A NOVEL SECURE MULTI-KEYWORD RANKEDAND TREE BASED SEARCH SCHEME FOR MULTI DATA OWNERS IN CLOUDCOMPUTING

M.Swetha M.E(CSE), V.Abinaya M.E(CSE), PG Scholar
Mr.R.Sudhakar M.E., Assistant Professor(CSE)
Nandha College of Technology
swethanethra94@gmail.com, sudhakarcs87@gmail.com, abinaya101195@gmail.com

Abstract

Cloud computing has been considered an enterprise for IT infrastructure, which can organize huge resource of computing, storage and applications, and enable users to enjoy ubiquitous, convenient on-demand network access to a configurable computing resources with great efficiency and minimal economic overhead for shared pool. Attracted by these appealing features, both individuals and enterprises are motivated to contract out their data to the cloud, instead of purchasing software and hardware to manage the data themselves. So far, most of the works have been proposed under different threat models to achieve various search functions, such as single keyword search, similarity search, multi- keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi-keyword ranked search achieves more attention for its practical applicability. propose a secure and ranked multi-keyword search protocol in a multi-owner cloud model over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. A special tree-based index structure is constructed and efficient multi keyword ranked search is proposed. Due to the use of the special tree- based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. As a result, different data owners use different keys to encrypt their files and keywords. Authenticated data users can issue a query without knowing secret keys of these different data owners. Extensive experiments on real-world datasets confirm the efficacy and efficiency of our proposed schemes.

I.INTRODUCTION

1. General Background – Cloud Computing

Cloud computing, or something being in the cloud, is an expression used to describe a variety of different types of computing concepts that involve a large number of computers connected through a real-time communication network such as the Internet. In science, cloud computing is a synonym for distributed computing over a network and means the ability to run a program on many connected computers at the same time. The phrase is also more commonly used to refer to network-based services which appear to be provided by real server hardware, which in fact is served up by virtual hardware, simulated by software running on one or more real machines. Such virtual servers do not physically exist and can therefore be moved around and scaled up (or down) on the fly without affecting the end user-arguably, rather like a cloud.

2. Cloud Computing Security

As cloud computing is achieving increased popularity, concerns are being voiced about the security issues introduced through adoption of this new model. The effectiveness and efficiency of traditional protection mechanisms are being reconsidered as the characteristics of this innovative deployment model can differ widely from those of traditional architectures. An alternative perspective on the topic of cloud security is that this is but another, although quite broad, case of "applied security" and that similar security principles that apply in shared multi-user mainframe security models apply with cloud security. The relative security of cloud computing services is a contentious issue that may be delaying its adoption. Physical control of the Private Cloud equipment is more secure than having the equipment off site and under someone else's control. Physical control and the ability to visually inspect data links and access ports is required in order to ensure data links are not compromised. Issues barring the adoption of cloud computing are due in large part to the private and public sectors' unease surrounding the external management of security-based services. It is the very nature of cloud computing-based services, private or public, that promote external management of provided services. This delivers great incentive to cloud computing service providers to prioritize building and maintaining strong management of secure services. Security issues have been categorized into sensitive data access, data segregation, privacy, bug exploitation, recovery, accountability, malicious insiders, management console security, account control, and multi-tenancy issues. Solutions to various cloud security issues vary, from cryptography, particularly public key infrastructure (PKI), to use of multiple cloud providers, standardization of APIs, and improving virtual machine support and legal support.

II. EXISTING SYSTEM

Searchable encryption schemes enable the clients to store the encrypted data to the cloud and execute keyword search over cipher text domain. Due to different cryptography primitives, searchable encryption schemes can be constructed using public key based cryptography or symmetric key based cryptography. These early works are single keyword Boolean search schemes, which are very simple in terms of functionality. Afterward, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, and multi keyword Ranked search etc. A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi- keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection. These are significant works as it is highly possible that the data owners need to update their data on the cloud server. A secure multi keyword search method which utilized local sensitive hash (LSH) functions to cluster the similar documents. The LSH algorithm is suitable for similar search but cannot provide. In this scheme, different data owners use different secret keys to encrypt their documents and keywords while authorized data users can query without knowing keys of these different data owners. exact ranking.

Drawbacks

1. All these multi-keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality.
2. Some early Works have realized the ranked search using order- preserving techniques, but they are designed only for single keyword search.
3. Huge cost in terms of data usability. For example, the existing techniques on keyword based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.

III.PROPOSED SYSTEM

Present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF * IDF model are combined in the index construction and query generation. construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation beten encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.

Advantages

1. The proposed scheme should allow new data owners to enter this system without affecting other data owners or data users, i.e., the scheme should support data owner scalability.
2. The proposed scheme should allow multi-keyword search over encrypted files which would be encrypted with different keys for different data owners. It also needs to allow the cloud server to rank the search results among different data owners and return the top-k results.

IV.SYSTEM METHODOLOGY

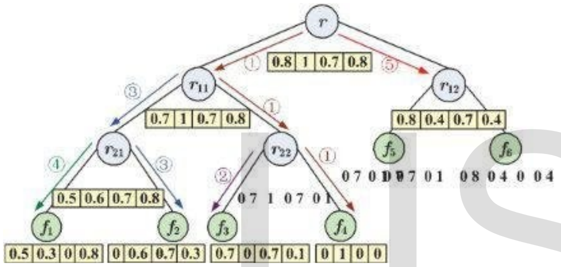
Despite of the various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers brings privacy concerns. The cloud service providers (CSPs) that keep the data for users may access users' sensitive information without authorization. A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Hover, this will cause a huge cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.

1. Dynamic Multi-Keyword Ranked Search

To enable secure, efficient, accurate and dynamic multi-keyword ranked search over outsourced encrypted cloud data under the above models, our system has the following design goals.

Dynamic - The proposed scheme is designed to provide not only multi- keyword query and accurate result ranking, but also dynamic update on document collections.

Search efficiency - The scheme aims to achieve sub-linear search efficiency by exploring a special tree-based index and an efficient search algorithm.



Tree based index structure

Algorithm for Build Index Tree

Input: the document collection F ¼ ff1 ; f2 ; ... ; fn g with the identifiers FID ¼ fFIDjFID ¼ 1; 2; ... ; ng.
Output: the index tree T

```
for each document fFID in F do
Construct a leaf node u for fFID , with u:ID ¼ GenIDðÞ, u:Pl ¼ u:Pr ¼ null, u:FID ¼ FID, and D½I ¼ TF fFID ;wi for i ¼ 1; ... ; m;—
Insert u to Current Node Set;
end for
while the number of nodes in
CurrentNodeSet is larger than do if the number of nodes in
for each pair of nodes u0 and u00 in CurrentNodeSet do
Generate a parent node u for u0 and u00 , with u:ID ¼ GenIDðÞ, u:Pl ¼ u0 , u:Pr ¼ u00 , u:FID ¼ 0 and D½i ¼ maxfu0 :D½i ; u00 :D½i g for each i ¼ 1; .. . ; m;
Insert u to TempNodeSet;
end for
else
```

2.Privacy-Preserving

The scheme is designed to prevent the cloud server from learning additional information about the document collection, the index tree, and the query. The specific privacy requirements are summarized as follows,

3. System Model

The system model in this project involves three different entities: data owner, data user and cloud server.

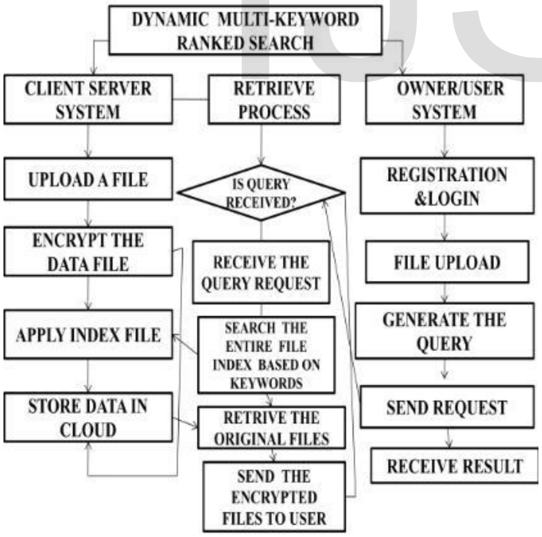
Data owner

Data owner has a collection of user wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. In our scheme, the data owner first builds a secure searchable tree index I from document collection F, and then generates an encrypted document collection C for F. Afterwards, the data owner outsources the encrypted collection C and the secure index I to the cloud server, and securely distributes the key information of trapdoor generation (including keyword IDF values) and document decryption to the authorized data users. Besides, the data owner is responsible for the update operation of his documents stored in the cloud server. While updating, the data owner generates the update information locally and sends it to the server.

Data users - Data users are authorized ones to access the documents of data owner. With t query keywords, the authorized user can generate a trapdoor TD according to search control mechanisms to fetch k encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key.

Cloud Server - Cloud server stores the encrypted document collection C and the encrypted searchable tree index I for data owner. Upon receiving the trapdoor TD from the data user, the cloud server executes search over the index tree I, and finally returns the corresponding collection of top-k ranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection C according to the received information.

V.SYSTEM ARCHITECTURE



1.User Revocation

User revocation is performed by the authentication server via a public available revocation list, based on which group user's can encrypt their data files and ensure the confidentiality against the revoked users. The authentication server compute the revocation parameters and make the result public available by migrating them into the cloud. Such a design can significantly reduce the computation overhead of users to encrypt files and the cipher text size.

2.Traceability

Anonymity guarantees that group users can access the cloud without revealing the real identity. Although anonymity represents an effective protection for user identity, it also poses a potential inside attack risk to the system. The authentication server should have the ability to reveal the real identities of data owners.

3.Data Owner/User Module

Owners/users are a set of registered users that will store their private data into the cloud server and share them with others in the group. The group membership is dynamically changed, due to new user's participation in the system. Data owners have a collection of files F. To enable efficient search operations on these files which will be encrypted, data owners first build a secure searchable index I on the keyword set W extracted from F, then they submit I to the administration server. Finally, data owners encrypt their files F and outsource the corresponding encrypted files C to the cloud server.

4.Cloud Storage Server Module

The cloud server provides data storage and search services to data owners and data users. After verify the user connection under signature, user can able to access the particular owner's data with respect to owner's private key which is reference of user identity (ID data). So the cloud verifies whether the request user is in the revoke list which is send by group manager under signature if so, it provide permission to access the data else throw unauthorized user request.

VI.CONCLUSION

For privacy concerns, secure searches over encrypted cloud data has motivated several research works under the single owner model. Hover, most cloud servers in practice do not just serve one owner; instead, they support multiple owners to share the benefits brought by cloud computing. In future work, focus schemes to deal with secure data sharing for dynamic groups in the cloud, expect to combine the group signature and dynamic broadcast encryption techniques. Specially, the group signature scheme enables users to anonymously use the cloud resources, and the dynamic broadcast encryption technique allows data owners to securely share their data files with others including new users. The authentication server compute the revocation parameters and make the result public available by migrating them into the cloud. Such a design can significantly reduce the computation overhead of users to encrypt files and the cipher text size. Specially, the computation overhead of users for encryption operations and the cipher text size is constant and independent of the revocation users. To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, will systematically construct a novel secure search protocol.

VII.REFERENCES

[1] Boneh D, Di Crescenzo G, Ostrovsky R, and Persiano G,(2004),„PublicKey Encryption with Keyword Search‟ , in Proc. Adv. Cryptol.-Eurocrypt,pp. 506–522.

[2] Boneh D, Kushilevitz E, Ostrovsky R, and Skeith W E III,(2007),„PublicKey Encryption that Allows pir Queries‟, in Proc. Adv. Cryptol, pp. 50–67.

[3] Cao N,Lou W, Li J,Wang C and Wang Q,(2010),„Fuzzy Keyword Search Over Encrypted Data in Cloud Computing‟, in IEEE Proc. INFOCOM, pp.1–5.

[4] Chang Y C and Mitzenmacher M,(2005),„Privacy Preserving Keyword Searches On Remote Encrypted Data‟, in Proc. 3rd Int. Conf. Appl. Cryptography Network Secur, pp. 442–455.

[5] Curtmola R, Garay J, Kamara S and OstrovskyR,(2006),„Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions‟ in Proc. 13th ACM Conf. Comput. Commun. Secur., pp. 79–88.

[6] Gentry C,(2009),„A Fully Homomorphic Encryption Scheme‟, Ph.D. dissertation, Stanford Univ., Stanford, CA, USA.

[7] Goh E J,(2003), „Secure Indexes‟, IACR Cryptol. ePrint Archive, vol. 2003,p. 216.

[8] Goldreich O and Ostrovsky R,(1996),„Software Protection and Simulation On obliviourams‟, J. ACM, vol. 43, no. 3, pp. 431–473.

[9] Golle P, Staddon J and Waters B,(2004),„Secure Conjunctive Keyword Searc h Over Encrypted Data‟, in Proc. Appl. Cryptography Netw. Secur.,2004, pp. 31– 45.

[10] Hou Y T , Lou Y, Wang B, Yu S,(2014),„Privacy-Preserving Multi-keyword Fuzzy Search Over Encrypted

Data in the Cloud‟ in Proc. IEEE INFOCOM,pp.2112–2120.

[11] Islam M S, Kuzu M, and Kantarcioglu M,(2012),„Efficient Similarity Search Over Encrypted Data‟, in Proc. IEEE 28th Int. Conf. Data Eng., pp.1156–1167.28

[12] Kamara S and Lauter K,(2010),„Cryptographic Cloud Storage‟, in Proc. Financ. Cryptography Data Secur, pp. 136–149.