International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data

Daya Katherin

Student, Dept. of Dual Degree Computer Applications, Sree Narayana Guru Institute of Science and Technology

---

Abstract : *Cloud Computing is that the long-dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so on enjoy the on-demand top quality applications and services from a shared pool of configurable computing resources .As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, like emails, personal health records, company finance data, and government documents, etc. To guard data privacy, sensitive cloud data possesses to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization. When sharing files to the cloud storage the keywords that are highly repeating are taken and saves them in database as ranked keyword vector. In the ranked keyword search we should provide a rank to the keywords. In ranked keyword search encrypted domain is going to be used. While searching rank is also given so that our search result will contain the documents that has the given keyword corresponding to the rank. Cyber security is provided i.e., the keyword is encrypted and send to cloud and is decrypted on reaching the cloud. The encrypted keyword is divided into 3 portions such as a, b, c and performs XOR operation for security*.

1. INTRODUCTION

In Cloud Computing, data owners may share their outsourced data with an outsized number of users, who might want to retrieve certain only specific data files. One among the foremost popular ways to try to retrieve data is thru keyword-based search which features a series of demerits such large processing overhead, less accuracy, not secure etc. Cloud data owners like better to outsource documents in an encrypted form for the aim of privacy preserving. Therefore, it's essential to develop efficient and reliable cipher text search techniques. One challenge is that the connection between documents are going to be normally concealed within the process of encryption, which can cause significant search accuracy performance degradation. Also, the quantity of knowledge in data centres has experienced a dramatic growth. This may make it even tougher to style cipher text search schemes which will provide efficient and reliable online information retrieval on large volume of encrypted data .During this project, a hierarchical clustering method is proposed to support more search semantics and also to satisfy the demand for fast cipher text search within an enormous data environment. Within the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. So as to verify the authenticity of search results, a structure called minimum hash sub-tree is meant during this project. Experiments are conducted using the gathering set built from the IEEE Explore. The results show that with a pointy increase of documents within the dataset the search time of the proposed method increases linearly whereas the search time of the normal method increases exponentially. Furthermore, the proposed method has a plus over the normal method within the rank privacy and relevance of retrieved documents.

The encryption algorithm uses within the application is that the Advanced Encryption Standard (AES). To search the file collection for a given keyword a licensed user generates and submits an enquiry request of the keyword to the cloud server. Upon receiving the search request, the cloud server is responsible to look the index and return the corresponding set of files to the user. Data users can access the system using an Android client application. The client must establish a connection to the server using an authentication mechanism. Once a connection is established, users can search data using keywords and may download required data files. The files are listed within the ascending order of the rank score of searched keywords. Once the file has been downloaded into the device of the users the files are decrypted to the first form.

2. EXISTING SYSTEM

It is performed by one of the most popular way that is the keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenario. The existing system will take the searchable keyword from the user and compare with each and every document of corresponding application or service in the cloud server thus retrieving matching documents. Even though the information retrieval is possible but it's restricted to the plain text.

- Simpler and Convenient method for plain text
- Information can be retrieved easily
- If directly applied on encrypted cloud data then the users have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing over-head.
- Since it prefers to search over unencrypted cloud data the security and the integrity of the data is a concern.

3. PROPOSED SYSTEM

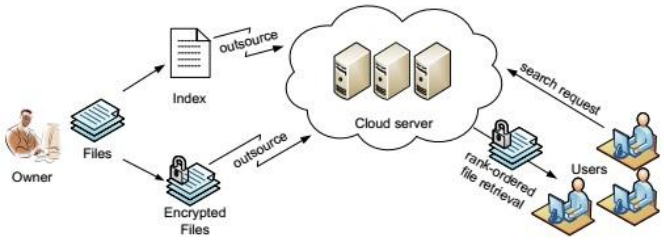
An encrypted cloud data hosting service involves three different entities: data owner, data user, and cloud server. Data owner features a set of data files that he wants to outsource on the cloud server in encrypted form while still keeping the potential to seem through them for effective data utilization reasons. To do so, before outsourcing, data owner will first build a secure searchable index from a gaggle of distinct keywords extracted from the file collection and store both the index and thus the encrypted file collection on the cloud server. The authorization between the info owner and users is appropriately done. To look the file collection for a given keyword a licensed user generates and submits an enquiry request of the keyword to the cloud server. Upon receiving the search request the cloud server is responsible to look the index and return the responding set of files to the user. The secure ranked keyword search is performed a touch just like the search result should be returned according to certain ranked relevance criteria (e.g., keyword frequency-based scores) to strengthen file retrieval accuracy for users without prior knowledge on the file collection. To reduce bandwidth, the user may send an optional value in conjunction with the trapdoor and cloud server only sends back the absolute best most relevant files to the user's interested keyword. The client application should initially need to register to the cloud server where the server provides with a username and password. The client application uses these username and password for authentication procedure

• The major advantage is that system uses some statistical measurement for evaluating relevance score in the information retrieval process for the user and that doesn't need to go through each of every retrieved file in order.
• Fulfils the secure ranked search functionality with little relevance score information leakage against keyword privacy.
• Symmetric encryption scheme indeed enables "as-strong-as-possible" security guarantee
• Another advantage is that it does not invariably sends back all files solely based on presence or absence of the keyword thus reduce large unnecessary network traffic.

4. ARCHITECTURAL DESIGN

The output of this process is the description of software architecture. This process is concerned with establishing a basic structural framework for the system. It involves identifying the major components of the system and communication between these components. The first phase of architectural design activity is usually concerned with decomposing a system into a set of interfacing sub systems. The system is structured into a number of sub systems. At its most abstract level, an architectural design may be depicted as a block diagram where each box in the diagram represents a sub system. Arrows means that, the control is passed to subsystem to subsystem in the direction of arrows. An architectural block diagram represents an overview of the system structure.

The entire system is designed to make the man machine interaction easy. The products of the design are models that enable us to reason about the structure, make trade-off when requirements conflict, and in general, provide a blueprint for implementation



Architecture for search over encrypted cloud data

4.1 MODULE DESCRIPTION

4.1.1 DATA OWNER

- Register the users and distribute username and password so that they can login and access files.
- Uploading the necessary files into the cloud server
- Calculating rank scores of the keywords in the files using various statistical procedures and store in the database
- New keyword can be added it will check the availability of the file and update also the latest rank score.
- Encrypting the files using Data Encryption Standard (DES) and stored inside the cloud server.
- Also, able to view the user lists, the rank list and delete the users and the file entries

4.1.2 USER

- The user will log into the application through the username and password provided by the data owner.
- User enters a keyword that he/she needed to search in the cloud server.
- A list will be generated based on the keyword and ascending order of the rank score of the particular keyword.
- The user can select the required files that from the list.
- The requested files will be downloaded into user's mobile device and which will be in decrypted form

5. ALGORITHM

Algorithm 1 One-to-many Order-preserving Mapping

```
1: procedure OPMK(D,R,m,id(F))
2: while |D| ! = 1 do
3: {D,R}← BinarySearch(K,D,R,m);
4: end while
5: coin R ←− TapeGen(K,(D,R,1||m,id(F)));
6: c coin ←−R;
7: return c;
8: end procedure
9: procedure BinarySearch(K,D,R,m);
10: M ←|D|; N ←|R|;
11: d ← min(D)−1; r ← min(R)−1;
12: y ← r +dN/2e;
13: coin R ←− TapeGen(K,(D,R,0||y));
14: x R ←− d + HYGEINV(coin,M,N,y−r);
15: if m ≤ x then
16: D←{d + 1,...,x};
17: R←{r + 1,...,y};
18: else
19: D←{x + 1,...,d + M};
20: R←{y + 1,...,r + N};
21: end if
22: return {D,R};
23: end procedure
```

Algorithm 2 Reversing One-to-many Order-preserving Mapping-ROPM

```
1: procedure OPMK(D,R,c,id(F))
2: while |D| ! = 1 do
3: {D,R}← BinarySearch(K,D,R,c);
4: end while
5: m ← min(D);
6: coin R ←− TapeGen(K,(D,R,1||m,id(F)));
7: w coin ←−R;
8: if w = c then return m;
9: end if
10: return ⊥;
11: end procedure
12: procedure BinarySearch(K,D,R,c);
13: M ←|D|; N ←|R|;
14: d ← min(D)−1; r ← min(R)−1;
15: y ← r +dN/2e;
16: coin R ←− TapeGen(K,(D,R,0||y));
17: x R ←− d + HYGEINV(coin,M,N,y−r);
18: if c ≤ y then
19: D←{d + 1,...,x};
20: R←{r + 1,...,y};
21: else
22: D←{x + 1,...,d + M};
23: R←{y + 1,...,r + N};
24: end if
25: return {D,R};
26: end procedure
```

6. SYSTEM IMPLEMENTATION

Implementation is that the process of converting a replacement or revised system design into operation. it's the key stage in achieving a successful new system because, usually it reveals tons of up heal. It must therefore be carefully planned and controlled. Aside from planning the 2 major tasks of preparing for implementation are education and training of users and testing of the system. Implementation is that the stage of project where the theoretical design is becoming working system or it's the key stage in achieving a successful new system. Therefore, it must be carefully planned and controlled. It also can be considered to be the foremost crucial stage in achieving a successful new system and in giving the user confidence that the new system will work and be effective.

7. SECURITY

The source node transmits data to destination once destination received the data it has to know the data transmission path. So the destination send the request to server with IP address and token (Sequence Number).The server already keeps all the records about the nodes, after destination request received in server it will validate the IP address and token if the IP address and tokens are valid server sends the response as Trace-back IP Path to destination.

8. CONCLUSIONS

In this project, we investigated ciphertext search in the scenario of cloud storage. We explore the problem of maintaining the semantic relationship between different plain documents over the related encrypted documents and give the design method to enhance the performance of the semantic search. We also propose the MRSE-HCI architecture to adapt to the requirements of data explosion, online information retrieval and semantic search. At the same time, a verifiable mechanism is also proposed to guarantee the correctness and completeness of search results. In addition, we analyses the search efficiency and security under two popular threat models. An experimental platform is built to evaluate the search efficiency, accuracy, and rank security. The experiment result proves that the proposed architecture not only properly solves the multi-keyword ranked search problem, but also brings an improvement in search efficiency, rank security, and the relevance between retrieved documents.

9. FUTURE SCOPE

- We can do blocking of attacked IP

If the Symantec Endpoint Protection client detects a network attack, it can automatically block the connection to ensure that the client node is safe. The client activates an Active Response, which automatically blocks all communication to and from the attacking node for a set

period of time. The IP address of the attacking node is blocked for a single location. The attacker's IP address is recorded in the Security log. We can unblock an attack by canceling a specific IP address or canceling all Active Response.

- Optimizing node relationships in a super-router network

In structured routed systems, consistency maintenance and load balancing can be achieved through the super node topology, which exploit the heterogeneity of nodes in a routing network by assigning additional responsibilities to high capacity nodes called super nodes. Every node is assigned to a fixed, very small number (usually one) of super nodes. Consequently, super nodes become bottlenecks in terms of fault tolerance. Restoring the system structures such as routing tables back to a consistent state after a super-node crash requires a considerable effort.

REFERENCES

[1] S. Grzonkowski, P. M. Corcoran, and T. Coughlin, "Security analysis of authentication protocols for next-generation mobile and CE cloud services," in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp. 83–87.
• D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44–55.
• D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. EUROCRYPT, Interlaken, SWITZERLAND, 2004, pp. 506–522.
• Y. C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. 3rd Int. Conf. Applied Cryptography Netw. Security, New York, NY, 2005, pp. 442–455.
• R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, Alexandria, Virginia, 2006, pp. 79–88.
• M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Proc. 27th Annu. Int. Cryptol. Conf. Adv. Cryptol., Santa Barbara, CA, 2007, pp. 535–552.
• D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proc. 4th Conf. Theory Cryptography, Amsterdam, NETHERLANDS, 2007, pp. 535–554.
• E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol. 2003, pp. 216. 20