Achieving Privacy Preserved Content Retrieval with Assured Rank Integrity

.K. Santhosh / PG Scholar
Department of Computer Science & Engineering
SNS College of Technology
Coimbatore, India

Abstract: **Migration of users into the cloud environment is increased with the high popularity of services provided by the cloud providers. When large number users outsourcing their files into the cloud environment privacy becomes the most important issue. So the users outsource their data's after encryption. The searching and retrieval of files becomes most complex when the files are stored in the encrypted format. In the previous work Multi- keyword ranked search over encrypted data (MRSE) is implemented to assure the privacy enhanced searching mechanism. The ranking mechanism is used to retrieve the most similar values over the encrypted files. However one cannot assure that whether the all retrieved results are having most similar fields. The rank test mechanism can be implemented to find out the files are having similar fields or not. Our proposed system is used to retrieve the files with the most similarity values. It mainly concentrates on checking whether the retrieved queries are similar group of files or not.**

Keywords: Multi-Keyword Ranked Search Over Encrypted Data, Searchable Encryption.

I. INTRODUCTION

Cloud computing means providing services over the internet by software and hardware for the individuals and business people .Eg. Google, social networking. The basic concepts of cloud computing is deployment models and service models. Deployment models define the type of access to the cloud and there are four types of access: Public cloud , Private cloud, Hybrid cloud and Community cloud. The public cloud provides services to the commercial purposes. Public cloud provides free access to the providers. Private cloud is used for the special organization and it is maintained for the third party users. Community cloud is used for a group of organization. Hybrid cloud is used for combine all types of cloud .The data can be moved from one cloud to another. Service models are of three types: software as a service, platform as a service and infrastructure as a service. Infrastructure as a service is a basic services provided for the physical structure. It also provides physical storage and networking as a service. Platform as a service which provides the user to deploy the service. The consumer does not manage the operating system and network access. Software as a service is able to access the service or application. Eg. Salesforces.com .There are four types of cloud computing- technologies Virtualization, Service-Oriented Architecture (SOA), Grid Computing and Utility Computing. Virtualization means to allow a single physical instance of a application or resources. Service-Oriented Architecture allows to use the applications as a service for other applications. Grid computing allows a group of computers from multiple location which are connected to a common device. Utility computing is based on Pay per Use model means fully based on the cost or price and It offers computational resources on demand as a metered service. The multi-keyword ranked search is a method in which it explains about the inner product similarity and the coordinate matching. MRSE is a scheme in which allows two types model, first model is system model, and second model is thread model. MRSE is fully based on the coordinate matching and the inner product similarity. In the privacy preserving there are two types of methods to prevent, one of the method is data privacy. [1], [2] Data privacy is a method in which the data owner can restore the traditional symmetric cryptography in which the data is encrypted before outsourcing. The next method is the Index privacy, if any two cloud server has deduces the association with the keywords or encrypted from the cloud document then the content is used as searchable index.

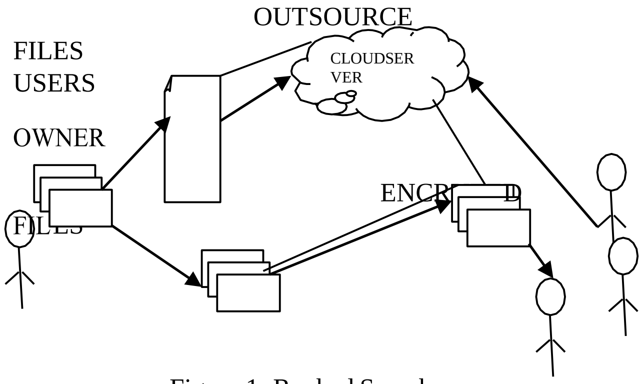

Figure 1: Ranked Search

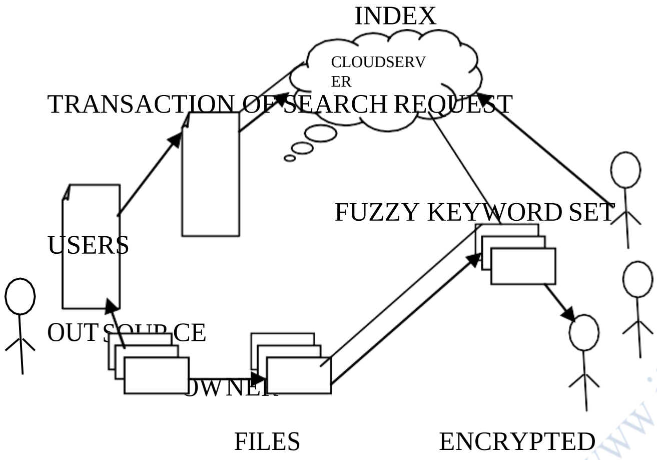

Figure 2: Fuzzy keyword search

The multi-keyword ranked search is a method in which it explains about the inner product similarity and the coordinate matching. Figure 1 explains about the ranked search method. The owner searches for the data using keyword while the data is outsourced and sent to the server. Then the user receives the necessary details about the data by the rank order [1]. Figure 2 explains about the fuzzy logic keyword search[6].

The owner searches for the data using keyword while the data is outsourced and sent to the server. Then the user receives the necessary details about the data by the pre-set edit order. A trapdoor function is generated in this fuzzy logic keyword search.h

A. *Searchable Encryption*

A Searchable encryption is a scheme in which is able to encrypt a set of search index, in which the contents are hidden except to the appropriate tokens. Each searchable encryption index is encrypted in a such a way that the token consist of a keyword in which one user can retrieve pointers to the encrypted files [1],[2].The Searchable encryption index cannot be used without tokens and the secret key function. There are two types of the Searchable encryptions, Symmetric Searchable encryption is a method in which the user can search the data and to generate the data. Asymmetric searchable encryption is a method in which the user can search the data from the different user and to generate the data from the other user.

B. *Multi-keyword Ranked Search over Encrypted data (MRSE)*

Multi-Keyword Ranked Search is method used to design a search for query required by the user which allow the multi-keyword to search for the particular details and provides results according to the need. In this multi- keyword search they have used coordinate matching and inner product similarity [1], The process of coordinate matching is used to find the matches as possible that are relevance to the data in the search query and the work of inner product similarity is to find the number of query keywords associated with the quantitative value in the document.

There are two types in the MRSE schemes .Scheme_1 tells about the cipher text model in which a new number is assigned to the query vector. Then a dummy keyword is introduced in the search and the trapdoor is generated. Then a random variable is created and the multiple keywords are encrypted for searching the data. Scheme _2 tells about the Background model effect is a best model compared to the cipher text model [1], [6]. Because when the data gets outsourced there is no privacy so to avoid this scale analysis attack is used in this Background model effect.

C. *Single Keyword Searchable*

The Single Keyword Searchable Encryption is one of the traditional methods in the searchable encryption method, and then the proposed system explains about how the data is hidden in the server and to search the contents from the data. Trapdoor is generated through secret key [1].It is a function in which an input is taken with a binary vector with a q-bits which represents about the true or false value and thus a trapdoor is generated. The system works with the concept of public and private key [2]. Any user with the public key can write the data to store on the server and the any authorized user with the public key can search the data. So with the public key it has a low level of privacy and not secure, thus a trapdoor is generated for a cipher text.

D. Boolean Keyword Searchable Encryption

The proposed system explains about the conjunctive keyword search over the encrypted data. The conjunctive keyword is a process in which the search returns about the all-or-nothing function which the search returns about the document in the query [1], [3] and [7]. Disjunctive keyword search does not return the accurate search results instead in returns about the subset of the specific keyword .The Boolean Keyword Searchable Encryption supports the Multi-Keyword Ranked Search Over Encrypted Data were the privacy is not secure in this encryption.

E. Public Key Encryption with Keyword Search

This Proposed system explains about the secure public key encryption in which the public key refers to the cipher texts are created by the various users. Consider there are two users A and B. Suppose if the user B is about to send an encrypted email to the user A with the keywords. Then the user B replies and sends a message [5], [1], user A is a public key. Our goal is that user A needs to send a secret key to the mail server to locate the keyword while it generates a trapdoor function with the private key. Then the server sends a reply message to the user A.

F. Private Information Retrieval

Private Information Retrieval explains about the secure message are kept from one to other user with a public key concept. The key is shared between two users only. When the user sends a message to the other user then that user is kept as storage provider and the user can collect, retrieve, search, delete the messages [4]. So the information is kept confidentially. This process is called as private information retrieval [1]. This concept is used for retrieving a plain or encrypted record of the database by address.

G. Predicate Privacy In Encryption System

Predicate encryption is a new encryption in which the user allows the fine grain control over access to encrypt data. The concept explains about when a user encrypts a message under a public key and the owner responds to that secret key by decrypting the cipher text. Predicate encryption allows symmetric-key setting and inner product queries [3]. In this tokens are associated with the predicates which can be evaluated over encrypted data. The predication is done with the public key in which the cipher text does not reveal information about the plain text.

H. Conjunctive Keyword Search

Conjunctive keyword search is done by the intersection to determine the correct set of documents or the user give the correct set of documents. The conjunctive keyword follows the decisional diffie Hellman assumption [4],[5]. The work of intersection is that for each set of documents a keyword is matched and the sets are found out by the user [3], [7].Server can also combine the information with the knowledge of user. This conjunctive keyword search partially solves the problem of the Boolean search keyword on encrypted data

II. PROPOSED METHODOLOGY

Fried man ranking test is implemented in this work to improve the search efficiency of the files. This mechanism is used to improve the search retrieval process by testing the rank fixed by the cloud servers. Whenever the user enters the multi key word for file retrieval, the trapdoor will be generated. Through that trap door value, the cloud server will match the user's keyword query with the searchable index of the encrypted files and will retrieve the most similar files. In this process, in order to assure retrieved files are similar to the queries and the result is retrieved by checking all the files in the group, rank test mechanism is used. The Fried man rank test mechanism is implemented in this work which can ensure that all files from the group is ranked properly and only the similar files are retrieved. This is done by ranking each block in file and comparing with the other files.

Fried man rank test is calculated as follows:

$$F_R = \frac{12}{rc\,(c+1)} \sum_{j=1}^{c} R_j^{\,2} - 3r(c+1)$$

Where, R_j^2 =Square of the total of the ranks for

group j (j=1,2…c)

r = number of blocks

c = number of groups

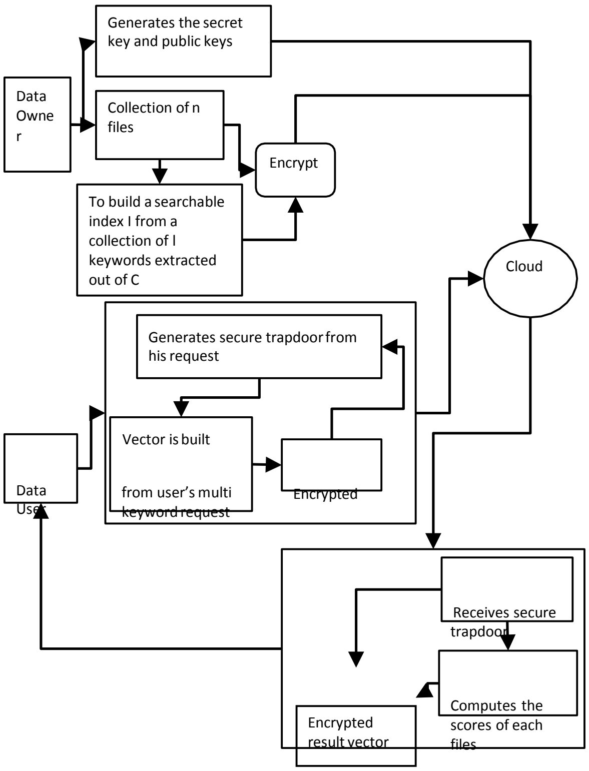
Generates the secret key and public keys
Data Owne r
Collection of n files
Encrypt
To build a searchable index I from a collection of l keywords extracted out of C
Cloud
Generates secure trapdoor from his request
Data User
Vector is built from user's multi keyword request
Encrypted
Receives secure trapdoor
Computes the scores of each files
Encrypted result vector

Figure 3: Architecture of MRSE Scheme

The Data owner first gets registered and logged into the process, then the same process is continued by the Data user. Then the file is encrypted and sent into the cloud (Figure:3) then the build index function is used to encrypt the files and sent into the cloud, while a trap door function is generated and thus the content is retrieved. The rank process is done, thus the searched keyword is received.

III. RELATED WORK

A. *Single Keyword Searchable Encryption*

In this Single Keyword Searchable Encryption system the Traditional single keyword searchable Encryption is used to hidden the content unless the server creates a trapdoor function through a secret keys [1]. So this model brings a well security than the other encryption standards. Then the related work explains about the secured rank keyword search method, which utilizes the rank results instead of different results [5]. But in public key encryption one user used to write a key to data in server while the authorized user can only accessed with the private key to search. So the privacy cannot be protected in this system and the key is encrypted easily.

B. Boolean Keyword Searchable Encryption

The Boolean keyword searchable encryption explains about the conjunctive keyword caused by the fundamental primitives [7] Eg: Communication caused by the sharing a secret key, Conjunctive keyword function describes about the all or nothing means it returns data were a keyword is specified by the search query, but the disjunctive search differs by the undifferentiated results. So the Boolean keyword searchable encryption does not support the multi keyword ranked searchable encrypted cloud data. Without the inner product similarity ranking test cannot be performed. Hence the proposed method solves the graph semantics.

IV.CONCLUSION

Thus the paper explains about all the types of keyword search used in the cloud data. It solves the problem of top-k retrieval data over encrypted cloud data and Privacy enhanced Rank test based query retrieval are used. In this mechanism the ranked queries that retrieved by the cloud server will be tested for similarity measure. It will output whether the files from the most similarity level that are retrieved. The privacy over mechanism is also provided by restricting the cloud server to learn the information from the data set.

V. REFERENCES

[1] Cao, Member, IEEE, Cong Wang, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data"Ning Member, IEEE, Ming Li, Member, IEEE, Kui Ren, Senior Member, IEEE, and Wenjing Lou, Senior Member, IEEE,vol 25, no.1, January 2014,

[2] S.Kamara and .Lauter,"Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[3] D. Boneh and B. Waters, "Conjunctive, Subset, and Range Queries on Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC), pp. 535-554, 2007 E. Shen, E. Shi, and B. Waters, "Predicate Privacy in Encryption Systems," Proc. Sixth Theory of Cryptography Conf. Theory of Cryptography (TCC), 2009.

[4] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[5] Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "PublicKey Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (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," Proc. IEEE INFOCOM, Mar. 2010.

[7] P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive KeywordSearch over Encrypted Data," Proc. Applied Cryptography and Network Security, pp. 31-45, 2004.