# Multi Keyword Retrieval On Secured Cloud

Hussain Abo Surrah
College of Computers and Information Technology
Taif University,
Kingdom of Saudi Arabia
salama366@yahoo.com

*Abstract*-**An efficient method of cryptography with multiple keyword searches is proposed in this paper. This paper aims to provide searching a file over cloud environment using multiple keywords representing the file with various probable situations. The aim is to provide the security to its maximum extent by including encryption and decryption methods. Authorization of the users directly by the administrators allows the files involved to transfer more securely. Encryption and decryption of both file name and file which uses asymmetric and symmetric key algorithms respectively. The secret key is generated for each user to prevent any other user to misuse the file. The implementation of the project in the simple manner provides an easy to understand environment for the user. By using server side process to hold the most processes, the user can reduce the systems inefficiency. Client side system uses less work corresponding to the task needed to perform the basic role like ranking and arranging the files in the requested order. The solution can be applied to various applications because of is simplicity.**

*Keywords:* **Cryptography, encryption, multi keyword**

## 1. INTRODUCTION

Information security, efficiency enhancement plays a vital role in computer technology. Due to various enhancements in the technologies, the need of the new ideas has been increased. Data security has become less reliable when using third party infrastructures like Cloud services. Cloud Storage of data provides various features for the customers by providing the reliable security. Even though they are used by trusting on basis of license agreement, there might have a chance of data leakage. The security provided by the trusted third parties may be misused by various methods. In order to overcome this kind of data abuse, the data owner can make the data as cipher and then can be uploaded to the server. By encrypting the data that is being uploaded is prevented from any harmful attacks. There are three types of cryptography that can be used such as symmetric, asymmetric and hash key cryptography to encrypt the data.

Symmetric key cryptography uses a single key, Asymmetric key cryptography uses two keys and Hash algorithm does not use any key to encrypt and decrypt data. The three types of cryptography are used in the project at various situations to provide high secure data with maximum efficiency that can be achieved.

The project provides user registration, authentication of users, uploading data with encryption, download data with decryption, generation of independent file code for each user, etc.,. The easy to use environment with sophisticated techniques is developed. The project will provide maximum efficiency by using secure as well as efficient algorithms such as MD5 and AES.

The project will support multi keyword searching which helps the user to find any related materials that they need. The ranking of the file based on the usage of the file is listed and then the list is show as per the maximum used files. By doing the ranking process at the client side the data leakage can be prevented.

Data security now a day's become a severe issue in the emerging cloud technology. The data that users are storing in the cloud database are vulnerable to various attacks. Hence a maximum security is needed. In order to reduce the data loss over cloud, there should be data in encrypted format. There are many possibilities that can leak the data by obtaining statistical leak aging, scheme robustness and similarity relevance. The data even in the encrypted format can be easily attacked by the attackers.

Cloud computing becomes part of internet based applications which is an emerging technology in various organisation. The data that are stored in the cloud has to be protected completely from any attack that is caused both by external and internal attackers. Most of the internal attacks are used by the cloud providers by using similarity relevance and analysing the statistical leakage. Based on the usage of the file over ranked manner, it is easy to get all the details of the most used files through probability prediction. This kind of data leakage should be completely avoided and maximum protection to the data is given. The solution suggests the same by applying some new concepts to increase the data security.

## 2. LITERATURE REVIEW

The system under working in today's environment stores the encrypted file on the server that allow user to identify the exact file work which has largely focused on searching consist single keyword. File encryption [1][2], however makes it hard to retrieve data from server. Public information retrieval technique provides a service for retrieve the data from public database.

Understanding the algorithm by using key generation, public keys, trapdoor, testing [3]. The user runs algorithm two times to generate two public/private key pairs. User produces trapdoor using the private key. The server simply sends the relevant emails back to user and call such a system non-interactive public key encryption with keyword search, or as a shorthand searchable public key encryption [4]. This mechanism uses multiple times the same process for every single key word provided hence it will take more time to execute. The security of the system also depends on hard assumption.

The searching of the existing system based on public key cryptography [5]. There is possibility of every user have the public key and different private key. The sender of the system uses the receiver's public key to encrypt the data [6]. On the other hand receiver with a keyword will search for the keyword in the data that is being sent to him without involving full data but only with the keyword that the receiver wish to provide. This kind of searching will only provide the data related to the keyword by extracting the encrypted data [7].

The homomorphism encryption scheme which is partially implemented [8], which is less security and less reliable. The system enhances the scheme into fully homomorphism using some basic modular arithmetic operations and Gentry's techniques [8][9]. This also reduces the security to check out the approximate data. The comparison of the system analyse the hardness of the developed system.

The inner product similarity with coordinate mapping to get as much possible related files as it can have by comparing the encrypted data [4]. The ranking is done after retrieving the most similar and related file. The number of usage of the files with various times difference has also used to do the rank based retrieval mechanism. The data enabled services are also monitored with security and the efficiency are maintained by simplifying the algorithms by considering the security no to go down.

## 3. SYSTEM DIAGRAM

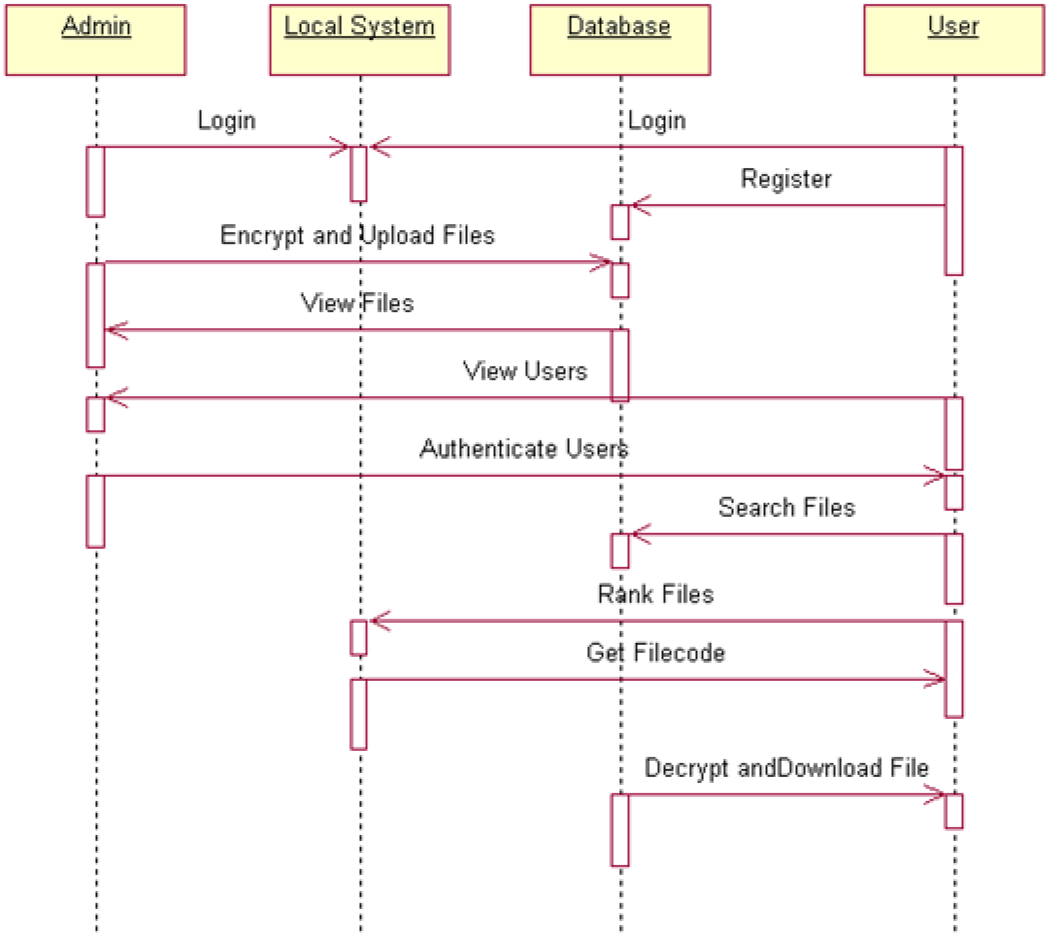

Figure 1: System flow diagram

## 4. METHODOLOGY AND APPROACH

### 4.1 USER AUTHENTICATION

The User registration and authentication of unregistered users comes under this module. When a new user wants to join the system then the user should be a member to access the file. In order to access the files the user has to be registered to the system by providing the basic details such as Username, password, Email Id and Mobile number. After a process of registration the user still not able to access their own account due to the unauthenticated status of the account. The user can only login to their account after the authentication of the administrator who can view the list of unauthenticated users.

The administrators can also delete the full record of any file. The administrator selects the user and by authenticating the particular user the account is activated and the account is ready to use. After the authentication of the user account, the user can log on to their account through the user account and not as administrator. The module contains another process that is the login process. There are two types of users to login in the system. One is the user by leaving the checkbox unchecked and the other one is the administrator. The login credentials are verified from the database and user or admin home page is displayed.

### 4.2 ENCRYPTED FILE UPLOAD

File Encryption and uploading of file in the database is done by the administrator only. The file is chosen to upload by using a file upload control and the name of the file is obtained by the text box control. The required details are checked for any data loss before uploading. The file chosen is then used to get the extension of the file which is used at the user end to save the file type, Content type of the file in order to know what kind of content is to be stored in the Database, File Name of the file, Encrypted Name of the File and File content as byes are stored in the Database.

Filename is encrypted before storing in the database. The encrypted filename is obtained by using hashing and MD5 encryption Algorithm. The Method used in the Algorithm is ECB (Electronic Code Book) provided by the .Net environment. The encryption algorithms provide high security to the file that is being encrypted. The file encrypted is then stored with the File into the database. The file can contain most used file types like text, image, application, etc.

File that is being encrypted uses AES (Advanced Encryption Standard) and stored in the system. The Encryption of the file uses symmetric key algorithm which uses single key to encrypt and decrypt data where encryption and decryption of file name uses asymmetric key algorithm which uses two keys.

File is read by bytes from the file uploaded. Key and IV are generated using derived bytes and symmetric key. Using the Key and IV the data is encrypted and written into byte array and the byte array is converted to base64String which is the encrypted message and is stored in the memory.

Files that are uploaded can be viewed by the administrator and if the file is not needed then admin can remove the file by deleting it from the database.

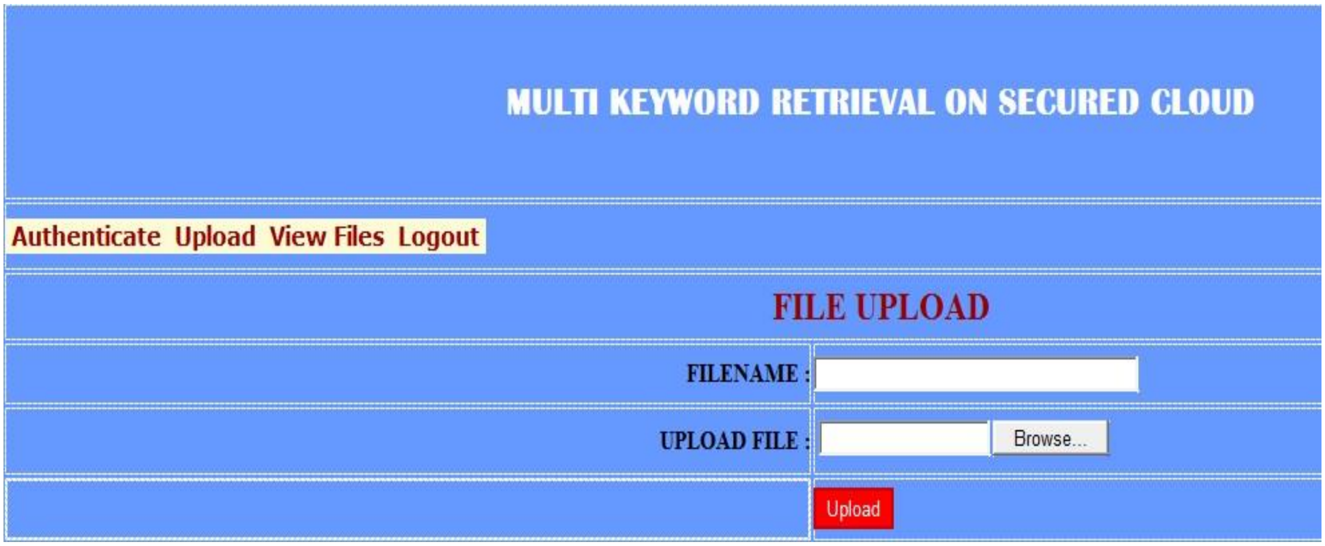
MULTI KEYWORD RETRIEVAL ON SECURED CLOUD
Authenticate Upload View Files Logout
FILE UPLOAD
FILENAME :
UPLOAD FILE :
Browse...
Upload

Figure 2: File upload

### 4.3 SEARCHING ENCRYPTED FILE

Users can login to their account after the authorisation of their account by the administrator. Users have the access right now to the data that are available throughout the system's database in the cloud. After the login process is completed by the users they can go to the search menu which will provide the facility to search the entire database. Searching process involves generation of the rank list based on the usage of the files which we are searching. The searching method which allows multi keyword, the user can type in any related words for the file which they are processed based upon the file usage. The ranking is done at the client side and the processes consuming more time and memory are diverted to the server. The server maintains the hardest processes. The client only does the ranking. The client side ranking prevents the data leakage and vulnerable attacks.

Searched files can now be displayed to the user. Users can now select the file appropriate to their search which always the first file in most case. Send code is the option provided to the users in order to send the file code which is the encrypted filename for the current user to the cloud storage. The cloud storage will then be used to send the File code to the current user's Email. Email account is used as cloud storage in the project. The file code that is sent to the user's email is the encrypted by combining the user's identification hence none other than the requested user can download the file. Users who logged in using the same identification can only download the file.

The download process does the decryption of the file using symmetric key algorithm.

### 4.4 FILE DOWNLOAD

User who requested the file can have the file code which is encrypted with the user identification prevents other from accessing the file. Requested user should be logged-in in order to download the file. Decryption process contains two stages. First the decryption of the file code is done and the user identification is verified and separated from the file code with the help of the variable user that is created at the time of login.

Users who use others file code will fail to download the file. The Encrypted filename is now separated from the user identification and then the decryption of the file name can be done. After the decryption of the file name using the similar algorithm used for encryption in a reversed manner, the file name is obtained. Now the file with the name decrypted can be used to query over the database and can be downloaded.

The file contains encrypted form of data, so it has to be converted to store as in a proper format. Every data about the file is extracted from the database such as File name, Extension, Content Type and Binary form of data after the extraction the data acquired as bytes are converted into the original file format through HTTP Response and the user can open or save the file. Before doing the byte conversion, the file encrypted using symmetric key algorithm has to be decrypted with the same AES algorithm. After decrypting the data, the bytes with clear form is obtained. It is then converted to its original format from bytes. Now the file is downloaded in a more secured way.

## 5. CONCLUSION

The solution, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements [10].Among various multi-keyword semantics[11], the chosen efficient principle of "coordinate matching"[1], i.e., as many matches as possible[12], to effectively capture similarity between query keywords and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity measurement[13]. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, The proposed solution provide secure inner product computation, and significantly improve it to achieve [12] privacy requirements in two levels of threat models. Thorough analysis investigating privacy [14] and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset [14] show our proposed schemes introduce low overhead on both computation and communication [15].

As our future work, we will explore supporting other multi-keyword semantics (e.g., weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in stronger threat model.

## 6. ACKNOLEDGEMENT

This work is part of my research and development; I sincerely thankful for the anonymous reviewers whose comment greatly helped clarify and improve this paper

## REFERENCES

1 .Jiadi Yu, Member, IEEE, Peng Lu, Yanmin Zhu, Member, IEEE, Guangtao Xue, Member IEEE Computer Society, and Minglu Li, "Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data"

2. S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.
3. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. of S&P*, 2000.

4. Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in *Proc. of ACNS*, 2005.

5.D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in *Proc. of EUROCRYPT*, 2004.

6.J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in *Proc. Of IEEE INFOCOM’10 Mini-Conference*, San Diego, CA, USA, March 2010.

7. William stallings,”Cryptography and Network Security, “in second edition

8. H. Hu, J. Xu, C. Ren, and B. Choi, “Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism,” Proc. IEEE 27th Int’l Conf. Data Eng. (ICDE), 2011.

9. D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, “Public- Key Encryption with Keyword Search,” Proc. Int’l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.

10. “Privacy Preserving Multi keyword Ranked” http://www.docstoc.com/docs/93989001/Privacy Preserving-Multi-keyword-Ranked-Search.ppt.

11. Ankatha Samuyelu Raja ,Vasanthi A,, “Secured Multi-keyword Ranked Search over Encrypted Cloud Data” International Journal of Advanced Research in Computer Science and Software Engineering. Volume 2, Issue 10, October 2012 ISSN: 2277 128X.

12. M.Sandhya, CH. Raja Jacob “Performance of SKSE and MRSE in Cloud Cache” ISSN: 0976-8491 (Online) | ISSN: 2229-4333 (Print) IJCST Vol. 3, Issue 2, April - June 2012.

13. Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data http://www.vidhatha.com/upload/ANDROIDPROJECTS/SYNOPSIS/ADP002%20-%20Privacy-Preserving%20Multi-keyword%20Ranked%20Search.doc.

14.Cong Wang , Li, Ming , Kui Ren , Wenjing Lou,Privacy-preserving multi-keyword ranked search over encrypted cloud data INFOCOM, 2011 Proceedings IEEE, DOI:10.1109/INFCOM.2011.5935306 ISBN:978-1-4244-9919-9.

15. Karapakula, A. ; Puramchand, M. ; Rafi, G.M. “Coordinate matching for effective capturing the similarity between query keywords and outsourced documents” IET Chennai 3rd International on Sustainable Energy and Intelligent Systems (SEISCON 2012),DOI: 10.1049/cp.2012.2246 , ISBN :978-1-84919-797-7.