experience, the EMRS[8] should not only support multi-keyword The efficient multi-keyword ranked search helps the server to get the data relevant to the given keywords to the user. The multi-keyword ranked search not only supports multi-keyword search over the encrypted data but also should support relevance based ranking result. It will return only data what the user is needed in order make better a ranked search has been proposed which will return the data in the ranked form related to the keyword. To overcome the problem of Boolean search Ning Cao [4] proposed multi-keyword ranked search over encrypted data in the cloud environment. The search index based on term frequency and vector space model with cosine similarity [7] has been used to achieve the higher search result accuracy. The K-nearest neighbour and relevance score technique [8] has been used in the blind storage for data retrieval in the cloud which will retrieve the data based on the multiple-keyword given by the user.

5.2. PROPOSED SYSTEM OVERVIEW

Cloud environment allows the data owner to store their data remotely on the cloud as to enjoy the on-demand high-quality applications and services from the shared pool of configurable computing resources. Its economic flexibility attracts many individuals and the business people. To provide the accurate data for the search user, the method keyword search is used. For efficient and secure search, multi-keyword ranked data searching is done in the context of cloud. The multi keyword search resultant based on the ranked search with efficient accuracy by using K-nearest neighbour techniques. The search efficiency can be improved to adopt Blind storage system to conceal access control issues in searchable encryption technique. The security analysis scheme demonstrates to achieve authentication and confidentiality of the document. The proposed scheme RSA encryption focuses on the security and access control

to enhance the security of data retrieval through blind storage using cloud environment.

5.3 TECHNIQUES USED IN PROPOSED SYSTEM

RSA algorithm is used for encrypting and decrypting the file in the cloud environment by the data owner. The RSA algorithm is a block ciphers in which the text are integers between 0 to n-1 for some n. RSA public key and RSA secret are the two point of integers in the scheme. Encryption and Decryption of the message is done using the RSA algorithm for making the communication secure. It is used for digital signature and key distribution. The strength of RSA is based on difficulty in factoring large integers. Especially those formed as product of two integers. The algorithm uses Number Theory concepts and modulo exponentiation, the Euler‟s function and the decryption is based on the Euler‟s theorem.

In public key cryptography the key has a public part and a private part. The public part is made known to everybody whereas the private key is kept secret by the receiver. Anyone who intends to send a message to the receiver encrypts the plain text using the public key corresponding to the receiver. Once encrypted using the public key, the cipher text can only be decrypted using the private key which is safe with the receiver. The algorithm has three steps:
1. Key generation,
2. Encryption,
3. Decryption.

Key generation
Key generation i) two prime numbers were selected such that X ≠ Z. ii) predict n = X*Z. iii) – Predict Ø (n) = (X-1) (Z-1). Ø (n) is called the Euler‟s Totient function. Two integers say X, Z are relatively prime if and only if common positive integer factor is one. iv) choose any number e when gcd (Ø (n), e) =1; 1 < e < Ø (n). v) Predict the value of d – de ≡ 1 mod Ø (n) or d ≡ e^-1 mod Ø (n). In perdition of „d‟ we need the multiplicative inverse of „e‟ modulo Ø (n). vi) Public key of RSA is {e, n}. vii) Private Key of RSA is {d, n}.

Encryption
The data owner needs to encrypt the data before outsourcing them to the cloud environment. For data sharing among the users the data owner need to generate two keys which also should be encrypted. If M is the original message then

Cipher text(C) = M^e mod n.

Decryption
For decrypting the data, the private key of the data owner must be given to the data user. If the cipher text is C then the original message or the plain text will be decrypted by using the mod function.

M = C^d mod n.

6. RELATED WORKS

Searchable encryption is a technique which provides the search service over the encrypted cloud data. It is classified as searchable public key encryption (SSE) and searchable symmetric encryption (SPE).The first SSE scheme has been introduced by song et al[1] which builds the searchable way but it only supports single keyword in the encrypted data. Naveed et al[2] constructs a blind storage system to achieve a searchable encryption however it supports only single keyword search in the encrypted data in the cloud. Block cipher AES algorithm has been used for encryption and decryption. Cong Wang [3] overcomes the problem of Boolean search which is the traditional method of searching technique which will meet the effective data utilization. This paper will assure “as strong as possible” security guarantee.

Moreover it explore relevance score from information retrieval to build a secure searchable index and develop a one-to-many order preserving mapping technique to properly protect those sensitive score information. Ning cao [4] proposed an efficient similarity measure of co-ordinating matching which results as many as possible data which is related to the multiple keywords given by the data user. Cong Wang [5] focused on retrieving the matching files in a ranked order regarding to their relevance criteria and it can be done through indexing. This will enable the quick search of documents that contain a given keyword. Bing wang [6] brings the solution for spelling error during the keyword search. The proposed scheme achieves fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need for predefined dictionary and effectively supports multi-keyword fuzzy search without increasing the complexity or index file. Wenhai sun [7] proposed a tree based index structure and multidimensional algorithm in order to improve the search efficiency in the practical world. The vector space model with cosine similarity and search index has been used in order to support multi-keyword search and ranked efficiency. Hong wel li [8] utilize the relevance score and K-nearest neighbour techniques to develop an efficient multi- keyword search that will return the ranked search result based on accuracy. This multi-keyword search efficiency has been used in the blind storage in cloud system in order to conceal the access pattern. Though the SSE helps in faster computation, but the security of the data is not sure since

the shared key is used for both sender and receiver. Jiadi yu [10] overcomes the problem of Boolean keyword search by using two round searchable encryption(TRSE) which will retrieve top K data which is related to the given keyword communication overhead is reduced by using vector space model and homomorphic encryption. Diffiehelman algorithm is used for encrypting and decrypting the data. The vector space model provides sufficient search accuracy and homomorphic encryption enables user to involve in ranking. Curtmola [11] proposed an idea of key sharing among the group of users. It is quite risk and it computes more time for sharing the data among the group of user. Hence Baojiang [12] solves the problem of key sharing among the group of users. The traditional method uses single key for every document sharing in a group, key aggregate searchable encryption has been proposed in which a single a single aggregate key is given to the user vector in a group to retrieve all the documents which is published by the data owner which will reduce storage, complexity and provides secure communication. Zhi-Hua Zhang [13] presents a identity-based authentication scheme foe application like e-business, this paper avoids the issue of revocation and key escrow problem in the authentication scheme based on the public key certificate. Qin Liu [14] addresses two issues privacy and efficiency and reduces the communication cost. Based on Aggregation and distribution layer (ADL) a middleware layer between the user and the cloud, the paper presents a scheme, termed efficient information retrieval for ranked query (EIRQ) to reduce querying costs incurred in the cloud.

7. CONCLUSION AND FUTURE WORK

In this paper we analysed various security issues in cloud computing based upon the previous works which is described in related work. The data security in blind storage is incorporated by using data access control technique in which the data are encrypted using RSA public key encryption algorithm. This algorithm is more efficient to secure data in blind storage whereas the server does not know the stored information about the files. The data are retrieved from the server by using efficient multi-keyword ranked search so that it can optimize search efficiency by reducing the time duration. The future work of this paper is to minimize the computation time needed to retrieve the data with essential security.

REFERENCES

[1] D. X song, D. Wagner, and A.Perrig, “Practical techniques for searches on encrypted data, “ in proc. IEEE symp. Secur.Privacy. May 2000, pp. 44-55.

[2]M.Naveed,M.Prabhakaran and C.A.Gunter,”Dynamic searchable encryption via blind storage,” in proc.IEEE symp.secur.privacy,may2014,pp.639-654.

[3]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.

[4]N Cao, C. Wang, M. Li,K.Ren, and W. Lou, “Privacy preserving keyword ranked search over encrypted cloud data,” IEEE Trans. Parallel Distrib. Syst., vol, 25, no. 1, pp. 222-223, Jan. 2014.

[5]C. Wang, N. Cao, J. Li, K. Ren, and W. Lou,”secure ranked keyword search over encrypted cloud data,” in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), jun.2010, pp. 253-262.

[6] B. Wang, S. Yu, W.Lou, and Y.T.Hou, “privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud,” in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112-2120.

[7]W Sun, B Wang, N Cao, M Li, W Lou, Y. Thomas Hou, H Li, “Verifiable privacy preserving multi-keyword text search in the cloud supporting similarity based ranking”, ,” IEEE Trans.Parallel Distrib. Syst., vol.25,no.11, pp. 3025-3034, Nov 2014..

[8]H li,D liu,Y dai,T H.Luan and Xuemin “Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage”.

[9]D. Bonch, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Proc. EUROCRYPT, 2004, pp. 506-522.

[10] J. yu, P. Lu, Y. Zhu, G. Xue, and M. Li,”Toward secure multi-keyword top-k retrieval over encrypted cloud data,” IEEE Trans.Dependable Secure Comput.,vol. 10, no 4,pp.239-250,Jul./Aug. 2013.

[11]R. Curtmola, J. Garay, S. Kamara, R. Ostrpvsky. “Searchable symmetric encryption: improved definitions and efficient constructions,” In: Proceeding of the 13th ACM Press, pp. 79-88, 2006.

[12]B. Cui, Z Liu, L Wang,"Key-aggregate searchable encryption for group data sharing via cloud storage," IEEE Trans.on computer., vol.6,no.1.

[13] H. Li, Y. Dai. L. Tian, and H. Yang, "Identity-based authentication for cloud computing," in cloud computing. Berlin, Germany:springer-Verlag, 2009,pp. 157-166.

[14] Q.Liu,C.C Tan, J. Wu, and G.Wang, "Efficient information retrieval for ranked queries in cost-effective cloud environments," in Proc. IEEE INFOCOM, Mar. 2012, pp.2581-2585.