IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

Preserving Text search privacy through blind storage towards secure Storage and retrieval of data

D.Kavitha[1], S.Hemavathy[2]

Assistant Professor, Department of Computer Science and Engineering, Valliammai Engineering College[1]

PG Scholar, Department of Computer Science and Engineering, Valliammai Engineering College[2]

Abstract: The blind storage is to preserve the outsourced data in cloud through gateway encryption and to implement multi-keyword ranked search over the encrypted data in a secure way by NLP process without downloading and decrypting the entire group member file contents. The access rights of the user are specified by the data owner in order to keep track of the files which are uploaded. The blind storage helps the data owner to save the data securely and it is not known to the server or user until they get the index file. Ranked search over the files help the server to retrieve what the file needed by the search user.

Keywords: Cloud Computing, Multi-keyword search, Blind Storage, Access Control.

I. INTRODUCTION

Research in cloud computing is receiving a lot of attention from both academic and industrial worlds. In cloud computing, users can outsource their computation and storage to servers called clouds using Internet. Clouds can provide several types of services like applications e.g., Google Apps, Microsoft online, infrastructures e.g., Amazon's EC2, Eucalyptus, Nimbus, and platforms to help developers write applications e.g., Amazon's S3, Windows Azure. Much of the data stored in clouds is highly sensitive, for example, medical records and social networks. Security and privacy are thus very important issues in cloud computing. In one hand, the user should authenticate itself before initiating any transaction, and on the other hand, it must be ensured that the cloud does not tamper with the data that is outsourced.

To make it secure the blind storage is adopted in the proposed system[2]. The blind storage helps the data owner to store a set of files with the remote server, revealing to the server neither the number nor the sizes of the files. The server would learn about the existence of a file only when the user retrieves it later. In the blind storage, all the files 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 the cloud server, it can only see the blocks of encrypted files uploaded and downloaded. Thus it leaks little information to the cloud and protects the files.

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[3]. 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.

II. RELATED WORKS

2.1 Dynamic Searchable Encryption via Blind Storage

Muhammad Naveed, Manoj Prabhakaran proposed a new storage scheme called Blind storage, which allows a client to store a set of files on a remote server in such a way that the server does not learn how many files are stored or the length of the individual file. Block cipher AES algorithm is used for encrypting and decrypting the file. To satisfy the client need single keyword search is implemented to retrieve the files related to the keyword. A new dynamic SSE scheme that is more efficient and simpler than prior schemes, achieving fully adaptive security. Dynamic SSE, which supports adding and removing document at any point during the life-time of the system[2].

Disadvantage

It supports only single keyword which will retrieve only small amount of file which will not satisfy the search user

DOI 10.17148/IJARCCE.2016.5291

fully. It does not include the scheme secure against corrupt servers. The access right of the search is not determined in the system.

2.2 Key-Aggregate searchable Encryption (KASE) For Group Data Sharing Via Cloud Storage

Baojiang Cui, Zheli Liu proposed that the traditional method of sharing the data among the group of users for different documents needs large number of keys to encrypt and decrypt the file. In order to solve this problem, the Key Aggregate Searchable Encryption (KASE) has been proposed to distribute the single key to the user in the group top share large amount of file and the user needs to submit a single trapdoor to cloud owner for querying the shared file. Diffie Hellman is used to prove the security of some broadcast encryption scheme. It is an effective solution for building practical data sharing system based on public cloud storage [11].

Disadvantage

KASE cannot be applicable to the federated cloud. Federated cloud is the deployment and management of multiple external and internal cloud computing services to match business needs. Combination of union of several smaller parts that perform a common action.

If a user wants to query over documents shared by multiple owners he must generate multiple trapdoors to the hence this problem is yet to be solved.

2.3 Toward Secure Multi-keyword Top-K Retrieval Over Encrypted Cloud Data

Jiadi Yu, Peng Lu observed that the data privacy issue using SSE is discussed in this system. The order preserving encryption (OPE) which is done in server side for ranking the file leaks data privacy and hence the security is not guaranteed. To eliminate the leakage they proposed two-round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval. The TRSE employs a vector space model and homomorphic encryption. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while majority of the computing work is done on the server side by operations only on cipher text. By which the information leakage is eliminated and data security is ensured[10].

Disadvantage

The practical efficiency of the system is difficult in the real world environment compared to other system. The data leakage is reduced thereby protecting the data but it is done at some level and need to give high security for the data.

2.4 Privacy Preserving Multi-keyword Ranked Search over Encrypted Cloud Data

Ning Cao, Cong Wang observed that the usage of single keyword and Boolean search will give only some amount of file which is related to the keyword. To solve this issue and make the search user satisfy, for the first time they proposed Multi-keyword ranked search over encrypted data in cloud computing (MRSE). The search user will give multiple keywords to the server and in turn the server will retrieve the files which are related to the keywords.

The server uses Multi-keyword semantics called coordinate matching is used to find as many matches as possible, to capture the relevance of data documents to the search query. Inner product similarity is also used to quantitatively evaluate such similarity measure [4].

Disadvantage

The files which are related to the keyword will be given to the user and hence it will be confusing to the user to choose which file. Hence the ranking of such files help the user to know which is relevant to the keywords given. Hence the ranking and relevance scoring is main disadvantage of this system.

2.5 Verifiable Privacy-Preserving Multi-keyword Text Search in the Cloud Supporting Similarity Based Ranking

Wenhai Sun, Bing Wang proposed privacy preserving multi-keyword text search (MTS) to solve the practical efficiency of the search user by retrieving the files which are related to the keyword.

To support multi-keyword search and search result ranking, search index based on term frequency and vector space model with cosine similarity measure is used to achieve higher search result accuracy. Tree based index has been used to improve the search efficiency and adaptive methods for multi-dimensional (MD) algorithms is used to improve the practical efficiency [12].

Disadvantage

The algorithm takes long time to get back the results and the time taken for retrieval of the files are longer and hence need to improve the efficiency of the algorithm is the important.

2.6 Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data

Cong Wang, Ning Cao observed that the traditional searchable encryption techniques allows the users to securely search over encrypted data through keywords, they support only Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud.

Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. The statistical measure of relevance score is introduced to improve the search efficiency. Further they developed the one-to-many order-preserving mapping technique to properly protect sensitive score information. The data security system is “as strong as possible” [8].

Disadvantage

The efficiency of the system is difficult and the data retrieval takes lot of time so usability and reliability is not supported well.

DOI 10.17148/IJARCCE.2016.5291

III. OVERALL ARCHITECTURE

The figure [1] shows the overall architecture of the system. The proposed system has three entities: 1.Group owner, 2. Search user 3. Cloud server. The group owner will have a large collection of file F which will be encrypted before storing it in the cloud server. The encrypted data is E. Then the data will be saved in the blind storage which the data is splitted and saved in the different blocks each block will be indexed by numbers and it will also be saved in the blind storage. The group owner creates an index file (I) for the stored data which contains important keywords which will be used by the server to deliver the result of search user. If the search user needs any file then he gives keywords d which contain multiple words. The cloud server in turn search the keywords in the index file and if the file match with the keyword d then the list of files will be given to the search user. If any file match with the keyword then he needs to get permission from the group owner, then the group owner will add him to the group and share the index file I to the search user. For providing more security the group owner uses Discremenational Access control where read and write access will be given to the search user.

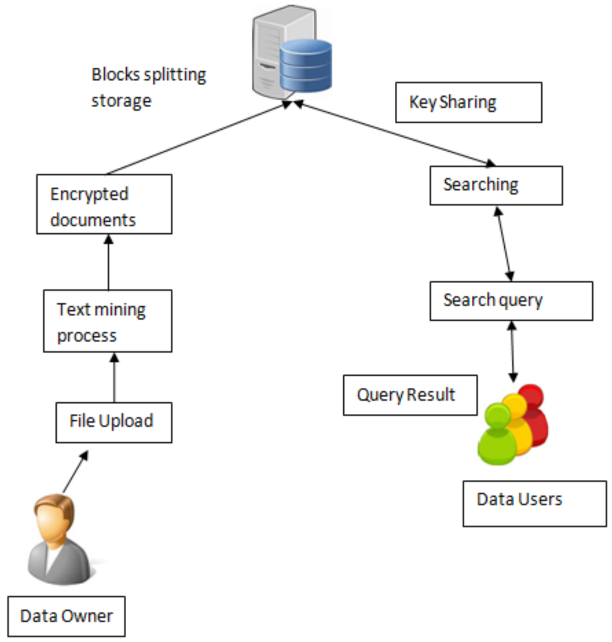


Fig 1 Architecture Diagram

IV. GROUP CREATION AND KEY GENERATION

The data will be saved in the cloud and if the owner wants to keep it secret need to share only to the user which the owner trust and wish to share the data. Data owner should be register in this environment and create a group. If the search user needs the some file he need to give request to the server and if it match with the private owner data then the server will connect the user with the owner, if the owner wish the share the data with the user he need to add him to the group. Data users also register and give request to group owner to add a group user. Data owner accept the request from the user. Multiple groups can be created. Data user only can access the respective data owner documents. Data user cannot access the webpage until the data owner accepts the request. Only with the acceptance of the owner the user will able to get the index file and view the contents of the file which is stored.

The RSA algorithm is used for generating key for sharing the data among the users. There will be two keys public and private. If the owner wants to share the data with the user then the public key is given and the file is encrypted by using this key. The user will decrypt the key by using his private key. Hence the data share is done by using the key and the file get encrypted and decrypted by this key. RSA algorithm is more secure and it consume some time for computation and the security of the file will be more and thus it is difficult to hack the file of the user.

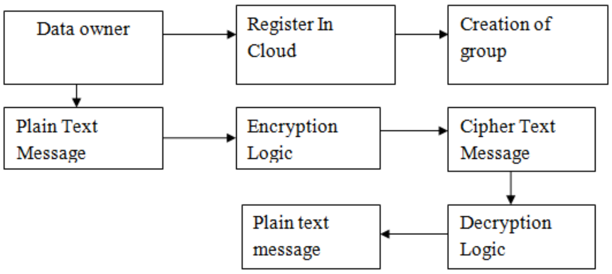


Fig 2 Group creation and key generation

V. INTERESTING TEXT MINING AND PROVIDING ACCESS RIGHTS

This module is used for extracting the file from the cloud related to the keywords given by the search user. The search user will give keywords of what he wants in turn the server will search in the cloud and this process is called as text mining. The server uses NLP and word net tool for mining the text and retrieve the file which are related to the keywords. The access right of the user is important for keeping the files more secure and thus without the knowledge of the owner the data will not be known to the other search user.

The proposed system uses discreminational access control in order to give read and write access to the search user. The owner can give either read or write access to the user and thus making the file in his control. If it is read access then the search user can view the content of the file by using the index file and cannot download the file by this access right. If it is write access then the user can download the file and by using the index file he can edit it and send to the data owner.

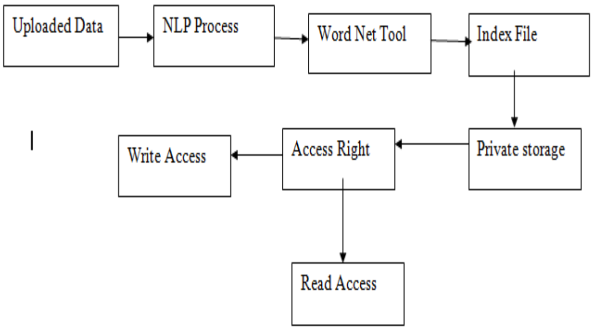


Fig 3 Mining of text

DOI 10.17148/IJARCCE.2016.5291

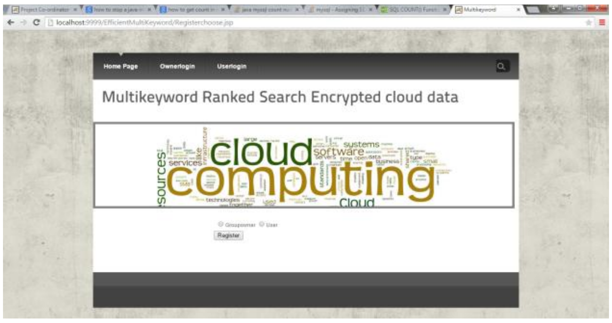
VI. ALGORITHM

Step 1: Creation of Group by Data Owners.
Step 2: User Registration And Group Owner registration.
Step 3: Addition of data user through owner to group.
Step 4: Owner upload in public or private cloud.
Step 5: Data in public mode is visible to all.
Step 6: If the data is in private mode owner assigns access rights.
Step7: Access rights given by the owner will be in discrimination fashion.
Step8: User open and view that data through read access download and rewrite is allowed in write access.
Step9: Interesting keywords are mined through NLP process and Word net tool.

Implementation process:

Registration For owner and user:

Registration page for registration of data owner and search user.



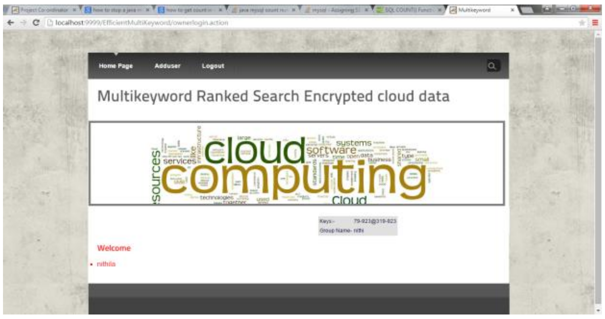
localhost:9999/EfficientMultiKeyword/Registerchoose.jsp

Multikeyword Ranked Search Encrypted cloud data

Fig 4 Registration

Key Generation:

Key will be generated by using RSA algorithm using which the data can be shared.



localhost:9999/EfficientMultiKeyword/ownerlogin.action

Multikeyword Ranked Search Encrypted cloud data

Keys:- 79-923@319-923
Group Name- nithi

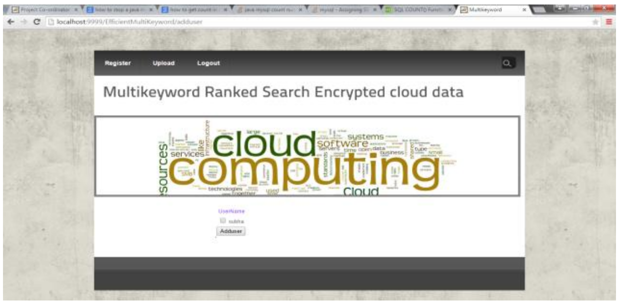
Welcome

- nithila

Fig 5 Key Generation

Adding User to group:

Data owners can add the members to the group in order to share the data with the search user.



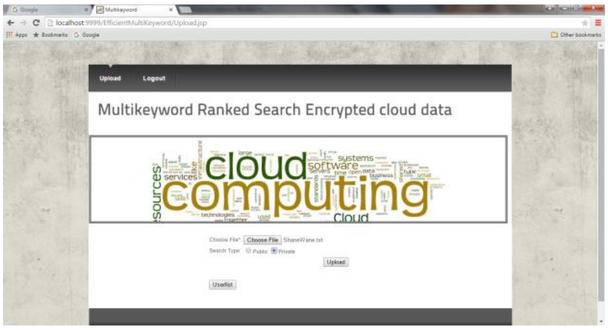
localhost:9999/EfficientMultiKeyword/adduser

Multikeyword Ranked Search Encrypted cloud data

Fig 6 Adding User to group

Uploading in Private mode:

Data can be uploaded in private mode which can be used by the group users.



localhost:9999/EfficientMultiKeyword/Upload.jsp

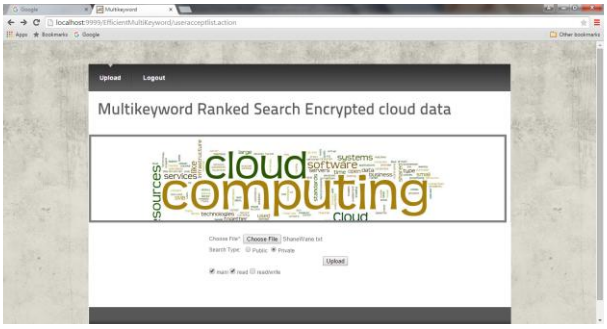
Multikeyword Ranked Search Encrypted cloud data

Choose File*: ShaneWane.txt
Search Type: ○ Public ● Private

Fig 7 Uploading in private mode

Access rights to user:

Discreminational access control is given to data user to provide either read or write access.



localhost:9999/EfficientMultiKeyword/useracceptlist.action

Multikeyword Ranked Search Encrypted cloud data

Choose File*: ShaneWane.txt
Search Type: ○ Public ● Private
☑ mani ☑ read ☐ read/write

Fig 8 Access right of user

VII. CONCLUSION AND FUTURE WORK

In the proposed system the problem of sharing a file in the cloud environment is discussed, the system uses blind storage which helps the user to store the file in a secure way and indexing is given for each file for easy retrieval. An attempt was made to improve the data discovery and user searching experience by supporting secure multi-keyword ranked search (SMRS) scheme. It is used for efficient and secure search over the encrypted cloud data.

The proposed scheme can effectively achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. 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. The resulting scheme is more efficient by reducing the communication overhead and effectively achieves confidentiality of files and index. The future of the system is to minimize the computation speed, to support a ranked search, which allows users to search even with misspelled keywords and achieve it in the multi-cloud environment.

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.

DOI 10.17148/IJARCCE.2016.5291

[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] 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.

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

DOI 10.17148/IJARCCE.2016.5291