AN EFFICIENT KEYWORD SEARCH RETRIEVAL BASED ON THE AES ENCRYPTION ALGORITHM OVER DIFFERENTIAL QUERY SERVICES

Prof C.SIVAKUMAR
Asscio.Prof dept of CSE,
Anurag Engineering College, Telangana.

Abstract— *The needs of Cloud computing is increasing due to massive increase of user access to the cloud databases. The more number of users are trying to access the cloud databases to fulfill their storage requirements where the cloud service providers need to focus on providing efficient services. In the existing work, EIRQ technique is implemented where it aims to retrieve the documents based on user requirements and also focus on reduction of communication cost. The EIRQ doesn't concentrate on retrieving most similar documents to the users. Hence it needs to be concentrated to improve the user friendly environment. in this work, the page ranking scheme is introduced which concentrates on retrieving the most similar documents to the users. This approach improves the user friendly environment as well as it tries to focus on the reduction of communication cost.*

Index Terms— **Cloud Computing, AES algorithm, Page ranking**

1. INTRODUCTION

Cloud computing as an emanate technology to imperative information technology process in future. Many organizations choose to out-source their data for sharing in cloud. An organization supports the cloud services and authorizes its staff to share files in the cloud, its typical in cloud application. Each file is related by set of keywords. The staff as authorized users for retrieving files. They can retrieve files of their interests by querying the cloud with certain keywords. Here the key problem is that user privacy. The user privacy is a third party outside the security boundary.

The User privacy is classified into two types.1) Search privacy 2) Access privacy. The cloud knows nothing about what the user is searching for is called Search privacy, and the cloud knows nothing about which files are returned to the user is called access privacy.

Cooperate private searching protocol (COPS) as a proxy server, called as Aggregation and Distribution layer (ADL).The ADL is intermediate between the users and the cloud. The ADL expand two main functionalities inside the organization, which is aggregating user queries and distributing search results. Under the ADL, the computation cost on the cloud can be widely reduced, since the cloud only needs to execute a combined query once, no problem how many users are executing queries. The files are shared by the users need to be returned only once. Most importantly, COPS can protect user privacy from the ADL, other users and the cloud by using a series of secure functions.

The existing scheme, termed Efficient Information retrieval for Ranked Query (EIRQ), in which each user can choose the rank of his query, which is used to determine the percentage of matched files to be returned. The idea of EIRQ is before returning to the ADL to construct a privacy-preserving mask matrix that allows the cloud to filter out a certain percentage of matched files .This is not a trivial work, as the cloud needs to set rank of queries without knowing anything about user privacy correctly filter out files.

2. BACKGROUND

2.1 SYSTEM MODEL

The system model consists of three entities. They are Aggregation and Distribution layer, the cloud and the many users. Figure 1 shows that the only one ADL in this paper.

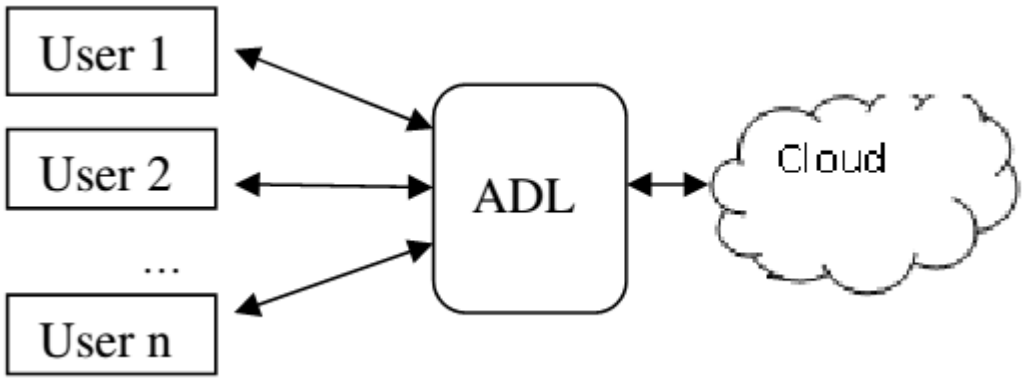

Figure 2.1 System model

The queries are sending to the ADL by the authorized users. The ADL aggregate users queries and send as combined query to the cloud. Then, the combined queries are processed by the cloud on the file collection and send a buffer. The buffer involve of all matched files to the ADL. The ADL will distribute the search results to each user. In this method the organization may require the ADL to wait for a period of time before running our schemes, which may get a certain querying delay.

3. SCHEME DESCRIPTION

In this section, the EIRQ scheme described in three schemes.1) EIRQ Efficient,2) EIRQ Simple and 3) EIRQ privacy scheme .By comparing all the scheme the EIRQ Efficient scheme provide less communication cost.

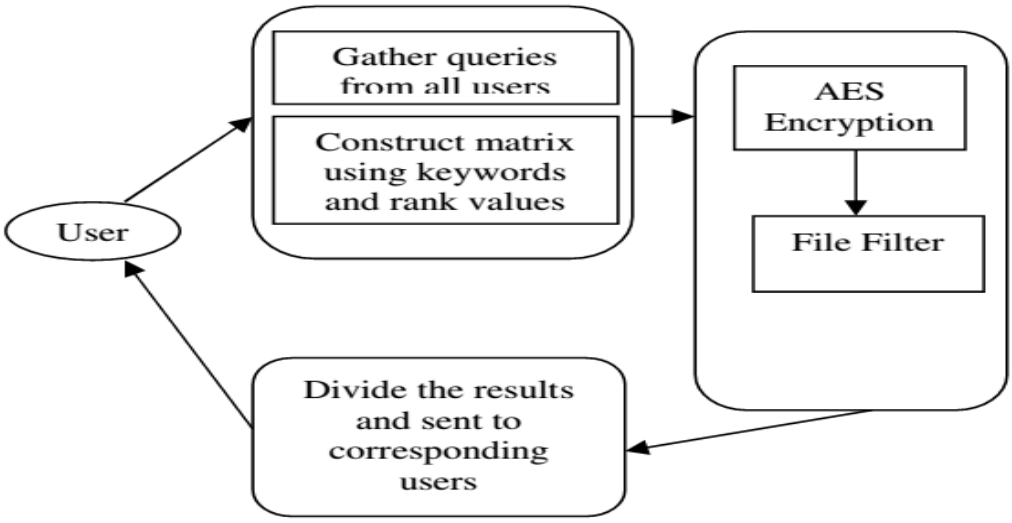

Figure 3.1 Architecture Diagram

3.1 THE EIRQ-EFFICIENT SCHEME

The EIRQ-Efficient scheme should be resolved two fundamental problems. First, we should determine the relationship between query rank and the percentage of matched files to be returned. Else that queries are classified into 0 to r ranks. Rank-0 queries have the highest rank and the Rank-r queries have the lowest rank. This relationship by allowing Rank-i queries to retrieve ð1 _ i=rÞ percent of matched files. Finally Rank-0 queries can retrieve 100 percent of matched files, and Rank-r queries cannot retrieve any files.

Secondly, we should determine which matched files will be returned and which will not. In this paper, we simply fix the probability of a file being produces by the highest rank of queries matching this file. Specifically, we first rank each keyword by the highest rank of queries selecting it, and then rank each file by the highest rank of its keywords. If the file rank is i, then the possibility of being filtered out is i=r. Therefore, Rank-0 files will be mapped into a buffer with probability 1, and Rank-r files will not be mapped at all. Since unneeded files have been filtered out before mapping, the mapped files should survive in the buffer with probability 1. we will illustrate how to adjust the buffer size and mapping times to achieve this goal.

EIRQ-Efficient mainly consists of four algorithms. The algorithms are 1) 2) QueryGen 2) Matrix Construct 3)Fle filter and 4) ResultDivide are easily under-stood.

Step 1: The user sends the keyword and the rank of the query to the ADL by using QueryGen algorithm.
Step 2:The ADL runs the MatrixConstruct algorithm after aggregating enough user queries,to send a mask matrix to the cloud. The mask matrix M consists that d-row and r-column matrix, where d is the number of keywords, and r is the lowest query rank.
Step 3. The cloud runs the FileFilter algorithm to return a buffer. The buffer contains a certain percentage of matched files to the ADL. Here the DES algorithm used.
Step 4. To distribute search results to each user by the ADL runs the Result Divide algorithm.We require the cloud to attach keywords to the file content to allow the ADL to distribute files correctly.By executing keyword searches the ADL can find out all of the files that match users' queries.

4. PROPOSED METHOD

Step 1:The user sends the keyword and the rank of the query to the ADL by using QueryGen algorithm.
Step 2:The ADL runs the MatrixConstruct algorithm after aggregating enough user queries,to send a mask matrix to the cloud. The mask matrix M consists that d-row and r-column matrix, where d is the number of keywords, and r is the lowest query rank.
Step 3: The cloud runs the FileFilter algorithm to return a buffer. The buffer contains a certain percentage of matched files to the ADL. Here the AES algorithm used.
Step 4: To distribute search results to each user by the ADL runs the ResultDivide algorithm.We require the cloud to attach keywords to the file content to allow the ADL to distribute files correctly.By executing keyword searches the ADL can find out all of the files that match users' queries

4.1 ADVANCED ENCRYPTION SECURITY ALGORITHM

The AES encryption algorithm will provide more security when comparing to the DES encryption algorithm. The AES algorithm provides less communication cost. This algorithm works after the file filtering process, and it will works on the cloud.
Algorithm Steps: The AES Encryption algorithm will work based on the following steps.

Step :1 KEY Expansion:

To generate ten sub keys for each ten AES rounds the 512 bit input key of the new AES-512 algorithm is used.

Step 2: INITIAL ROUND:

1) **ADD ROUNDKEY**

Each byte of the state is merging with a block of the round key using bitwise xor.

2) **ROUNDS**

a) **Subbytes**

A non linear substitution step where each byte is alternated with another according to a lookup table.

b) **ShiftRows**

A transposition step where the last three of the state are shifted exactly a certain number of steps.

c) **MixColumns**

A mixing operation which operates on the columns of the state,grouping the four bytes in each column.

d) **AddRoundkey**

To make the relationship between the key and the cipher text more complicated and to satisfy the confusion principle, the Add Round Key operation is performed.

5. RESULT ANALYSIS

5.1 TRANSFER TIME

Figure 5.1 shows that the AES based encryption scheme takes less time when compared with the DES based encryption scheme. The AES based encryption scheme detects 20% of transfer time. So the AES based encryption is reduce the transfer time and fastly provide the query results.

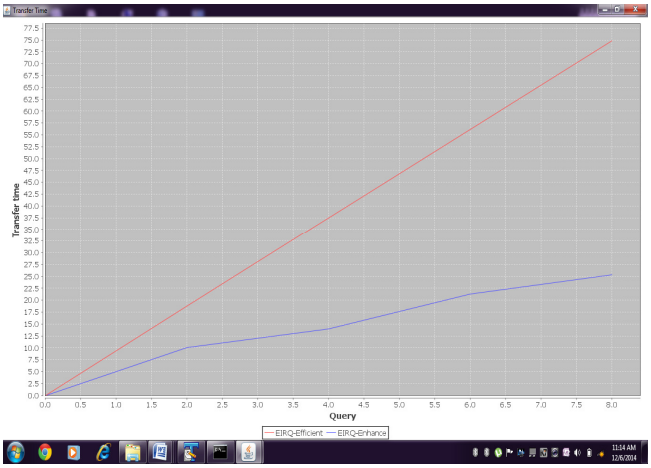

Figure 5.1 Number of Queries Vs.Transfer Time

5.2 COMUNICATION COST

Figure 5.2 shows that the AES based encryption scheme takes less communication cost when compared with the DES based encryption scheme. The AES based encryption scheme detects 70% of Communication cost. So the AES based encryption is reduce the Communication cost and fastly provide the query results.

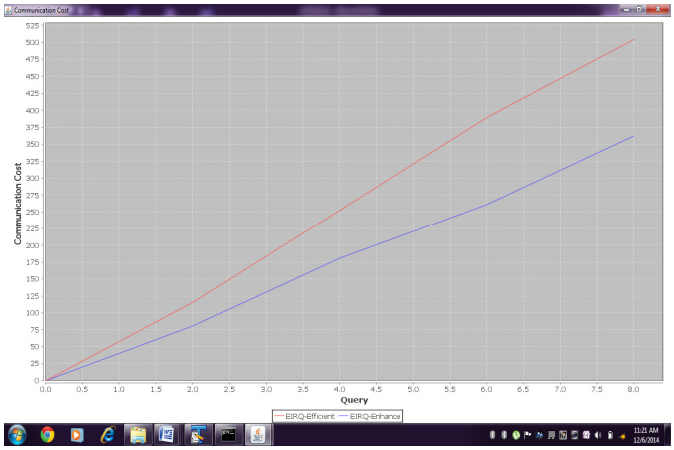

Figure 5.2 Number of Queries vs.Communication Cost

6. CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

The user privacy is an important issue in the cloud computing when requesting for an contents stored in the cloud storage. It will become burden for cloud service providers for handling the differential query service from the uers. The EIRQ scheme provides a differential query services with the user privacy. It works based on the ranking of users query. The cloud sends the result based on the DES algorithm. The DES algorithm takes more time to provide the result and also it provide the large decrypting data results. The AES algorithm takes less time when comparing with DES algorithm. The communication cost is also reduced. The memory space also decreased. The security is automatically increased when using AES algorithm. The experimental result proves that the AES algorithm provided an better optimized resource provisioning in which cost and time can be reduced considerably than the DES algorithm.

6.2 FUTURE WORK

In the future information discovery is used to support the differential queries from the users where the ranking of files can be done by using the page ranking method. This ranking is done based on the information discovered in order to retrieve the most similar files to the users, which may improve the user environment.

To retrieve the most similar documents, for work page ranking scheme is introduced which will retrieve the contents from the most popular web sites.

REFERENCES

1. *Qin Liu,Chiu C,Jie Wu,and Guojun Wang,(2014) Towards Differential Query services in Cost-Efficient Clouds,'IEEE Transactions on parallel and Distributed Systems.*
2. *2. Boneh.D,Crescenzo.D,Ostrovsky.R,and Persiano.G,(2004) Public- Key with Keyword Search,' Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques.*
3. *3. Cao.N,Wang.C, Ren.M Li, K. and Lou.W,(2011)'Privacy-Preserving Multi keyword Ranked Search over Encrypted Cloud Data,' Proc. IEEE INFOCOM.*
4. *Coron.J.S,Mandal.A, Naccache.D and Tibouchi.M, (2011) 'Fully Homomorphic Encryption over the Integers with Shorter Public Keys,' CRYPTO '11: Proc. 31st Ann. Conf. Advances in Cryptology.*
5. *5. Curtmola.R, Garay.J.A, Kamara.S, and Ostrovsky.R, (2006) 'Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,'Proc. ACM 13th Conf. Computer and Comm. Security.*
6. *Golle.P, Staddon.J, and Waters.B, (2004) 'Secure Conjunctive Keyword Search over Encrypted Data,' Proc. Second Int'l Conf. Applied Cryptography and Network Security (ACNS), pp. 31-45.*

7. *Hu.H, Xu.J, Ren.C, and Choi.B, (2011) 'Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism,' Proc. IEEE 27th Int'l Conf. Data Eng. (ICDE).*
8. *Song.D, Wagner.D, and Perrig.A, (2000) 'Practical Techniques for Searches on Encrypted Data,' Proc. IEEE Symp. Security and Privacy.Huseyin Ozg ur Tan and Ibrahim Korpeo,IEEE, "Power Efficient Data Gathering and Aggregationin Wireless Sensor Networks",December 2003.*