Enhanced Secure Multi Keyword Top-K Retrieval in Cloud

Ahmad Tasnim Siddiqui, Syed Ajaz, Isbudeen Noor Mohamed

College of Computers and Information Technology, Taif University, Saudi Arabia

***Abstract*—*This research is capable to do cryptography with multi keywords search. This research is target to provide search files from cloud network using multi keywords. This paper is target to provide a security at the maximum level by includes encryption and decryption. The administrator has control of authorization and allowing files to move more secured. Encryption and decryption of files and file names which is used symmetric and asymmetric algorithm respectively. The unique key is generated for every users to protect other user cannot access the files. While implementing this project the user can understand very simple environment. The user can reduce incapable systems in server side process to hold most of the processes. The client side system has used less work for the corresponding task to perform the necessary role like arranging and ranking the files from requested order. This project can apply in various applications for this user friendly.*

***Keywords*— Cryptography, encryption, multi keyword.**

I. INTRODUCTION

Information security, efficiency improvement plays a very crucial part in computer technologies. Due to the diverse improvements and advancements in the field of technology, the need of the new thoughts has been greater than before. Data protection and safety has become very crucial and less reliable due to the use of third party services like Cloud services. Cloud Storage provides a range of features for the consumers by providing the consistent safety measures. Normally they are used by trusting on basis of license agreement; still there is possibility of data outflow. The security provided by the trusted third parties may be distorted by different type of methods. In order to prevent from this kind of data exploitation, the data vendor can build the data as cipher data and then they can upload at the server. By using encryption technique the uploaded data can be prevented from any attacks. There are three types of cryptography techniques which we can use. They are symmetric cryptography, asymmetric cryptography and hash key cryptography to encrypt and protect the data.

In symmetric cryptography, it uses a single key, where asymmetric cryptography uses two keys but hash key cryptography algorithm does not use any key to encrypt or decrypt data. These are three types of cryptography techniques which are being used in the project at various points in order to provide high level security to data with utmost possible efficiency.

This project provide modules for the user registration, user authentications, encrypted data uploading, downloading data using decryption, generation of independent file code for every user, etc.,. The project uses user friendly environment with sophisticated techniques. This will provide utmost efficiency by using secure as well as competent algorithms such as MD5 and AES.

As it supports multi keyword searching which helps the user to find any related materials that they need? The ranking of the file based on the usage is listed and then the list is shown as per the maximum used files. By the ranking method at the client side we can prevent and reduce the data leakage.

Now a day's data security has turn out to be a crucial issue in the emerging field of cloud technology. The data that users are maintaining in the cloud are exposed to different types of attacks. Therefore an utmost security measures are desirable. In order to reduce the risk of data loss over cloud system, data should be in some encrypted format. There are several possibilities that can cause the data leakage by obtaining statistical leak aging, scheme robustness and similarity significance. The data still in the encrypted format can be effortlessly attacked by the hackers.

In various organization, cloud computing has become part of the internet based applications which is a rising technology. The data which are being stored in the cloud has to be protected completely from any type of expected or unexpected attacks that can be due to both external and internal attackers. Maximum of the internal attacks are done by the cloud service providers by using similarity significance and analyzing the statistical leakage. Based on the usage of the file over ranked manner, it is easy to obtain entire details of the most used files through probability technique. This type of data leakage should be totally avoided and utmost protection to the data is given. The solution suggests the same by applying some new concepts to enhance the data prevention and security.

II. LITERATURE REVIEW

These days system working under cloud environment stores the encrypted file at the server which permits user to recognize the exact file which has largely focused on searching single keyword. File encryption [1] [2], though makes it tough to access and download data from the server. Public information retrieval technique provides a service to regain the data from public database storage.

This algorithm is used to produce key generation, public keys, trapdoor, testing [3]. User runs this algorithm twice to generate two public or private key pair. User can produce trapdoor by using private key. The server then sends the appropriate 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 technique uses the same process multiple times for every single keyword given therefore it will require more time to perform the task. The security of the system also depends on hardware configuration.

While searching of the existing system based on public key cryptography [5], there is always option for every user has the public key and different private key. Users use the receiver's public key to encrypt data [6]. On the other side, receiver having a keyword will look for the keyword in the data which is being sent to him without linking full data but just with the keyword that the receiver wish to grant. In this sort of searching, the data related to the keyword by extracting the encrypted data will only be provided [7].

The homomorphism encryption system which is partly implemented [8] is less secure and less reliable. The system improves the scheme into fully homomorphism using basic modular arithmetic operations and Gentry's techniques [8] [9]. This also reduces the security to check out the estimated data. The comparison of the system analyzes the hardness of the developed system.

The inner product similarity with the coordinate mapping to get as much possible related files as it can be done by comparing the encrypted data [4]. The ranking can be done after getting the most likely and related file. The number of usage of the files with a range of time variation has also been used to perform the rank based retrieval system. The data enabled services are also under observation. Security and efficiency are being maintained by simplifying the algorithms by considering the security no to go down.

III. METHODOLOGY AND APPROACH

A. User Authentication

The new user verification is comes under this module. Whenever the new user can join to the systems the user must be a member of the systems to access the files. In order to access the files the user has to be in providing the details of his name, secret code, Email address and contact number. After the new user registration if the user still cannot able to access his account due to unauthenticated status of his account. The new user can log into the system after his account has authenticated by the administrator .The administrator have rights to view the list of new registered users.

The administrators have rights to delete the records at any time. The administrator can select the particular user by authenticating on his account to be activated and the account is ready to use. After the authentication of the user account, the user can log in to his account ass a regular and he doesn't have the administrator rights. This is module contains other part that is user login part. There are two types of users to login in the system. The user is leaving the checkbox as unchecked for regular user and another part is for the administrator login. Once login credentials have been verified from the server and the user or admin home page will be displayed.

B. Encrypted File Uploading

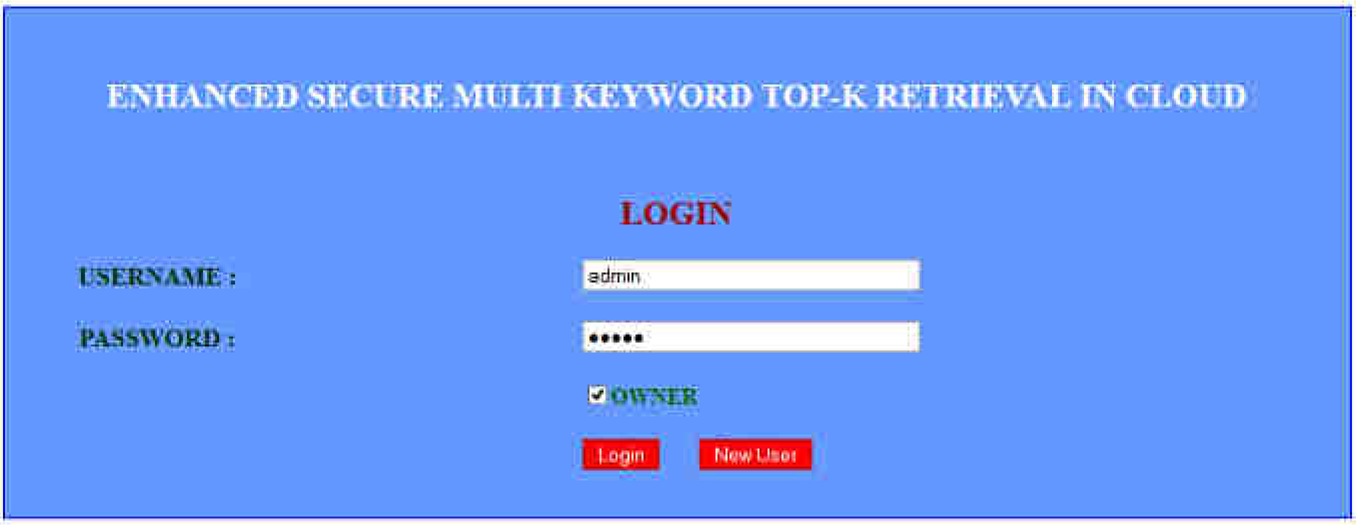
The Encryption and uploading of files in the data server is controlled by the administrator. The file is chosen for upload used by a file upload control button and contains name of the file is entered by the text box control. The required details have checked before uploading files. Chosen file is then used to get the extension of the file which is used by the user is try to save the file type and content type of the file in order to know what kind of content is to store in the data server, file name of the file, encrypted name of the file and file content as byes are stored in the data server.

The filenames are encrypted before entering in to the data server. The encrypted filename is obtained by using hashing and MD5 encryption algorithm. This method used in this algorithm is ECB (Electronic Code Book) provided by the .Net environment. This encryption algorithm provides highly secured to the file that is being encrypted. The file encrypted is then stored with the file into the data server. The files could contain most used file types like text, image, application, etc.

The files that are being encrypted by use AES (Advanced Encryption Standard) and stored in the system. The Encryption of the files uses symmetric key algorithm which is uses single key to encryption and decryption data where encryption and decryption of file names uses asymmetric key algorithm which is uses two keys.

The file is read by bytes from the file uploaded. Key and IV are generated by using derived bytes and symmetric key. Using the Key and IV the data is encrypted and written into byte array and this byte array is converted to base 64 String which is the encrypted message and is stored in the memory.

These files are uploaded and could be reviewed by the administrator and if the files are not needed then admin could remove the files by deleting it from the data server.


ENHANCED SECURE MULTI KEYWORD TOP-K RETRIEVAL IN CLOUD

LOGIN

USERNAME : admin

PASSWORD : •••••

☑ OWNER

Fig.1: File upload

C. Searching Encrypted File

The users can login to their account after the authorization of their account by the administrator. The 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 user can go in to the search menu which will provide the facility to search the entire database. The searching processes involve generate the rank list based on the usage of the files which we are searching. The searching method which is allowed multi keyword, the users could type in any related words for the file which they are processed based upon the files use. The ranks are done at the client side and the processes consuming more time and memory are diverted to the database. The database is maintains the hardest process. The client only does the ranking. The client side rank is protecting external attacks like the data leak and vulnerabilities.

Searched files can now be displayed to the user. The user could select the files from his search which always the first file in most case. Send code is the option provided to the user in order to send the files code which is the encrypted filenames for the current user to the cloud. The cloud storage will then be used to send the file code to the current user email. This email account is used for cloud storage in the project. The file code that is sent to the user email account and it’s encrypted by comparing of the user’s identification hence none other than the requested user can download the files. The user who logged in used the same identification can only download the file.

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

D. File Download

The user who requested the files can have the file code which is encrypted with the user identification prevents other from getting access of file. User must be logged-in to download any file. Decryption process has two stages. First is decryption of the file code which 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.

If the users are using other file code they cannot download the file. The Encrypted filename is now separated from the user identification and then the decrypted file name can be done. After the decrypted file name using the similar algorithm can be used to encrypt file in a reversed manner and we can obtain file name. Now the file with the decrypted name can be used to query over the database and user can download it.

This file contains data in encrypted form, so it has to be converted to store as in a specified format. Entire data related to the file is extracted from the database such as file name, extension, content type and binary form of data. After the extraction of information, the data acquired as bytes has to be converted into the original file format by using HTTP response and at this stage users have the option to open or save the file. Before doing the byte conversion, the file encrypted using symmetric key algorithm must be decrypted with the similar AES algorithm. After decrypting the data, the bytes with clear form is obtained. It is then converted to original format from bytes. Now the file is downloaded in a more secured way.

Below is the system flow diagram of proposed system:

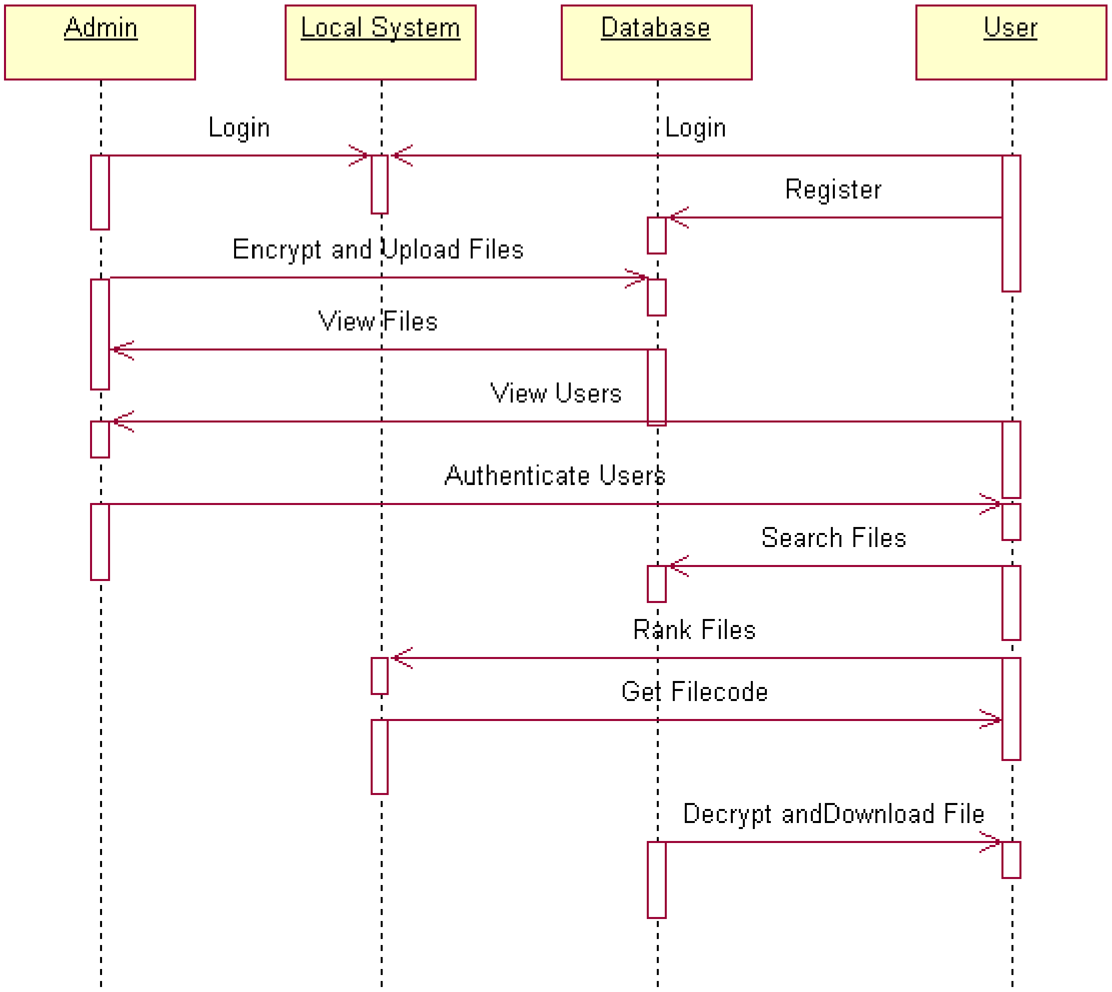


Fig.2: System flow diagram

IV. 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 range of privacy requirements [10]. Along with the variety of multi-keyword semantics [11], the chosen efficient principle of "coordinate matching" [1], i.e., as many matches as possible [12]. To capture the similarity between query, keywords, 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 provides secure inner product computation, and considerably improves it to achieve [12] privacy requirements in two levels of threat models. Thorough analysis investigating privacy [14] and efficiency guarantees of proposed schemes, and our experiment on the real world dataset [14] shows that proposed scheme introduces 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.

ACKNOLEDGEMENT

This work is part of our research and development; we 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