Private-Searching Techniques in Cloud Environment

Mr. Suryabhargav R, Smt. Kavitha M

***Abstract*— The Cloud computing needs are increasing due to more of user access to the cloud databases. The cloud service providers should provide efficient services since more number of users are trying to access cloud databases for their storage requirements. But, in this scenario the issue that is to be addressed is Security. Hence User privacy must be protected in addition to querying of cloud. Therefore different private searching techniques are developed by researchers to provide privacy. This paper presents description of different private searching techniques like Ostrovsky, COPS, EIRQ (with trusted ADL) and EIRQ (No trusted ADL) protocols which are currently available for retrieving information from clouds. The Pallier cryptosystem is used by the secure search protocols as an encryption scheme.**

***Index Terms*— Efficient information retrieval for ranked query (EIRQ), Cloud computing, User privacy, Aggregation and Distribution Layer (ADL), Co-operative private searching (COPS).**

I. INTRODUCTION

Cloud computing is an emerging technology which is adopting rapidly. Because of overwhelming merits of cloud computing like cost-effectiveness, flexibility and scalability, many organizations out-source their data for sharing in cloud. In this environment, an organization allows its staff to share files by subscribing to cloud services and the organizations will pay the amount only for their usage time of server. Here keywords are used to represent different, and the files can only be accessed by its authorized staff. The staff can query cloud with these keywords to retrieve files of their interest. In such a scenario, User privacy must be preserved, which is outside the security boundary of the organization and this becomes a key problem. Search privacy and Access privacy are the two forms of User privacy [12]. Search privacy is, what user is searching should not be known to cloud, and access privacy is, that the files that are returned to users are not known to cloud.

II. RELATED WORK

Many algorithms were proposed for private searching in cloud. Private searching is proposed by Ostrovsky [1], pallier cryptosystem is used to encrypt queries and the data will be stored in clear form. The compact buffer is used by cloud to store all files, and user can recover files of their interest successfully with high probability. In the following work, private searching techniques [2] reduced the communication cost in Ostrovsky scheme [1] by solving a set of linear programs. Both the computation and communication costs increases with the number of searches made by the user become the main limitation of the current private searching techniques. Querying costs will be extensive when applying these schemes in cloud environment whose scale is large. Ranked searchable encryption allows users to fetch most matched files from the cloud. The work by secure ranked keyword search [8], supports only single-keyword searches, Order Preserving Symmetric Encryption (OPSE) [9] is used to encrypt files and queries and uses frequencies of keywords to rank results. The scheme [10], Privacy-preserving multi-keyword ranked search that supports multiple-keyword searches, secure KNN technique [11] is used to rank results. The main limitation of these approaches is that user privacy [6] will not be preserved. Let us look at the algorithms like Ostrovsky scheme, COPS protocol, EIRQ scheme (with Trusted ADL) and EIRQ scheme (No Trusted ADL).

III. LITERATURE SURVEY

A. *Ostrovsky Scheme*

Ostrovsky scheme proposed by Ostrovsky [1] is the first Private searching scheme introduced. This scheme allows a user to retrieve files of their interest from the server which is untrusted without leaking any information

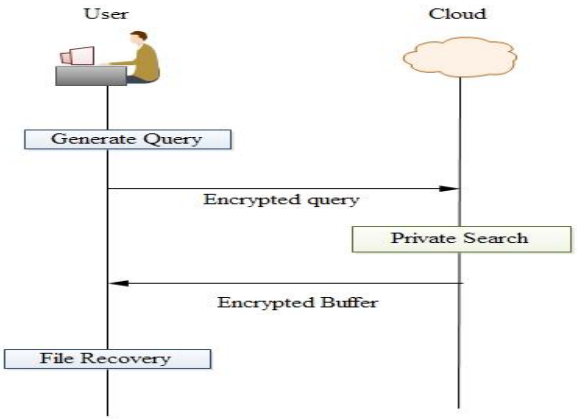

Fig: Ostrovsky Scheme

Ostrovsky Scheme mainly consists of user and cloud. Only authorised users are permitted can access files and access is denied for unauthorised users. The initial step is to send request from the user to cloud for establishment of a connection. Then authorized user should have own login name and passwords. After login, the user generates a query [2]. This query is encrypted in the form of 0‟s and 1‟s and then this query is sent to cloud. At the cloud side Private Search has been done to find out matched files. Then Cloud

***Manuscript received June, 2015*.**

***Suryabhargav R**, M.Tech Student, Dept. of Computer Science, Siddaganga Institute of Technology, Tumkur, India, Phone/ Mobile No: +91-9591548748.*

***Kavitha M**, Assistant. Professor, Dept. of Computer Science, Siddaganga Institute of Technology, Tumkur, India, Phone/ Mobile No: +91-9611449444.*

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

sends the matched files to encrypted buffer. Then Files are recovered at the user side. This scheme involves query overhead is more costly to access files at every query.

Merits:

- Ensure privacy by using Paillier cryptosystem.

Demerits:

- Ostrovsky protocol does not permit aggregation of queries.
- The cost that is incurred by the customers is not reduced in this scheme.

B. COPS

COPS (Cooperative private searching) protocol introduces an aggregation and distribution layer (ADL), a middleware layer that exists between the users and the cloud was proposed by Qin Liu et al. in [3].

Aggregation of queries that are received from the users and distribution of results to the users from the cloud will be done by intermediate layer, ADL.

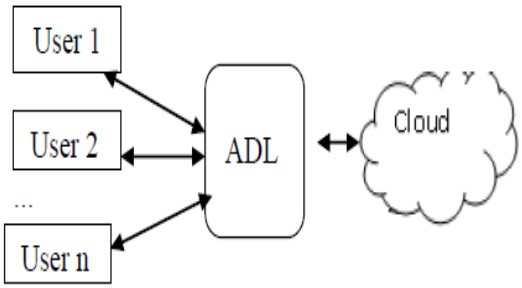

Fig: COPS Scheme

In COPS Protocol, each individual user generates a query. These queries are sent to ADL without encryption since it is trusted. The ADL merges all the user queries that are received and sends requests to the cloud after aggregation. The public key of organisation is used for encryption here. The cloud will find files matching the combined query and sends matched files to the ADL and then the appropriate files are distributed to the users by ADL.

Merits:

- Costs are reduced because of merging of queries and also the user privacy is preserved.

Demerits:

- CPU consumption will be more because COPS protocol sends all matched files to users and also bandwidth used is more for all matched files to be transferred.
- When users cannot tolerate delays, then multiple ADLs to be deployed within the organisation.

C. EIRQ (With Trusted ADL)

EIRQ (Efficient Information Retrieval for Ranked Queries), where users send the queries to the cloud and cloud processes queries and generates results and then sends it to users. In this case lot of files matches users query. But most of the times, users are interested on certain percentage of files.

The EIRQ model consists of the cloud, organization and ADL. ADL is placed inside the organization based on requirement of users.

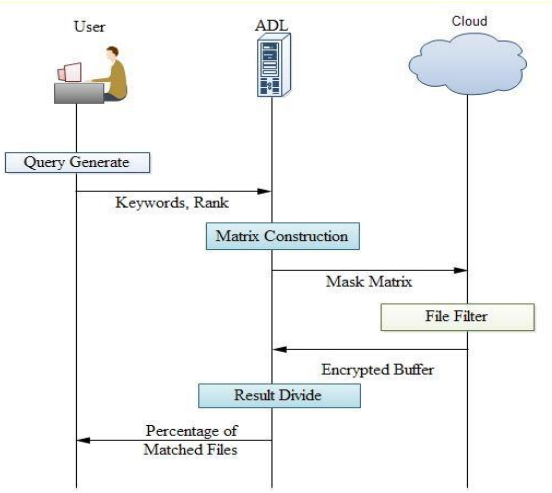

Fig: EIRQ Scheme (with Trusted ADL)

In this scheme, only authorised users are permitted to access files from the cloud. Here first user sends request to ADL for establishment of a connection from the ADL. Then authorized user should have own login name and passwords. After login user generates a query. This query is encrypted into 0"s and 1"s and then sends to ADL. Then the ADL runs Matrix Construct Algorithm [4] based on the Keywords and Ranks. This process is called as Aggregation. After the aggregation process, ADL sends the Mask Matrix to Cloud. Then the cloud runs FileFilter Algorithm to filter out the files based on the Ranks and keywords and sends encrypted buffer to ADL. The ResultDivide algorithm is used by ADL to distribute appropriate files to each user and then FileRecover algorithm is run by individual user to retrieve matched files.

Merits:

- Lowers communication cost and reduces computation overhead by aggregation of queries.
- Ensure User privacy by constructing mask matrix.

Demerits:

- The ADL as an aggregator will collect all messages between the users and the cloud, and may become the biggest attacking target.

D. EIRQ (No Trusted ADL)

To avoid possible information leaking, it may be required to hide user privacy from the ADL. Here Combiner and the Distributor substitute the ADL. Three kinds of secure functions are used, the secret seeds of which are shared by the

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

users and the cloud: Shuffle function is used to shuffle a dictionary, Pseudonym function is used to calculate file pseudonym and Obfuscate function is used to obfuscate file content.

In this scheme, Alice and Bob send their mask matrices, denoted as MAlice and MBob to the Combiner and send their shuffled queries denoted as QAlice and QBob to the Distributor.

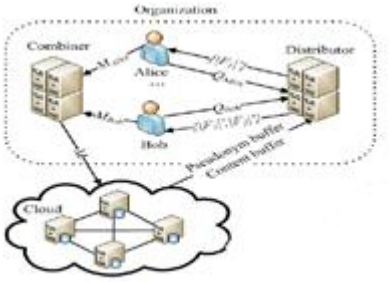

Fig: EIRQ (No Trusted ADL)

Each user sets the values of mask matrix elements as the Matrix Construct algorithm in EIRQ-Efficient, and encrypts each element under the Distributor's public key pk. The shuffled queries are 0-1 bit strings of d bits. Each user first shuffles the dictionary and then sets the i-th bit of the query to 1 only if ith keyword is chosen from shuffled dictionary. Then the Combiner generates a combined mask matrix M by performing multiplications on MAlice and MBob element by element and sends M to the cloud. The cloud processes M on the whole file collection, and returns two buffers, pseudonym buffer and content buffer, to the Distributor. The Distributor recovers from the pseudonym buffer and the content buffer it distributes to Alice and Bob. Each user recovers file content by dividing the obfuscated file content by the obfuscated factor.

Merits:

- User privacy is protected by constructing several mask matrices and encrypting them and using several secret functions.

Demerits:

- Combiner colludes with the user or the cloud. The Combiner will know the shuffle function, with which it can deduce which keywords each user is searching for.

- The Distributor will know the obfuscate function and the pseudonym function, with which it can deduce which files are returned to each user.

IV. CONCLUSION

Cloud computing is used for sharing and retrieving information. In this paper, different techniques for searching over outsourced encrypted data are described. This study concludes that retrieval that is based on rank is most efficient for searching on data that is encrypted, because it is more secure, search access is fast and information will not be leaked to authorities which are not trusted. It is required to get desired information from the cloud with optimal communication cost and minimal computation overhead..

REFERENCES

[1] R. Ostrovsky and W. Skeith III, "Private searching on streaming data," in Proc. of ACM CRYPTO, 2005.

[2] J. Bethencourt, D. Song, and B. Waters, "New techniques for private stream searching," 2009.

[3] Q. Liu, C. Tan, J. Wu, and G. Wang, "Cooperative Private Searching in Clouds," J. Parallel Distributed Computing, Aug. 2012.

[4] Q. Liu, C. C. Tan, J. Wu, and G. Wang, "Efficient information retrieval for ranked queries in cost-effective cloud environments," IEEE INFOCOM, 2012.

[5] P. Mell and T. Grance, "The nist definition of cloud computing" NIST Special Publication, 2011.

[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," ACM CCS, 2006.

[7] G. Danezis and C. Diaz, "Improving the decoding efficiency of private search," in IACR, 2006.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data, IEEE ICDCS, 2010

[9] A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill, "Order-preserving symmetric encryption, EUROCRYPT, 2009.

[10] Ning Cao, Cong Wang , Li, Ming , Kui Ren, Wenjing Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" INFOCOM, 2011

[11] W. Wong, D. Cheung, B. Kao, and N.Mamoulis, "Secure knn computation on encrypted databases," ACM SIGMOD, 2009.

[12] Qin Liu, Chiu C. Tan, Jie Wu and Fellow (2013) "Towards Differential Query Services in Cost-Efficient Clouds" IEEE Transactions On Parallel and Distributed Systems, vol. 20, no.10, pp-1-11.

Suryabhargav R received the Diploma in Computer Science and Engineering from Govt. Polytechnic Davangere and B.E. in Computer Science and Engineering from BIET Davangere and currently pursuing Masters in Computer Science and Engineering from Siddaganga Institute of Technology, Tumkur.

Kavitha M received B.E. in Computer Science from SIT, Tumkur and M.Tech. in Computer Science from SIT, Tumkur and currently working as Assistant professor in Siddaganga Institute of Technology, Tumkur

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET