Keyword-based Ciphertext Search Algorithm under Cloud Storage

Ren Xunyi[1,a] ,Yan Shiyang[1]

[1] College of Computer Nanjing University of Posts and Telecommunications, 210003 Nan Jing, China

Abstract. With the development of network storage services,cloud storage have the advantage of high scalability,inexpensive,without access limit and easy to manage. These advantages make more and more small or medium enterprises choose to outsource large quantities of data to a third party. This way can make lots of small and medium enterprises get rid of costs of construction and maintenance,so it has broad market prospects. But now lots of cloud storage service providers can not protect data security.This result leakage of user data ,so many users have to use traditional storage method.This has become one of the important factors that hinder the development of cloud storage. In this article,establishing keyword index by extracting keywords from ciphertext data. After that,encrypted data and the encrypted index upload cloud server together.User get related ciphertext by searching encrypted index,so it can response data leakage problem.

1 Introduction

Cloud storage has the advantages of cost-effective,high scalability,inexpensive,without access limit and easy to manage.These advantages can enable small businesses spend less cost getting rid of the high cost of construction of the storage system, which greatly reduce their storage costs.Therefore, cloud storage has a high market prospect.The behavior of confidential data is saved to third parties caused the company's great worry,which limit the development of cloud storage.Many enterprises have to adopt the traditional storage model.This has become an important factor of restricting the development of cloud storage.

Under these conditions,encrypted storage produced.In the conditions of large amount of ciphertext data stored in server-side ,users need to put all of the ciphertext data downloaded to the local and retrieve after decrypting data. Although this strategy protects the privacy of data, but users need to waste a lot of resources, consume large amounts of network bandwidth, which also poses a serious obstacle to the development of cloud storage.At the same time, a huge cloud server computing capacity were idle, and became a simple data pool.In order to exert cloud server computing power and huge storage resources, keyword-based search technologies emerged.[1]After encrypting ,client upload documents and inverted index to the cloud server.When searching, users only need to enter a keyword.Cloud Server can search for relevant ciphertext file and then return to the client.Local users can use the key to decrypt the ciphertext file.This program not only protect the security of data,but also use computing power and large storage resources.

2 The Method of Ciphertext Search

First,The user selects a file and extracts the keywords in the file.Then, the keyword and the document identification combined to form an inverted index.Files use symmetric key encryption algorithm to encrypt and inverted indices also use a pseudo-random function to encrypt. Ciphertext and the encrypted indices upload to the server with storage.When the user queries, client hosts use the search algorithm to process the keyword to a query pointer and submit to the server.The server uses the pointer to call the user's encrypted search index.Server use the user's encrypted index to find the corresponding ciphertext and returns to the client, the client can use the key to decrypt the ciphertext.

With this idea, the cloud server will not get any useful information about files.Server just know the ciphertext, encrypted index, and search pointer and can not get any effective plaintext and keyword information.Therefore, using this search method will give the ability of transparently querying to cloud server .

3 Implementation Model of Ciphertext Search Method

This article discusses the method involving two entities.The one is the owner of the confidential data.They want the data to be stored in the cloud server, they need to protect the confidentiality of data, such

[a] Corresponding author: renxy@njupt.edu.cn

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

entities are called users.The other is a provider of cloud storage services.They offer storage interface for users, and perform some operations of retrieving on the data , such entities are called servers.

From the foregoing, in order to protect the confidentiality of data, the cloud server can not get any useful information throughout the search process.All data processing on locally, including file encryption, encryption of the index,and the process of generating search pointer by keywords.

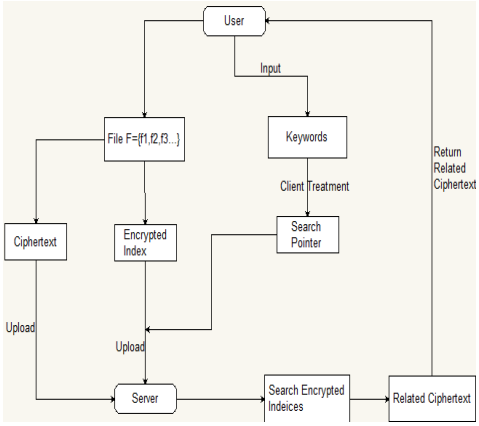

Figure 1. Search model under the cloud

From the figure above,the users select a file to be stored with the client and pretreat the file.Then,users upload this treated file.Pretreatment is divided into two portions.The one is to use a symmetric key encryption algorithm to encrypt the plaintext.The other is to encrypt indices composed by a set of keywords and get encrypted index. After that,users upload encrypted files and encrypted indices to server and store.It is not necessary for users to care store details. When the user queries, client hosts use the search algorithm to process the keyword to a query pointer and submit to the server.The server uses the pointer to call the user's encrypted search index.Server use the user's encrypted index to find the corresponding ciphertext and returns to the client, the client can use the key to decrypt the ciphertext.

The key to the searchable encryption algorithm is to encrypt file index.This programme use pretreatment method.First,we extract a few keywords from each file you want to upload, and then use the symmetric key encryption algorithm to encrypt the plaintext.Each keyword contain files generate list called keyword list.The node structure of linked list include file identification, the location of the previous node and the location of last node.All keyword lists consist of index called inverted index[3].Using pseudo-random function encrypt inverted index and get encrypted index.This method make server can not get effective information.Then,these encrypted indices are stored in random locations on the array(A_s).The first node for each keyword list is stored uniformly in the dictionary called(T_s).Encryption array and dictionary are stored on the server and the server can not get effective information.When the user queries, client hosts use the search algorithm to process the keyword to a query pointer and submit to the server.The server uses the pointer to call the user's encrypted search index.Server use the user's encrypted index to find the corresponding ciphertext and returns to the client.By these steps can complete the search process[4].

4 The Design of Algorithm

In this article,we use Random oracle to generate search index and use Pseudo-random function to generate encrypted index. We choose $\{0,1\}^m \to \{0,1\}^m$ as Random oracle,and use H_1、H_2 to express respectively.We choose $\{0,1\}^n \times \{0,1\}^s \to \{0,1\}^s$ as Pseudo-random function ,and use F、G、P to express.Using η to express search pointer.In this program,we design Symmetric Searchable Encryption(SSE).[2]Algorithm provides file encryption, key generation, encryption index,pointer search, file decryption and other functions.

This algorithm can be expressed as $SSE = \{Enc, Gen, IndEnc, IndGen, Search, Dec\}$. Enc represent the algorithm of encrypting files.Client operate the algorithm to encrypt plaintext file. Gen represent the algorithm of generating key,and generate the key of symmetric key encryption algorithm and pseudo-random function. $IndEnc$ represent the algorithm of encrypting keyword and generate encrypted index.Client perform this algorithm to generate encrypted index and upload encrypted index and ciphertext to server. $IndGen$ represent the algorithm of generating search pointer.Client perform this algorithm to transform keyword to search pointer,and upload to server. $Search$ represent the algorithm of searching encrypted indices. This algorithm run in server,and server can find related ciphertext when client submit search pointer. Dec represent the algorithm of decryption.Client decrypt ciphertext return from server by using key.

4.1 The Generation of Key

The process of generating key: $IndEnc$:The algorithm of encrypting keyword,three randomly selected m-bit key length string ,they are k_1, k_2, k_3 as the key of pseudo-random function. $Gen(s)$:Given a system security parameter named s as input parameter and generating $K_4 = Gen(s)$ as key of symmetric encryption algorithm.Then,outputting $K = \{K_1, K_2, K_3, K_4\}$ .The key generating by this algorithm is stored in client,and encrypt files and keywords.

4.2 The Encryption of Files and Indices

Files encryption are to encrypt the plaintext document by using symmetric key encryption algorithm.[6]

The process of index encryption: $Enc(K,F,\lambda):K$ is key, $F=\{f_1,f_2,f_3.....\}$ is files collection , λ is inverted index, A_s is the array that store encrypted index, T_s is the dictionary that store first node of keyword list, V is the collection of keywords, $\$V$ is the number of keywords, L_f is the list consist of all keywords by file f , $\$L_f$ is the number of keywords in file f , L_v is The file list consist of keyword v , $\$f_{v_i}$ is the number of files including v .

(1) Initialize array A_s and dictionary T_s .

(2)Perform following operations for each $v_i \in V$:(a) Create linked list named L_{v_i} , the node of linked list are $(N_{i,1},N_{i,2}....N_{i,\$f_i})$,including $\$f_{v_i}$ nodes.These nodes are stored in random position of array A_s .The structure of node is $N_{i,j}:=(<loc_s(N_{i-1,j}),loc_s(N_{i+1,j}),id_i>\oplus H_1(P_{K_1}(v)))$. $loc_s(N_{i-1,j})$ represent the front location of $N_{i,j}$, $loc_s(N_{i+1,j})$ represent the back location of $N_{i,j}$, id_i represent file identification of node.(b) The first node of every linked list are stored in T_s .The structure of node is $T_s:=<loc_s(N_{i,1})>\oplus G_{k_2}(v)$, $loc_s(N_{i,1})$ represent location of first node of every linked list.

(3) These free nodes of A_s and T_s are filled with random strings.

(4) Each f_i is encrypted and generate $c_i=Enc(K,f_i)$.

(5) The algorithm output $\mu:=(A_s,T_s)$ and $c=\{c_1,c_2,.....c_{\$F}\}$.

Encrypted files and encrypted indices are uploaded to cloud server.Specific storage operations are done by cloud server, the client is not involved in this work.[7]

4.3 Implement of Search

When user want to search files,they upload keyword and client transform keyword to η by using $IndEnc$ algorithm .Server use search index to read encrypted index. Then,server use $Search()$ algorithm to search encrypted index,and get related ciphertext collection.These ciphertext all return to client,and client decrypt these ciphertext to plaintext.[8-10]This completes the whole search process.Specific search process is as follows:

(1) $IndEnc(K,v)$: Input v and K ,output $\eta:=\{F_{k_1}(v),G_{k_2}(v),P_{k_3}(v)\}$

(2) $Search(c,\eta,\mu)$:Input μ , η ,after following these steps:(a)Find the location of v_i in T_s ,this location name α_i .(b) Using α_i find the coordinate of keyword in A_s and find N_i ,then find id_i and find related ciphertext.(c)Using $I_{v_i}=\{id_1,id_2,.....id_m\}$ represent searched file mark collection.Then,finding every mark's corresponding ciphertext and outputting $\{c_i\}$.

(3) $Dec(K,c)$:Input K and $c=\{c_1,c_2,.....c_m\}$,using $f_i=SSE.Dec(K,c_i)$ to get plaintext.

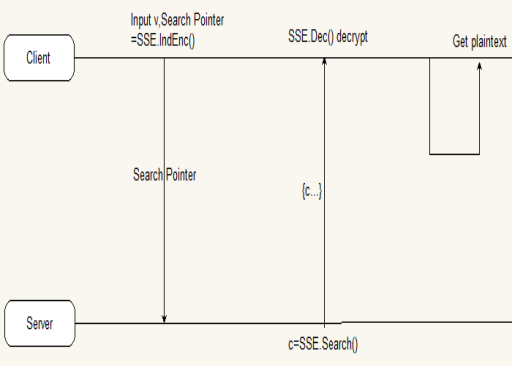

Figure 2. The whole search process of chart

5 Conclusion

Through the above analysis,programme extract keywords from user's files,and transform plaintext files and keywords to ciphertext and encrypted index respectively. These two steps can be done on the client.Ciphertext and encrypted indices are stored on cloud server.When the user queries, client hosts use the search algorithm to process the keyword to a query pointer and submit to the server.The server uses the pointer to call the user's encrypted search index.Server use the user's encrypted index to find the corresponding ciphertext and returns to the client, the client can use the key to decrypt the ciphertext.

References:

1. Fei X, Liu C Y, Fang B X, et al. Research on ciphertext search for the cloud environment[J]. Journal on Communications, 2013, 34(7):143-153.
2. KAMARA S,PAPAMANTHOU C,ROEDER T.Dynamic searchable symmetric encryption[A].Proceedings of the 2012 ACM conference on Computer and Communications Security(CCS 12)[C].New York,NY,USA,2012.965-976
3. Liu A F, An effective and Dynamic Ciphertext Search Method for Cloud Storage[D],2013
4. Neng-Hai Y U, Hao Z, Jia-Jia X U, et al. Review of Cloud Computing Security[J]. Tien Tzu Hsueh Pao/acta Electronica Sinica, 2013, 41(2):371-381.

5. Qi W U, Wan C X. Multi-user Conjunctive Keyword Search Scheme over Ciphertext[J]. Computer Science, 2013.
6. Yuan F,Zhou C,Zhang Q L, et al. Ciphertext search method based on word-for-word indexing: CN, CN 101937464 B[P]. 2012.
7. Qi W U, Wan C X. Multi-user Conjunctive Keyword Search Scheme over Ciphertext[J]. Computer Science, 2013.
8. Borisov N, Mitra S. Restricted queries over an encrypted index with applications to regulatory compliance[C]// Proceedings of the 6th international conference on Applied cryptography and network security. Springer-Verlag, 2008:373-391.
9. Peng L, Li R, Wang H, et al. An Encrypted Index Mechanism in Ciphertext Retrieval System[C]// Proceedings of the 2011 Eighth Web Information Systems and Applications Conference. IEEE Computer Society, 2011:131-136.
10. Kang H C, Mun D W, Kim Y H. Method for indexing encrypted column: US, US8281153[P]. 2012.