Enabling Secure Ranked Multiple Keyword Search Over Outsourced Cloud Data

Pooja Ranjan

Abstract— As Cloud Computing becomes popular, sensitive information are being increasingly centralized into the cloud. To protect data and to keep privacy, sensitive data has to be encrypted before outsourcing. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. In this paper, we define and solve the problem of secure ranked keyword search over encrypted cloud data.To increase the security a secure and efficient keyword search can be done using ranked keyword index search scheme. System should provide security guarantee, ranked keyword search and efficiency. Data and index is stored in cloud server in an encrypted format. Whenever user wants to retrieve data submit request in secret format. The server search the data stored in database and returns the result according to the rank.

Index Terms— cloud computing, confidential data, cloud security, keyword search, keyword index search, searchable symmetric encryption

1 INTRODUCTION

Cloud computing is internet based computing where virtual shared devices and other resources and hosting to customers on a pay-as-u use basis. The cloud users can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. The main advantage outsourcing is it increases data availability and flexibility. The data owners upload their data and later they can retrieve from cloud. But the data is not secure since cloud server is an untrusted entity. The cloud server may leak data information to unauthorized entities or even be hacked. To avoid this sensitive data have to be stored in an encrypted format. The traditional keyword search algorithms works on unencrypted data. Thus, enabling an encrypted cloud data search service is of paramount importance.

Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. Therefore, how to enable a searchable symmetric encryption system with support of secure ranked search is the problem tackled here.

It has the following advantages:

1) Explore the problem of multikeyword ranked search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system.

2) It supports keyword search in encrypted form. The cloud server could determine which all documents contain the specified keyword without revealing anything about the contents of document or the keyword searched.

3) In this scheme, same keywords are encrypted to different cipher text for different documents thus reducing redundancy and avoiding the chance of statistical attack on keyword cipher text.

In general, keyword index search schemes consist of setup and searching processes. In the setup process, a user uploads encrypted data together with its indexes or searchable information on a database in cloud server, the index contains encrypted keywords for searching the data. To prevent cloud server from reading the keywords user generates trapdoor and sends it to the server. The trapdoors are nothing but encrypted keywords. When a user wants a document he sends the trapdoor to the server. Sever finds same value and returns. Receiver decrypts the document with his/her own decryption key.

This paper is structured as follows: First we review some related works in section 2. Then give some related definitions in section3. Section 4 contains the proposed method and analyzes its security and privacy in section 5. Finally we conclude this paper in section 6.

2. RELATED WORKS

It is an important research problem to enable the cloud service provider to efficiently search the keyword in encrypted form on encrypted files and providing user data privacy at the same time. Boneh et al introduced a public key encryption with keyword search (PEKS)[1] which supports encrypted keyword search. Here the document is encrypted with any public key encryption algorithm and the user needs to decrypt it completely by himself. It will depletes too much CPU and memory power of the client if documents are decrypted frequently and loose the critical value of cloud computing. Qin liu introduced an efficient privacy preserving keyword search[2] which allows the service provider to participate in partial decipherment of the searched

documents thus reducing the computational overhead of the user. But this scheme supports only single user. That is the user searching for data is same as the user who store data in cloud. Y.H.H et al proposed the scheme of public key encryption with conjunctive keyword search[3] and extended it to a multi-user system. In this scheme, the keyword encryption uses public key of all share users. The trapdoor is made as the keyword query with the partial of every computed public key. When the number of users increases, it has low efficiency and it is not so efficient for the actual application of cloud. Changy u Dong et al proposed a scheme shared and searchable encrypted data for untrusted servers[4]. It supports multiple users. In this scheme encryption is not based on public key. In this scheme service provider performs partial decryption. But it can't resist the statistical attack on keywords. Here same keywords are encrypted to same cipher texts only for different documents. Wang and Cong proposed a scheme order preserving symmetric encryption scheme[6]. It supports multiple users and single keyword search. But it doesn't guarantee access privacy. Wang and Cong also proposed another scheme which uses searchable symmetric encryption along with order preserving symmetric encryption[7]. It provides security guarantee compared to previous SSE schemes. However it supports only single keyword search.

3. Related Definitions

Definition 1. One-Way Hash Key Chain

Many protocols need to commit to a sequence of random values. For this purpose, repeatedly use a one-way hash function to generate a one-way chain. One-way chains are a widely-used cryptographic primitive. It is generated by selecting the last value at random and apply function h. anybody can deduce earlier value i from the chain using later value j. but reverse process is not possible i.e. with earlier value ki we cannot predict later value kj.

Definition 2. PRF

Pseudo Random Function (PRF) defined over (K,X,Y):

$F: K \times X \rightarrow Y$

such that exists "efficient" algorithm to evaluate F(k,x).

Let $F: K \times X \rightarrow Y$ be a PRF

Funs[X,Y]: the set of all functions from X to Y

$SF = \{ F(k, \cdot) \text{ s.t. } k \in K \} \subseteq Funs[X,Y]$

A PRF is secure if a random function in Funs[X,Y] is indistinguishable from a random function in SF.

4. Proposed Method

4.1. System Model

The system consist of four different entities: data owner, data user, key server and the cloud server

Data Users : Data users are authorized persons who can retrieve the file stored in the cloud servers. They search for the files using trapdoor for keywords and can retrieve the file and decrypt it using the secret key.

Key Server: Key server is a trusted server which manages the group session keys and the search keys of all groups,for secure communication and secure keyword index search.

Cloud Server : Cloud server is an untrusted server which provides the storage service. The data owners store their documents along with encrypted keywords in the cloud storage server in encrypted form. Cloud server does not get any information about the document or the keywords.

4.2. Algorithm

- KeyGen(l). Taking l as an input, this algorithm generates users' group session key set {gk}, index generation key set {ik}, and file encryption key set {fk}.
- BuildIndex(ik,W). Inputs of this algorithm are an index generation key ik and a keyword set W. Output is index list table.
- FileEnc(fk, D). Given a file encryption key fk and a document D, this algorithm outputs an encrypted document.
- TrapdoorGen(w, ik). This algorithm takes a keyword w and index generation key ik. It encrypts the keyword w with index generation key ik and returns the encryption value, which is the trapdoor Tw for the keyword w.
- Retrieval(Tw,k). When the cloud server receives request with trapdoor and the encrypted keyword, it performs the ranked search on the index I with the help of trapdoor Tw, and finally returns the ranked id list of top-k encrypted documents sorted by their similarity with keyword.

• Decrypt(E(D), fk). Given a file encryption key fk and encrypted document E(D), it outputs a plaintext document D.

4.3. Working Process

This scheme largely comprises of two parts;
(1) Setup phase
(2) Retrieval phase

The setup phase consists of three algorithms of KeyGen; BuildIndex; FileEnc. The retreival phase is composed of three algorithms of TrapdoorGen; Retrieval; Decrypt.

4.3.1 Setup phase

4.3.1.1 KeyGen(l)

In this construction, group search keys are generated. With system parameter l, GM generates group session keys {*gk*}, index generation keys {*ik*}, and document encryption keys {*fk*}, where index generation keys and file encryption keys are called as search keys. The search keys are reversely generated by one-way hash key chains. At first, the last key of a key chain is selected. The key server applies the last key to a hash function repeatedly and computes all other keys until the first key comes out.

The property of One-way hash function can prohibit a leaving member from computing new keys after leaving the group. But any newly joining member can obtain all previous keys through applying the current key to hash function h repeatedly. This eliminates decryption and re-encryption of the previous documents. These search keys are distributed to all of the group members every membership change.

4.3.1.2 BuildIndex() and FileEncrypt()

Before uploading data to the cloud server, the data owner encrypts the documents and keyword. He encrypts the files using the algorithm FileEncrypt with file encryption key. Using the index key the keywords are encrypted and stored in index table. Then user sends this to key server.

4.3.1.3 Uploading to cloud server

The key server re-encrypts the received documents with his key. Then it uploads whole documents to the database in cloud server. The database consists of two tables each for storing index and document. By using this tables server can easily process the queries comes to it.

4.3.2 Retrieval phase

4.3.2.1 Trapdoor(w, ik)

The user generates and sends a Trapdoor(w,ik) to the key sever for an interested keyword w. Upon receiving the trapdoor Tw, the server reencrypts the keyword and sends to cloud server.

4.3.2.2 Retrieval(Tw) and Decrypt(E(D), dk)

Upon receiving trapdoor, the cloud server search in the database for the keyword. If match is found it sends the encrypted document to key server. The server decrypts the document and calculates the rank based on the similarity and sends it to the user.

5. Privacy and Security Analysis

In this scheme security is ensured by setting group key, index key and document key. Group Key is provided to the members of that group only. Any member who leaves from the group can't access the files. Also members of a group cannot predict key of another group. A new member to a group can also access previous documents of that group. The cloud server cannot predict the content of the encrypted files. It cannot find anything related to the document from the keywords because the keywords itself are encrypted. Suppose cloud server somehow find the search keys then also it cannot decrypt the data because it doesn't know the secret key of key server.

The scheme guarantees data privacy, index privacy, search privacy, keyword privacy and trapdoor unlinkability. The data owner encrypts the data before uploading to cloud server. Hence it assures data privacy. The index contains the encrypted keywords and encrypted identifiers of the document. To hide what they are searching the users send request as trapdoors. Hence keyword privacy can be achieved. Unlinkability means that when services and resources are used by someone, the others cannot link these being correlated or used together. Trapdoor unlinkability means no one can predict the document's content from the trapdoor. In particular, the cloud server should not be able to deduce the relationship of any given trapdoors, e.g., to determine whether the two trapdoors are formed by the same search request.

6. Conclusion and Future Works

Cloud computing is one of the current most important and promising technologies. A data owner can store his data in cloud and could retrieve whenever needed. User can send query to cloud server for retrieval of data. For this purpose a secure scheme is needed. In this paper secure system for keyword search is proposed. It enables the user to retrieve documents without any leakage of data. This scheme enables multiple people to access the database. Only authorized users are allowed to access database. The encryption and decryption is partially done by data owners. To give more security encryption at a trusted server is also done. When user leaves from a group his permissions to access database is also revoked. For future works, a new scheme could be developed to avoid the encryption and decryption overhead for the users.

References

[1] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano ,Public Key Encrption with Keyword Search In proceedings of Eurocrypt 2004, LNCS 3027, pp. 506-522, 2004

[2] Qin Liu,Guojun Wang ,Jie Wu Proceeding CSE '09 —An Efficient Privacy Preserve Keyword Search Scheme in Cloud Computing,Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 02 pp 715-720

[3] Yong Ho Hwang and Pil Joong Lee, —Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-user System,Lecture Notes in Computer Science, 2007, Volume 4575/2007, 2-22

[4] Changyu Dong, Giovanni Russello and Naranker Dulay ,Sharable and Searchable Encrypted Data for Untrusted Servers Proceeedings of the 22nd annual IFIP WG 11.3 working conference on Data and Applications Security pp 127 – 143

[5] Liu Hong-xia, —Research on privacy preserving keyword search in cloud storage ,Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on 9-11 July 2010 ,pp 444 – 446

[6]Wang,Cong, Secure keyword Search over encrypted cloud Data,Distributed computing Systems(ICDCS),2010 IEEE 30th International Conference on 25-june 2010

[7] Wang,Cong, Enabling Secure and Efficient Ranked keyword Search over outsourced cloud Data,Distributed computing Systems(ICDCS),2010 IEEE 30th International Conference on august 20122