Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 5, Issue 2; April-June, 2018, pp. 60-63

http://www.krishisanskriti.org/Publication.html

A Survey on Multi-Keyword Ranked Fuzzy Keyword Search over Encrypted Cloud Data

Saba[1] and Prof. S. C. Karande[2]

[1,2]Department of Computer Engineering Maharashtra Institute of Technology, Pune
E-mail: [1]saba.cs14@gmail.com, [2]shridevi.karande@mitpune.edu.in

Abstract—*As cloud computing has grown in various fields; it is used to store bulky data. The storage helps in anytime and anywhere access of the data that is stored over the cloud. The cloud servers help in sharing data amongst the people within an organisation or outside it. As security has become a very important aspect, providing security for the data that is being stored over the cloud has become a major challenge. With the proposed method, the data that would be stored would be encrypted which would ensure the data integrity. This data is then uploaded by the data owners. The data files are ranked according to the search by data users. The ranking helps in defining the newer files that are added and frequently searched.*

Keywords*: Cloud, Multi-keyword, Encryption, Ranked search.*

1. INTRODUCTION

With the growing demand of cloud, the basic day to day softwares are making its way up. As its popularity is on a rise, almost every organisation tries to store their data over the cloud. This has helped in the anywhere and anytime access concept of data retrieval. The same is done over Google, Amazon and many other websites. The major difference between these websites is the deployment of such software on a public cloud. When the above is done on a private cloud, the difficulty arises. As the cloud is private, it has to provide authentication of the Data owners, Data users, Administrators, etc. The use of this technology has spread to the many major and minor companies and organisations. As the data is saved over the internet, it can be misused, stolen or even used for receiving some amount of ransom. Many data over the internet are too sensitive and cannot be compromised at any cost. Therefore to overcome such a problem, encryption of the data can be done.

The data that gets uploaded over the internet has divided the set of users into two; namely; the data owners and the data users. The data owners are the group of users which upload their files or folders over the internet so that it can be used by the authorized people. The second group of the people are the data users. These are the group of users which use these set of data or information that is saved over the internet to analyze or summarize the data. Any third party that tries to access the data saved is termed as an intruder.

A. Multi-Keyword

When a file can be searched using more than one keyword, the system can be termed as a multi-keyword search system. The keywords can be also the name of the file or the various keywords which were pre-defined by the data owner for the data user to use and search the file.

B. Fuzzy Keyword

Fuzzy keyword is defined as the keywords which are misspelled. These keywords are changed or substituted so as to create keywords which act as possible defined keywords. This is discussed in detail in the following sections.

C. Ranked Search

When a user searches a file over the cloud, the files that are retrieved are quite bulky. To give better results, they can be further refined by ranking the documents using some parameter. The parameters can be set by the either the data owners or the users. This can be done by considering the following parameters; recently used documents, recently uploaded document, most popular documents, etc. The users can also define the number of entries of documents that they would like to see.

D. Cloud Computing

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The different essential characteristics are as follows:

- On-demand self-service.
- Broad network access.
- Resource pooling.
- Rapid elasticity.
- Measured service.

2. MOTIVATION

As most of the organisations have a lot of their data over the internet, they have expanded their storage over the cloud. This has made it difficult in providing the security to the data that is being stored. Hence to provide it security, the encryption of the data that is being uploaded has to be done. This will ensure the data security. This also ensures that no other authority knows about the data that is being uploaded.

3. LITERATURE SURVEY

Michael Armbrust *et al* [2] has explained about the cloud computing. The paper mainly talks about the advantage and disadvantages of using the cloud server for the storage of data. It explains how cloud is a greener and better method of storage service. It explains various concepts of cloud which would encourage people to store their data over cloud.

Dawn Xiaodong Song *et al* [3] discusses about the four of the most important concepts of provable security, query isolation, controlled searching and hidden query. The provable security is defined as the security provided to the data that can be proved. The query isolation is defined as the query which is sent to search the files that are stored. The controlled search helps in showing files to the users who have the authority to access them. The untrusted server does not learn anything from the queries that are being sent to the cloud server.

Cong Wang *et al* [4] discusses about the pitfalls of the traditional file retrieval methods. In the earlier systems, the file that would be retrieved would be in any order and hence this would create a problem. The other issue to this would be the downloading of the files. The number of files would be very high as all the files that would be related to the keyword will be downloaded at a time and this would increases the bandwidth consumption. The authors put forward an idea about ranking of the files and documents that are saved over the cloud. The user has to download all the files, decrypt and then check if the file is of any use to the user. This would help in retrieving the files which are recent or most downloaded. The paper also discusses the shortcomings of the Searchable Symmetric Encryption (SSE).

Jin Li *et al* [5] has discussed the fuzzy keyword searching technique. With this technique of keyword search, the other keywords can be searched as well. The fuzzy keyword substitutes, adds or deletes the alphabets in the keyword that is being searched over the system. As the above techniques are applied, a larger range of keywords can be searched and this would increase the range of the files that can be used. As this is done, better search results are presented.

Qin Liu *et al* [6] discusses about the ADL. The aggregation and distribution layer (ADL) is a middleware layer between the users and the cloud. It was envisioned such that an ADL will be deployed in an organization that has outsourced the data operations to a cloud. The ADL will aggregate queries from multiple users and send a combined query to the cloud.

Due to this combined query, the cloud will need to execute the query only once and return all matched files to the ADL. The authors have discussed over the Efficient Information retrieval for Ranked Query (EIRQ). The EIRQ helps in the retrieval of the files which are the top k specified. The user has to define the value of the k percentage of the files that are required by the user. This will help in lowering the overhead communication cost and hence the computations required will be less. This would save a round trip time and hence would give better and faster results.

The author Hongwei Li *et al* [7] describes the system that is similar to the one that is described by the author Wei Zhang *et al* [8]. In this model, the author works over a single data owner and multiple data users. The data owner has the authority of sharing files, authenticating the data user and uploading the files over the cloud. The major drawback of this system is the retrieval of the files. The files can be sent to the data users by the data owner only. This would lead to a lag in the system, as the data owner is not present all the time.

This would be a threat as well for the security of the leakage of key would create havoc. If the key is compromised, the entire system would come to a crash.

Wei Zhang *et al* [8] takes the above mentioned system and extends it to a multi data owner, multi data user system. In this system, there would be multiple authorized data owners which would be able to upload their data over the cloud. The keywords that the data user would have to send to the cloud are encrypted by the keys of the data owner. This would mean that when the data user would want their keywords encrypted, it has be done by either of the data owner. Since the data owners cannot be online every time, this has been proved as a disadvantage in this system.

The author Wei Zhang *et al* [1] explains the system which has multiple data owners as well as multiple data users. As the data can be sensitive and hence when it gets into the hands of wrong people, it can turn out to be harmful for both the owners and the receivers. Hence to overcome such a situation, encryption and decryption of data can be done for the safe exchange of any amount and type of data. The data owners and the data users need to be authenticated on the cloud servers before hand to use any of the services provided. Then the sensitive data is outsourced to the cloud, so as to enable the easier accessing of the data by the data owners and the data users, it is encrypted. The data encrypted has a list of keywords which are sent to an administration server. This in turn is then re-encrypted and uploaded by the administration server.

4. SYSTEM ARCHITECTURE

The system consists of four components, data owners, data users, administrator server and the cloud server. The data users

have to register themselves and hence this authentication helps in allowing only registered users to avail the service provided by the system. The data owners can upload their files over the cloud server. This would help the data users to search among different files and find any data that is relevant. The above method is better as the data is circulated and accessed by authentic users only. To ensure, the security, AES algorithm is used to encrypt the files. To decrypt them, AES algorithm is used. The index of keywords that is maintained is done using the MD5 algorithm. This is then re-encrypted using the AES algorithm.

The keywords that are sent to the administrator server are also encrypted. These are then re-encrypted and sent to the cloud server. Here, the files are fetched and the relevant files are sent to the data users. These are in the encrypted form. The main motive behind the encryption of the file is to maintain security of the file content. The files are securely saved over the cloud server.

To help the searching of various keywords, the fuzzy keyword algorithm is used. With the help of the algorithm, various related keywords can be searched at a single time. This would result in a wider range of keywords searched within a single query given by the end user.

5. SYSTEM IMPLEMENTATION

The system has the four components, namely, data user, data owner, cloud server and the administrator server. The login and the registration of these users should be provided for the smooth working of the system. The registration would then allow the user to enter the OTP which would be sent to the e-mail address that is provided during registration.

Fig. 5.1 Home page

The homepage would consist of the different options of login or registration.

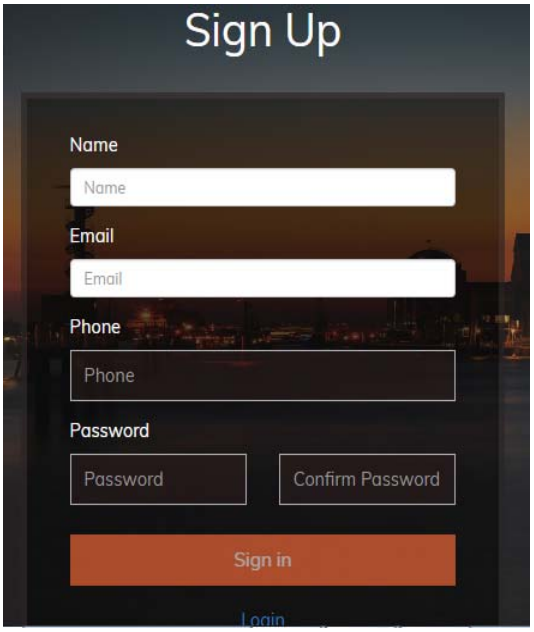

Fig. 5.2 Sign up page

After the signup page is filled, it would lead to the page that sends the OTP to the user that is trying to register itself.

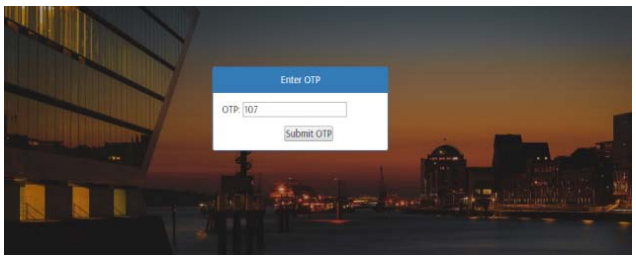

Fig. 5.3 One Time Password

Once the OTP is validated, the user can login using their ID and password. This would allow them to freely access files stored over the cloud server.

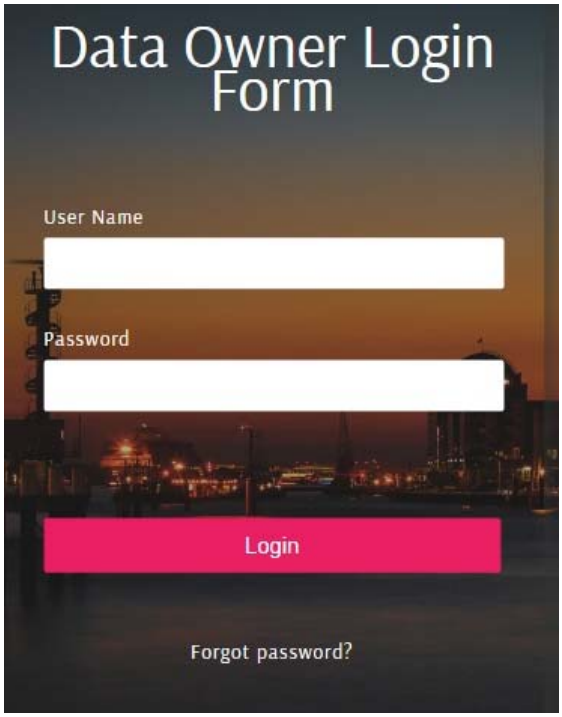

Fig. 5.4 Login page

6. CONCLUSION

As the data is being stored over cloud become very frequent, the searching of data has also become risky. To reduce such a risk, the encryption of the keywords has helped to keep the keywords under the wraps and hence made the search over cloud more secure. As the keywords are encrypted, the contents of the files are also hidden. Hence the cloud service provider will not be able to find any of the contents of the files.

REFERENCES

[1] W. Zhang, Y. Lin, S. Xiao, J. Wu, and S. Zhou, Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing, IEEE Transactions on computers, vol. 65, no. 5, May 2016.

[2] D. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data, in Proc. IEEE Int. Symp. Security Privacy, Nagoya, Japan, Jan. 2000, pp. 4455.

[3] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, Secure ranked keyword search over encrypted cloud data, in Proc. IEEE Distrib. Comput. Syst., Genoa, Italy, Jun. 2010, pp. 253262.

[4] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, Fuzzy keyword search over encrypted data in cloud computing, in Proc. IEEE INFOCOM, San Diego, CA, USA, Mar. 2010, pp. 15.

[5] 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, 2012, pp. 25812585.

[6] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, Dav id Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia, A view of cloud computing, Communications of the ACM, April 2010.

[7] H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou, and X. Shen, Enabling fine grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data, in IEEE Transaction on dependable and secure computing, vol 13, no. 3, May/June 2016.

[8] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, Secure ranked multi keyword search for multiple data owners in cloud computing, in Proc. 44th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw, Jun. 2014, pp. 276286.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, Privacy-preserving multi-keyword ranked search over encrypted cloud data, in Proc. IEEE INFOCOM, Shanghai, China, Apr. 2011, pp. 829837.

[10] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, Privacy-preserving multi keyword ranked search over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222233, Jan. 2014.

[11] W. Zhang, Y. Lin, S. Xiao, Q. Liu, and T. Zhou, Secure distributed keyword search in multiple clouds, in Proc. IEEE/ACM 22nd Int. Conf. Quality Service, Hong Kong, May 2014, pp. 370379.

[12] https://en.wikipedia.org/wiki/Cloud computing

[13] https://www.nist.gov/programs-projects/nist-cloud-computingprogramnccp

[14] https://en.wikipedia.org/wiki/Advanced Encryption Standard

[15] https://en.wikipedia.org/wiki/MD5

[16]http://www.cs.wustl.edu/jain/cse567006/ftp/encryption perf/index.html

[17] https://automationrhapsody.com/md5-sha-1-sha-256-sha-512-speedperformance/

[18] G.Singh and Supriya, A study of encryption algorithms (RSA, DES, 3DES and AES) for information security, International Journal of Computer Applications, Volume 67 No.19, April 2013.

[19] P. Gupta and S. Kumar, A comparative analysis of SHA and MD5 algorithm., International Journal of Computer Science and Information Technologies, Volume 5- No.3, 2014.

[20] A Survey on Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners over Encrypted Cloud Data, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 11, November 2017.