RELEVANCE SCORE BASED MULTI-KEYWORD SEARCH

P.Suganthi,
Associate Professor,
Department of Computer Science and Engineering,
K.L.N. College of Information Technology,
Pottapalayam,Sivagangai,Tamilnadu.

S.Dhivya,
UG Student,
Department of Computer Science and Engineering,
K.L.N. College of Information Technology,
Pottapalayam,Sivagangai,Tamilnadu.

M.Asrumathi,
UG Student,
Department of Computer Science and Engineering,
K.L.N. College of Information Technology,
Pottapalayam,Sivagangai,Tamilnadu.

E.Keerthiga
UG Student,
Department of Computer Science and Engineering,
K.L.N. College of Information Technology,
Pottapalayam,Sivagangai,Tamilnadu.

Abstract: These days the cloud storage has been used for many applications and the data owners upload their data to the cloud network so as to save storage and maintenance cost. Also they want the users of cloud to access their data in a secure way. The end users of the cloud can search and access the data and files from within the cloud using search keywords. The cloud should match the data with the searched keywords and provide with efficient search result. In most existing search methods only the single data owner model has been used. But the multi-owner model has not been efficient in many cases since each owner will use different private keys to encrypt their data and their keywords. So the cloud should handle this multiple keyword encryption when trying to make the search. In this project a multi-keyword rank based search model is proposed that used multiple data owners. The proposed approach also concentrates on the security of the data of the data owners and also tries to prevent unauthorized access during the search process. Initially an authentication mechanism is done to authorize valid users for the search. The end users provide their authentication data for this purpose and different key is used for each time. Finally after authentication the end user will generate a trapdoor that contains the list of all keywords for the search and sends that to the cloud. The cloud processes these keywords and identifies relevant data or files and ranks them based on a relevant score. Finally the result is displayed to the user with ranked search.

Keywords: Cloud computing, ranked keyword search, multiple owners, privacy preserving, dynamic secret key

I.INTRODUCTION

1.1 Cloud Computing

Cloud computing is a vast area and it has been used in many places and applications to improve their performance. Cloud Computing has been developed as a new computing environment for most of the fields and it provides easy, flexible and scalable on-demand services [1]. Anyone from anywhere can connect to the cloud and access these services and get their problems executed. The computing power of the cloud environment and the easy access makes the developers move their applications to the cloud for better performance [2]. The model of the cloud computing environment provides various services for the customers.

1.1.1 Various Cloud Services

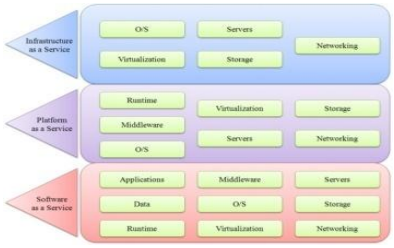

Figure. 1.Various cloud services

The various services that are provided are classified below in Fig. 1. By making use of these services, the workload of the clients is reduced by so much time and most of these are managed by the cloud providers or the cloud servers [3].

1.1.2 Storage Service in Cloud

The most important service provided by cloud environment is the use of storage space. Any user can avail the cloud storage to store their files and data in the cloud storage. This way the user can save their data from any unauthorized access. In some cases the data owner will upload their data to the cloud to provide multiple and easy access of these data to the respective clients of them. These clients or end users can access the uploaded data by making use of any search criteria. The data stored in the cloud are kept secure and access is only provided to those as described by the data owner. That kind of security policies is what makes the cloud so efficient and powerful. Apart from these, the cloud also provides other security related services for the data owners in case they upload the data. The owner can change his policies whenever possible and can also share the data with any other person as per his rights. Each data owner has his own key for accessing the data. In case the data is shared, then the user who tries to access the data is initially authenticated and only then the access is provided.

1.2 Keyword Search in Cloud

The keyword based search is the process of searching for various files or data in the cloud storage by using certain keywords as the input. In general the data owner will upload his data to the cloud and this will be shared to the various clients of the cloud. The cloud will encrypt and store these data so as to prevent them from being accessed by any attacker or someone with a malicious intention towards the data owner or the cloud. Also the authenticated user can provide any search keywords and the cloud should provide relevant and accurate search results for the same. These kinds of issues should be handled whenever we go for keyword based search in cloud.But further issues are seen when this is done using multiple keywords for a single search and when a single file or data has multiple keywords. One another issue is by using the multi-owner data model. In a single data owner model, the search made by the end user will be done on the data of only one owner. But in case of a multi-owner data model the search will be done in data of many owners. An efficient and accurate search mechanism should be used to make relevant search during these time.

1.3 Motivation of Work

The main focus here should be data privacy and authenticated access of data. The security of cloud computing should be efficient since it holds the key to its overall performance. Since the number of cloud users is large and since it is a large target for many attackers, an efficient security mechanism should be provided in case of multi-keyword based searching in cloud. For this purpose the users who try to do search should be authenticated in an effective manner before a search is made. The searching process should be so efficient in such a way that it would be almost impossible for any attacker to come through the authentication process. This is because in multi-user model the number of data is really large and there is a need to provide high security. Also since the data owners aim is to provide the valid and genuine end users with their data, an efficient search mechanism should also be provided that will provide better search results to the end users.

II. LITERATURE SURVEY

2.1 Encryption Algorithms

Many encryption algorithms are available for different types of need and in some cases new and hybrid algorithms are also used. In cryptography there are two types of encryption standards used and they are the public key cryptography and private key cryptography. Among this the public key encryption standard is the most used and this makes use of two keys one is the public key and the other is the private key. The public key is the one that is used to decrypt the data and the private key is the one that is used to encrypt the data. Here the security can be further improved by making use of the trapdoor function that makes sure that when someone knows the public key and the cipher text, he cannot be able to generate the original private key no matter what methods they employ.

The most commonly used public key algorithms are the Diffie-Hellman, RSA, ElGamal and DSA. The Diffie-Hellman is a system for exchanging cryptographic keys between active parties

2.2 Rank based Searching

Rank based searching is the process of providing an efficient searching mechanism that can identify the search intention of the user to provide efficient and relevant search results and also rank them based on higher relevance. During a search the number of items that match any given search criteria will be really high and the problem is to select only those items that will interest the user. For this purpose other parameters should be considered along with user given search criteria. Mostly the input provided by the user to make a search are one or more keywords and in other cases it could be a single word or sentence that can be split into many keywords. The rank based searching algorithm should consider all the aspects before making the respective search.To identify the user intentions and to provide accurate results other parameters are considered. Not much existing methods use such an efficient standard but some of the parameters or algorithm that can be used here are the similarity between the user and the searchable items based on his previous search history, the mutual information between the keywords used by the user and the files or data to be searched. In the proposed model in the project, the relevance score is taken and calculated between each keywords and the list of files or data to identify the most suitable search results. And finally these results are ranked based on the highest relevance scores.

III. EXISTING SYSTEM

3.1 Existing Method

The existing method as proposed by Deepali D.Rane et al. [ ] is the multi-user multi-keyword privacy preserving rank based search over encrypted cloud data. Here the data owners encrypt their data and then upload it to the cloud. Each data owner will have multiple keywords. The keywords are linked to their respective data by using the Lucene indexing mechanism. The data is encrypted using a symmetric key cryptography technique called as the TwoFish algorithm. Whenever a user search is made here, the cloud identifies the keywords in the search and then links them to the relevant data or files by using the Lucene index structure. Finally the top-k items that are retrieved are displayed to the use as the result. The data is decrypted using the TwoFish algorithm and then sent to the user. This approach makes use of multiple data owners and also the search is features by multiple keywords search.

3.1.1 Advantages

The encryption process is efficient and fast using TwoFish algorithm.Supports multi-owner multi-keyword model in rank based search.Search results as relevant to the keywords used.

3.1.2 LIMITATION

The authentication of user is not secure and results in unauthorised data access.The Lucene indexing mechanism is not accurate in case same keywords are used against

multiple files.The security of data owner's keywords is not considered.

3.2 Problem Description

3.2.1 Objectives

To enable cloud servers to perform secure and ranking based searching without knowing the keywords and trapdoors while also preserving the relevance between keywords and files.To prevent attackers from eavesdropping secret keys and pretending to be actual data users who perform searches.

3.2.2 Problem Statement

Preserving the privacy of the keywords, the trapdoor and the relevance between the keywords and the files during multi-keyword search performed by data owners and also to identify external attackers trying to act as data user for searching.

IV. PROPOSED SYSTEM

4.1 Proposed Method

The proposed system makes use of a privacy preserving rank based search model that uses relevance scores to rank the search results. This model uses an authentication approach to authenticate genuine users during search request. Any user who is not genuine is not allowed to enter the system for searching. A multi-owner model is used here where multiple data owners upload their files to the cloud. All these files are encrypted using their own data key. The list of keywords related to the files are also encrypted and sent to cloud. These keywords are again re-encrypted within cloud. For encryption purpose the RFC encryption mechanism is used that makes use of large hash values as encryption keys. This hash can be generated using the MD5 approach.During search, a user will authenticate him by providing the authentication data as given below,

Request Count	Last Request Time	Personal Identification	Random Number	CRC

A genuine user will provide correct data for request count, last request time and his personal ID. The Cyclic Redundancy Check (CRC) is used for checking data correctness and random number to confuse any attacker when a particular user does many searches. After successfully authentication the user will generate a trapdoor that contains his keywords and send to the cloud. The cloud processes these keywords and calculates relevance scores for each keyword with each file and top k files with high relevance are sent to the user.

4.2 Advantages over Existing Method

The security of data and the keywords of the users are enhanced by using double RFC encryption on keywords and RFC encryption on data.The search is made restricted for only genuine users by using authentication of user and trapdoor generation. The multi-keyword based ranked search is made accurate by making use of relevance score calculation using each keyword with each data or file.

4.3 System Architecture

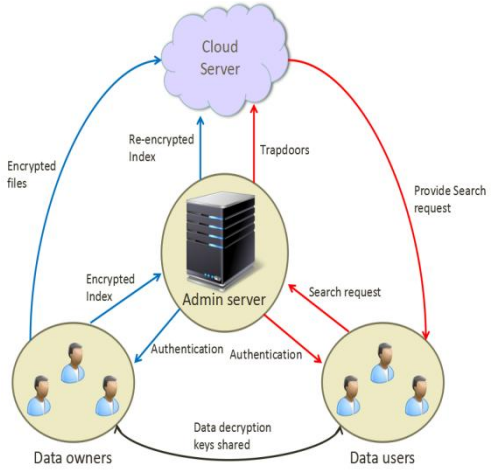

Figure.2. Overall multi-owner system model

The system model contains four entities, Data owner, Data users, Administration server and Cloud server. Data owners build a secure searchable index I on the keyword set W that is extracted from the collection of files F and them submits I to the administration server. After receiving I, the administration server re-encrypts it again and uploads the re-encrypted index to the cloud server. Then the owners encrypt their files F to obtain encrypted files C and upload them to the cloud server. Once data user makes a search using t keywords, he has to compute and submit the corresponding trapdoors to administration server. After authenticating the user, the trapdoors are then re-encrypted and uploaded to cloud server. After receiving trapdoor T, cloud server searches the encrypted index I of each data owner and returns the respective encrypted files to user.This system model is too complex to use normal search based methods like as in the existing system. Mechanism such as indexing will not be useful here since the number of keywords here is large and the user provided keywords is comparatively less. Also the number of data and data owners will also be more. So when using indexing, the number of search results will be really high and same keyword will provide multiple files or data as result. In such a way the relevant top-k files to be displayed to the user is hard to identify. To remove such issues the proposed method makes use of the relevance score calculation that is an efficient method for this model that uses multiple owners and multiple keywords.

4.4 Modules Identified

The proposed method has been implemented by identifying the following modules step by step.

4.4.1System Setup

Here the overall multi-user multi-keywords model is setup by creating multiple owners and providing them with set of files, keywords and their own private keys for encrypting the keywords and the data. Then the admin server is also setup by provide necessary keys for that and also providing genuine user's data and previous search history.

4.4.2.Data User Authentication

The data user authentication process is done by creating the authentication data and then sending it to the cloud admin.

The authentication process such as creating data, encrypting it and then generation of new keys are implemented here. The admin server identifies the data and authenticates the end users.

4.4.3Multi-keyword Search

Here the multi-keyword based search process is executed. Initially the authenticated user will generate the trapdoor that contains the list of keywords used for search. The cloud calculates the relevance score for all files with all keywords in trapdoor and returns the top-k items with the maximum relevance score. Here the multiple keywords of all data owners and their files are considered during search process.

4.4.4Simulation using Web App

The simulation of the proposed method is implemented and executed as a web application using the ASP.NET framework by taking the web browser as the user searching side and the SQL Server as the cloud server where all the data is stored as tables

V. CONCLUSION

In this project a multi-owner multi-keyword based searching model is proposed for cloud server that can search and rank the top-k items based on the relevance of the multiple keywords provided by the end user. The proposed searching is made accurate by using the relevance score between the files and the keywords that are used for searching. Compared to existing method that does not have a user authentication process, the proposed method makes use of an efficient and secure user authentication that not only avoids unauthorized users but also prevents them from entering the system by any illegal ways. The proposed model also handles the use of multiple data owners with multiple keywords from each data owner. The simulation is implemented as a web application that is created using the ASP.NET framework combined with the SQL Server. The results show that the authentication process is secure and the search results are more accurate in terms of relevance scores. In future this approach can be further enhanced by using a security mechanism that can avoid more attacks and can prevent the user data to be leaked.

VI. REFERENCE

[1]. Deepali D. Rane, V.R Ghorpade, "Multi-user multi-keyword privacy preserving rank based search over encrypted cloud data", IEEE International Conference on Pervasive Computing, 2015.

[2]. Bhushan Lal Sahu, Rajesh Tiwari; "A Comprehensive Study on Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2015

[3]. Abdelzahir Abdelmaboud, Dayang N.A. Jawawi, Imran Ghani, Abubakar Elsafi, Barbara Kitchenham; "Quality of Service Approaches in Cloud Computing: A Systematic Mapping Study", ELSEVIER, Journal of Systems and Software, Volume 101, March 2014.

[4]. W Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 11, pp. 3025–3035, 2014

[5]. Z. Xu, W. Kang, R. Li, K. Yow, and C. Xu, "Efficient multi keyword ranked query on encrypted data in the cloud," in Proc. IEEE Parallel and Distributed Systems (ICPADS'12), Singapore, Dec. 2012, pp. 244–251.

[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. IEEE INFOCOM'10, San Diego, CA, Mar. 2012, pp. 1–5.

[7]. M. Chuah and W. Hu, "Privacy-aware bed tree based solution for fuzzy multi-keyword search over encrypted data," in Proc. IEEE 31th International Conference on Distributed Computing Systems (ICDCS'11), Minneapolis, MN, Jun. 2012, pp. 383–392.

[8]. Z. Xu, W. Kang, R. Li, K. Yow, and C. Xu, "Efficient multi keyword ranked query on encrypted data in the cloud," in Proc. IEEE Parallel and Distributed Systems (ICPADS'12), Singapore, Dec. 2012, pp. 244–251.

[9]. 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'10, San Diego, CA, Mar. 2011, pp. 1–5.

[10]. M. Chuah and W. Hu, "Privacy-aware bed tree based solution for fuzzy multi-keyword search over encrypted data," in Proc. IEEE 31th International Conference on Distributed Computing Systems (ICDCS'11), Minneapolis, MN, Jun. 2010, pp. 383–392.