IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Multi-Keyword Search Methodology for Cloud Data

Mrs. Kavya B S[1], Shrilakshmi P V[2], Sushmitha N[2], Yamuna S[2].

Assistant Professor, Department of Computer Science and Engineering, NMIT, Bangalore, India[1]

Student, Department of Computer Science and Engineering, NMIT, Bangalore, India[2]

Abstract: Cloud computing provides a great platform to store data and provides many services on demand. This has encouraged many data owners around the world to store the data on cloud. To maintain the privacy and delicacy of the information stored on the cloud, encryption of the data must be done. Encryption is carried out before outsourcing the data. This makes it difficult to access data by the end users. Due to large number of information on the cloud and large number of users accessing it, there is a need for an efficient search scheme. This paper provides multi-keyword ranked search over the encrypted cloud data (MRSE).It is implemented using RSA algorithm through effective key generation. We implement "coordinate matching" as an efficient searching technique to synchronize pairing of query keywords. The efficiency of the searching is improved by tree-based index structure. The secure k-NN is used to encrypt index and query vector.

Keywords: Multi-keyword ranked search, coordinate matching, trapdoor, privacy preserving, secure k-NN.

I. INTRODUCTION

Cloud computing is a type of Internet based computing which provides many shared resources and data to computers and other devices. Cloud computing provides flexibility and it iscost efficient [1].Cloud computing is a shared pool of resources to access from or store data to a remote place. Cloud computing has a number of benefits like rapid elasticity, resource pooling , on-demand services, broad network access and pay as you use. Cloud computing does have some issues that has to be handled efficiently. The two main issues are security and privacy. Privacy is the obstruction that forbids the widespread adoption of the cloud by many end users. Even though sensitive data can be protected by using intrusion detection systems, firewalls and many other tools, privacy is still not fully attainable due to curious employees in cloud service providers[2].

With the advent of cloud computing data owners are encouraged to outsource their sensitive data from local sites to the cloud.To protect the data confidentiality, encryption of the data before outsourcing is the best approach. The existing techniques on keyword-based information retrieval cannot be applied directly on encrypted data. Downloading the data and decrypting it locally is highly impractical [3]. To meet the effective data retrieval need, the huge amounts of documents demand the cloud server to perform relevant ranking. Such rank based search system enables data users to find the most appropriate information faster. Rank search can also eliminate unwanted network traffic by sending back the most relevant information. To improve the accuracy of the search result as wellas to enhance the end user searching experience, it is also necessary for such ranking system to support multi-keyword search since single keyword search yields far too coarse results [4].

A secure tree based search scheme over the encrypted cloud data is proposed which supports multi-keyword ranked search over cloud data for end users.

The vector space model and term frequency (TF)* inverse document frequency (IDF) model are combined in the index construction and query generation to provide multi keyword ranked search. Based on this index tree, we construct a tree-based index structure and propose a "Greedy Depth-First Search" algorithm to obtain high search efficiency [4].To meet the challenge of supporting multi-keyword semantic without affecting privacy, we propose a basic idea for the MRSE using secure inner product computation which is adapted from a secure k-nearest neighbor(k-NN) technique [4].

Our contributions are summarized as follows:

1) We design the problem of MRSE and establish a set of strict privacy requirements for such a cloud data utilization system.
2) We use multi-keyword ranked search over encrypted cloud using co-ordinate matching to meet different privacy requirements with the help of threat models [4][5].

The remainder of this paper is organized as follows. In Section II, we discuss related work on both single and multi-keyword searchable encryption. In section III we describe proposed system in which goals, objectives and system

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

architecture are explained in detail. We have also explained two types of threat models, coordinate matching, trapdoor, k-NN algorithm and detailed description of RSA algorithm. Finally we conclude the paper in section IV.

II. RELATED WORK

From several years, searchable encryption which provides text search function based on encrypted data has been widely studied, especially in security definition, formalizations and improvement of efficiency.

1) Single keyword searchable encryption

The single keyword searchable encryption has a high searching cost due to the scanning of the whole data collection word by word. Secure index structure is explained by Goh and formulated a security model for index which is known as semantic security against adaptive chosen keyword attack. Due to lack of rank mechanism, users have to take a very long time to select what they want when large documents contain the query keyword. Boneh et al designed the first construction on searchable encryption, where anyone can use public key to access to the stored data on server but only authorized users having the private key can search. However, all the mentioned techniques only support single word search[6].

2) Multiple keyword searchable encryptions

To improve the search effectiveness for the users, a variety of conjunctive keyword search methods have been proposed. But these methods show large overhead, such as communication cost by sharing secret or computational cost by linear map. Pang and Mouratidis apply the Merkle hash tree based on authenticated structure to text search engines. But they only focus on the verification-specific issues ignoring the search privacy preserving capabilities that will be considered in this paper [6].

An encryption is an easiest way to secure the data privacy. It is extremely difficult to search documents over an encrypted data; therefore to improve the efficiency of searching many techniques are available. W. Sun, B. Wang, have given MTS scheme with similarity-based ranking. The authors build the search index based upon the term frequency and the vector space model to attain higher search result accuracy. To improve the search efficiency, a tree based index structure for multi-dimensional (MD) algorithms are proposed so that the search methodology is much better than that of linear search. In addition to this, two secure index schemes are considered to satisfy the stringent privacy requirements under strong threat models, i.e., known cipher text model and known background model. Zh ihua Xia, Xinhui Wang, present a secure multi-keyword ranked search scheme over encrypted cloud data. It also supports dynamic update operations of documents. Specifically, the widely-used TF*IDF model and the vector space model are combined within the index construction and query generation.

To surpass the disadvantages of sequential and binary search Pokom'y J proposed multi-dimensional Binary search tree for huge amounts of data. The multi-dimensional binary search tree is a data structure for storage of data to be retrieved by associative searches. The complete decryption of an RSA cipher text is unfeasible because no efficient rule presently exists for factoring large numbers. AES encryption is based combination of both substitution and permutation, and is fast in both software and hardware. Attribute-based encryption could be a kind of public key encryption in which secret key of a user and the cipher text are dependent upon attributes. In such a system, the decryption of a cipher text is feasible only if the set of attributes of the user key matches the attributes of the cipher text.The problem with the public key encryption with keyword search method is that every tag of all the files has to be processed in order to find the matches[7].

3) Boolean symmetric searchable encryption

The techniques mentioned earlier focused on single keyword matching but in practice users want to enter more than one keyword so as to get relevant data accurately. Tarik Moataz introduced a solution to handle challenges regarding search using multiple keywords over the encrypted cloud data. The conjunction search infers large overhead whereas disjunction search gives undifferentiated results.

4) Fuzzy Keyword Search

Traditional searching techniques for retrieving files only based upon exact keyword match, but the Fuzzy keyword search technique further extends by supporting common literal errors and format incompatibilities which would happen when the user types the keywords. Wild card based technique is used to build an efficient fuzzy keyword sets that are used for matching appropriate documents. The keyword sets are created using an Edit Distance algorithm which quantifies word similarity. The search result which is given back when search based on a fuzzy keyword data set is generated at every time the exact match search fails[2].

III. PROPOSED SYSTEM

A) Goal:

The key goal of the proposed system is to allow rank search for operational use of outsourced cloud data. Our system design instantaneously achieves performance and security assurances as given below:

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

- Multi-keyword Ranked Search: To design search scheme which allow multi-keyword query and provide ranking for effective data retrieval.
- Privacy-Preserving: To prevent the cloud server from knowing additional information from the dataset as well as index, and to meet privacy.
- Access control: Managing users' access on outsourced data using keys.
- Efficiency: Above goals on privacy and functionality should be achieved with low computation and communication overhead [5] [2].

B) Objective:

- The new scheme must be built in such a way that any authorized users can do a search on encrypted data on multiple keywords.
- The new scheme must facilitate users who can query the database provided that they possess so called trapdoors for the search terms that authorize the end users to include them in their queries.
- The new scheme must offer multiple keyword searches in a single query and ranks the results so the end user can retrieve the most relevant matches in an ordered manner.
- The new scheme must provide permissions to only authenticated owners to outsource the data to the cloud [5].

C) System Architecture

The problem that we consider is privacy preserving ranked search on private database model, where the documents are encrypted with the secret keys unknown to the holder of the cloud server. We consider three major roles:

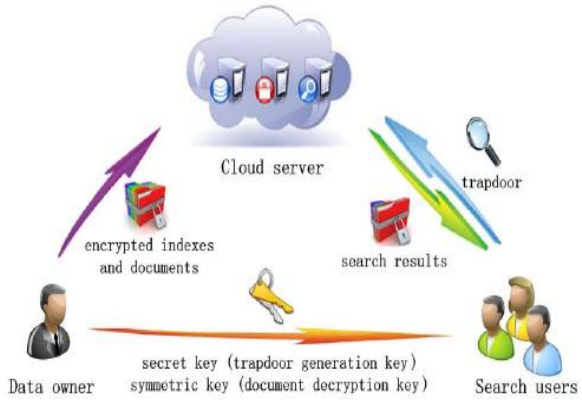

FIG.1 ARCHITETURE OF SEARCH OVER CLOUD DATA

- Data owner is the one who collects and/or generates information and outsource it to the cloud.
- Users are the members who are entitled to access the information from the cloud.
- Cloud Server, is a professional entity to offer information services to authorized users. It is often required that the server is oblivious to content of the database it maintains, the search terms in queries and documents retrieved.

The fig.1 shows the outline of the structure. The architecture involves three entities. The data owner is the one who generates and encrypt data, and upload them to cloud server. The owner can be either an organization or an individual. The cloud server belongs to a CSP which possesses significant storage and computation resources, and provides them to end user. The owner's data are encrypted end-to-end using secret keys. A searchable index is usually created and encrypted along with the outsourced data. To allow data access and search by end users, the data owner usually generates and distributes search tokens (or trapdoors), which are encrypted queries to users.

When a user wants to gain file access, he/she submits a corresponding token to the server, who then returns the matching set of documents. Finally, end user can download the decrypted files utilizing secret key provides by owner of the data [8] [9].

D) System Overview

1. Threat model

In this paper we explain two threat models.

Known cipher text model. In this model, Cloud server can get encrypted document collection, encrypted data index and encrypted query keywords. Known background model. In this model, Cloud server actually contains more information than that in known cipher text model. Statistical background information of dataset, such as term frequency information of a specific keyword and document frequency. These can be used by the cloud server to launch a statistical attack to infer or identify specific keyword in the query which further reveals the plain text content of the documents [6] [4].

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

2. Coordinate Matching

Coordinate matching is the process in which an intermediate similarity measure uses the number of query keywords appearing in the document to quantify the relevance of that document to the query. It is more elastic for the end users to identify a list of keywords indicating their concern and regain the most relevant documents with a rank order [5].

3. Trapdoor

The trapdoor generation function must be a randomized one instead of being deterministic. The generation of trapdoors incurs a vector splitting operation and two multiplication of a (m×m) matrix (m being the total number of keywords), thus the time complexity is $O(m^2)$. The cloud server should not be able to deduce the relationship of any given trapdoors, e.g., to know whether two trapdoors are created by the same end user search request. Otherwise the deterministic trapdoor generation would give advantage to the cloud server to accumulate frequencies of different search requests regarding different keyword(s), which may further violate the keyword privacy requirement [4].

4. K-NN algorithm

The kNearest Neighbors algorithm is a non-parametric method which is used for classification and regression. In both the cases, the input contains k closest training examples in the feature space. The result depends on whether k-NN is used for regression or classification. In k-NN regression, the result is a property value for the object. The value is taken as the average of the k nearest neighbors. In k-NN classification, the result is a class membership.

The object is classified based on the selection of majority of its neighbors. The object is assigned to the class which is most common among its k nearest neighbor. If k=1, then the object is assigned to the class of that single nearest neighbors. The k-NN algorithm is among the simplest machine learning algorithms. Both regression and classification is useful to assign weight to the contribution of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones [10][4].

5. RSA Algorithm

The RSA algorithm is used to encrypt and decrypt the file contents. It is used for both digital signatures and key exchange. It is an asymmetric algorithm which involves three steps: key generation, encryption and decryption. Key generation involves a public key and a private key where public key is used for encrypting data and known to everyone.

Private key is used to decrypt the encrypted data [11]. Following steps are used to create public and private key pair:

i. Choose two prime numbers, p and q. From these numbers we can calculate the modulus, n= pq
ii. Select the third number, e; such that it is relatively prime to the product (p-1) (q-1), the number e is the public exponent.
iii. Calculate an integer d from the quotient $\frac{(ed-1)}{(p-1)(q-1)}$. The number d is private exponent.
iv. The public key is the number pair (n,e).
v. To encrypt a message, M, with the public key, creates the cipher-text, C, using the equation
i. $C = M^e Mod\ n$
vi. The receiver then decrypts the cipher-text with the private key using the equation

$$M = C^d Mod\ n$$

Encryption algorithm:

Assume "A" as sender who sends a message to receiver "B". The sender will consider the following steps:

i. Obtain the public key (e, n) of the recipient B.
ii. The positive integer M represents the plain text message.
iii. Compute cipher-text$C = M^e Mod\ n$.
iv. The cipher-text C is send to B.

Decryption algorithm:

For the message sent by the sender "A", the receiver "B" will take the following steps:

i. The private key (n, d) is used to compute$M = C^e Mod\ n$.
ii. Plaintext is extracted from the integer representative M [12].

6. Modules

Our proposed system consists of following modules:

1) Data User Module
2) Data Owner Module
3) File Upload Module with Encryption
4) File Download Module with Decryption
5) Rank Search Module

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

1) Data User Module

In this module, the data users download files from cloud that are uploaded by data owners. The users must be able to do a multi keyword search on the cloud server. After specific search the users should be able to send a request to the respective data owners through trapdoor request for downloading files. If the request has been approved, the users should be able to download the decrypted file.

2) Data Owner Module

In this module the data owners must be able to upload the files and are also provided an option to enter the keywords for the files that are uploaded to the cloud server.

3) File Upload Module with Encryption

This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized users. The file should be encrypted with key so that the data users cannot just download it without this key. The end users request for the key through the Trapdoor.

4) File Download Module with Decryption

This module allows the user to download the file using a secret key. The data users will also be provided a request approval screen, where it will display if the data owner has accepted or rejected the request. If the request is approved, the users should be able to download the decrypted file with the help of a key.

5) Rank Search Module

This module allows the end users to search the files with multi-keyword rank searching. This module uses the frequently used rank searching algorithm. Co-ordinate matching principle will be adopted for the multi-keyword searching [5].

IV. RESULTS

A centralized location is provided where multiple users are created for data owners and data users. Once they log in either of the users can log into the system. The exchange of communication between data users and data owners is very secure. Public viewing is not possible as the contents are encrypted and sent to the cloud. The files can be viewed only by an authorized user with the consent of the owner.

- **Cloud Server:**

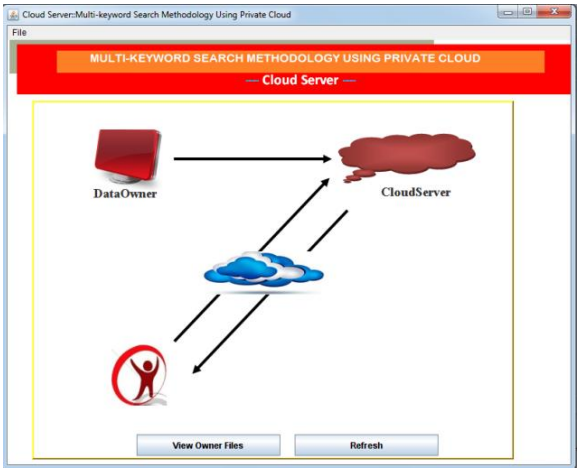

FIG.2 CLOUD SERVER PAGE

The above fig.2 shows a typical cloud server. It consists of data owner, Cloud and data user. Another important element shown in the above figure is the Trapdoor. It is mainly responsible for the security of the data present in the cloud. Through trapdoor only authorized users are allowed to access the files. Unauthorized users are immediately caught. The cloud server shows the effective movement of the data between the owner and the cloud, and between cloud and the owner.

- **Top-k Rank Search:**

The authorized users can download any file from the cloud. They can view the files according to a specific keyword. The files associated with that keyword will be displayed to the users. Users can also view the files based on the highest rank of the files. Rank of a file is incremented by each download of that file. Hence users can easily get the information about highest ranked files as shown in the fig.2.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

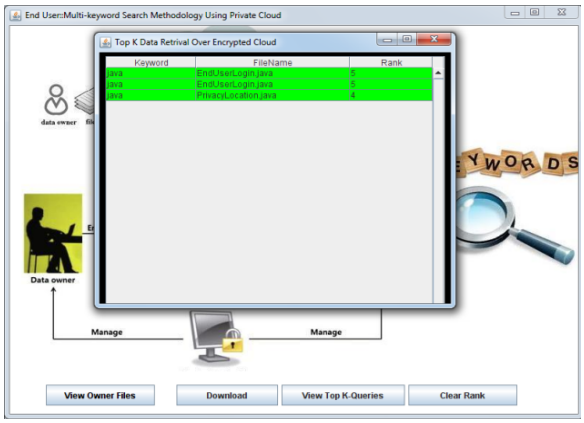

FIG.2 TOP-k RANK SEARCH

- **Data Owner:**

The fig.3 shows a file uploading process for an authorized data owner. They can upload file to the cloud server by browsing it from that particular system. To upload a file they must specify the filename, his/her valid name and enter a keyword for that particular file through which the users can view or download the file. Once the file is uploaded the secret key and public key is generated which is useful for the users to download the file. The file uploading is performed just by the owners. Deleting of a file is performed in the similar manner and it is also limited to the owner.

FIG.3 DATA OWNER

- **Cloud Attacker:**

End users are given the permission to download a file from the cloud server if he is an authorized user. Users must specify the appropriate file name and the secret key to download a file. If the end user specifies wrong filename or inappropriate secret key then the cloud server denies the user to download a file and displays a message “Hacker Found”. The details of the hacker are maintained in Cloud Attacker details where the invalid user name and the secret key are displayed as shown in the fig.4.

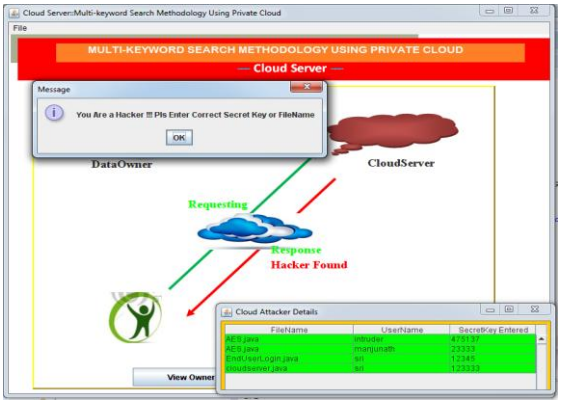

FIG.4 CLOUD ATTACKER

DOI10.17148/IJARCCE.2017.6422

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

V. CONCLUSION

In this paper, we motivate and solve the problem of multi-keyword ranked search on encrypted cloud data, and implement a variety of privacy requirements. Amongst the various multi-keyword concepts, we have used the efficient principle of coordinate matching. We established an effective ranking result using k-nearest neighbor algorithm to protect against two threat models. Access control for the different end users has been achieved. The integrity check for the data which is outsourced to cloud has been examined efficiently. We prove that our proposed method satisfies the security requirement. In our future work,we will explore the other multi-keyword semantics over encrypted cloud data and check for the integrity of the rank order.

ACKNOWLEDGEMENT

Dedicating the paper work to the esteemed guide, Asst. **Prof. Kavya B S,** whose guidance helped in We also thank **Dr. Thippeswamy M N**, Prof, and Head, Dept. of Computer Science and Engineering, NMIT, Bangalore-64 for his continuous support and encouragement and thanks to NMIT management.

REFERENCES

[1] Vaibhavi Kulkarni, Prof. Priya Pise, "Secure Multi-keyword Ranked Search over Encrypted Cloud Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 12, December 2016.
[2] Akshatha MS, Renita Tellis, "Cloud Data Encryption Using RSA, EnablingMulti-Keyword Ranked Search and AchievingPrivacy Requirements", International Journal of Advanced Research in Computer and Communication EngineeringVol. 5, Issue 5, May 2016.
[3] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, EEE, and Qian Wang, Member, IEEE, "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE transaction on parallel and distributed systems, 2015.
[4] Ning Cao, Cong Wang, Ming Li, Member, and Wenjing Lou, ―Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data‖ IEEE Transaction Parallel AND Distributed systems, vol. 25, no. 1, January 2014.
[5] Jyothi Koodi, G. Srinivasachar, "Privacy-Preserving Multi-keyword Ranked Search Over Encrypted Cloud Data",International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 03 | June-2015.
[6] Chi Chen, Xiaojie Zhu, Peisong ShenJiankun Hu, , Song Guo, Zahir Tari and Albert Y. Zomaya, "An Efficient Privacy-Preserving Ranked Keyword Search Method," IEEE transactions on parallel and distributed systems, VOL. 27, No. 4, April 2016
[7] Sonam.K.Darda , Prof. (Mrs). Manasi.K.Kulkarni, "Multiuser Multi-Keyword Ranked Search over Encrypted Cloud Using MHR andKP-ABE"International Journal of Computer Science Trends and Technology (I JCS T) – Volume 4 Issue 4, Jul - Aug 2016.
[8] G. Karthika Priya Dharshini, D.Viji, K.Saravanan,"Seclusion Search over Encrypted Data in Cloud Storage Services", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015.
[9] Jeniphar Francis, Ruchika Bansod, Chetna Getme, Priyanka Bagde, "A Secure and Encrypted Cloud Data with Multi-keyword Rank Search and Revocation of User", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 9, September 2016.
[10] Management Sinu V Agnes, Dr. K. Thamodaran,"Dynamic Multi-Keyword Ranked Search Based Security Mechanism for Cloud Data",International Journal of Advanced Research in Computer and Communication Engineering,Vol. 5, Issue 9, September 2016.
[11] Vanishree R, Mr.G.S Suresh, "Multi Keyword Ranked Search over Encrypted Cloud Data",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)Volume 4 Issue 5, May 2015.
[12] Nentawe Y. Goshwe, "Data Encryption and Decryption Using RSA Algorithm in a Network Environment",IJCSNS International Journal of Computer Science and Network Security, VOL.13 No.7, July 2013.