International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 11 3560 – 3564

Survey on Efficient Information Retrieval for Ranked Query in Cost-Efficient Clouds

Ms. Jyotsna T. Kumbhar[1]
ME Student, Department of Computer Engineering, TSSM'S, P.V.P.I.T., Bavdhan,
Pune University, Pune, Maharashtra, India, 411021
kumbharjyotsna90@gmail.com

Mr. Navnath D. Kale[2]
Assistant Professor, Department of Computer Engineering, TSSM'S, P.V.P.I.T., Bavdhan,
Pune University, Pune, Maharashtra, India, 411021
navnath1577@yahoo.co.in

***Abstract*—** Cloud computing technology redefines the advances in information technology. The most challenging research works in cloud computing is privacy and protection of data. Cloud computing provides an innovative business model for organizations with minimal investment. Cloud computing has emerged as a major driver in reducing the information technology costs incurred by organizations. Security is one of the major issues in cloud computing. So it is necessary to protect the user privacy while querying the data in the cloud environment, different techniques are developed by researchers to provide privacy, but the computational and bandwidth costs increased which are unacceptable to the users. This paper presents description and comparison of Ostrovsky, COPS and EIRQ protocols which are currently available for retrieving information from clouds. EIRQ protocol is the latest among these protocols and it addresses the issues of privacy, aggregation, CPU consumption and network bandwidth usage.

Keywords- *cloud computing, user privacy, encryption, ADL, mask matrix, Cooperative private searching protocol (COPS).*

*****

I. INTRODUCTION

Cloud computing is an emerging technology which is being used widely these days. Due to the cost-effectiveness, flexibility and scalability of cloud, more and more organizations are now using cloud to outsource their data for sharing. In a cloud computing environment, an organization subscribes the cloud services and gives access to its staff to share files in the cloud. Each file is identified with some keywords, and the staff, only authorized users can retrieve files, so they send query with those keywords to cloud and retrieve interested files. In such a scenario, protection of user privacy from the cloud, which is outside the security boundary of the organization, this becomes a key problem. User privacy can be classified into search privacy and access privacy [12]. Search privacy means that the cloud does not know about what the user is searching for, and access privacy means that the cloud does not know about which files are returned to the user. A naive solution is used to protect user privacy when the files are stored in the clear forms, so that the cloud cannot know which files the user is really interested in. user queries are classified into multiple ranks, so a new kind of user privacy that is rank privacy is introduced in cloud computing. Rank privacy is used to hide the rank of each user query from the cloud. While this does provide the necessary privacy, the communication cost is high. EIRQ protocol is the latest protocols and it addresses the issues of privacy, aggregation, CPU consumption and network bandwidth usage.

II. RELATED WORK

For the private searching in the cloud many algorithms were proposed. Private searching is proposed by [1], where the data is stored in the clear form, and the query is encrypted with the Paillier cryptosystem. The cloud stores all files into a compact buffer, with which the user can successfully recover all wanted files with high probability. In the following work, [2] reduced the communication cost in [1] by solving a set of linear programs; [7] presented an efficient decoding mechanism for private searching. The main drawback of the current private searching techniques is that both the computation and communication costs grow linearly with the number of users that are executing searches. Thus, when applying these schemes to a large-scale cloud environment, querying costs will be extensive. Ranked searchable encryption enables users to retrieve the most matched files from the cloud in the case that both the query and data are in the encrypted form. The work by [8], which only supports single-keyword searches, encrypts files and queries with Order Preserving Symmetric Encryption (OPSE) [9] and utilizes keyword frequency to rank results. Their following work [10], which supports multiple-keyword searches, uses the secure KNN technique [11] to rank results based on inner products. The main limitation of these approaches is that user access privacy [6] will not be preserved. Let us now see the three algorithms in detail i.e. Ostrovsky scheme, COPS protocol, EIRQ scheme. All these algorithms address to the private searching in the cloud environment.

III. LITERATURE SURVEY

Literature Survey analysis the project concept with the standard papers and journal. We can understand the methodology that follows and implementing some of new idea. To analysis the paper's main goal and find the possible way to applying in the different environment.

A. Techniques for searching over encrypted data

1) Paillier Cryptosystem

Public-Key Cryptosystems Based on Composite Degree Residuosity Classes
Publication: EUROCRYPT
Author : P. Paillier

All the secure search protocols presented in this paper use an encryption scheme known as Paillier cryptosystem. It is a public key cryptosystem with wide applications in cloud computing, electronic voting and other areas.

The Paillier Cryptosystem is a public key encryption scheme, developed by Pascal Paillier, with several interesting properties. This paper explores Paillier's work, this shows how to encrypt and decrypt messages using this cryptosystem, it uses mathematical principles that make the system work clearly outlined. In this scheme alphanumeric message is converting into a purely numeric message, which is broken into blocks, mi, such that, for each i, 0 < mi < n, for a predetermined value, n. in this the term plaintext is used to refer to a message that is numeric that is not encrypted, while the term cipher text is used to refer to plaintexts, that is not dencrypted. One property in particular, the addition of plaintexts through multiplication of cipher texts, it is looked at in terms of its potential application to a form of electronic voting, in order to illustrate the system's potential. Unlike RSA cryptosystem, Paillier cryptosystem results in a non-zero cipher text for a plaintext message of value 0. This feature facilitates masking the absence of certain keywords in the user queries. This paper presents a novel computational problem, called as Composite Residuosity Class Problem, and its applications to public-key cryptography. It presents a new trapdoor mechanism and three encryption schemes : a trapdoor permutation and two homomorphic probabilistic encryption schemes. These are computationally compared with RSA. This cryptosystems, based on usual modular arithmetics, are provably secure under appropriate assumptions in the Standard model.

2) Secure and ranked keyword search Cloud Data

Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data
Publication: IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS
Authors :Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou

Cloud computing economically enables the paradigm of data service outsourcing. However, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud to protect data privacy, it makes effective data utilization service a very challenging task. In traditional searchable encryption techniques users securely search over encrypted data through keywords, they use only Boolean search technique and it is not yet sufficient to meet the effective data utilization need that is inherently due to large number of users and huge amount of data files in cloud. This paper presents the problem and solution of secure ranked keyword search over encrypted cloud data. Ranked search enables search result relevance ranking and avoid undifferentiated results, and further ensures the file retrieval accuracy. Specifically, this work explores the statistical measure approach, like relevance score, from information retrieval to build a secure searchable index, and it develops a one-to-many order-preserving mapping technique to properly protect those sensitive score information. This design facilitates efficient server-side ranking without losing keyword privacy. Analysis shows that the proposed solution is compared to previous searchable encryption schemes. Experimental results of this work demonstrate the efficiency of the proposed solution. Paper presents a number of security-related research issues in Cloud data Access control. Early work concentrated on data authentication and integrity means how to efficiently and securely ensure that the server returns correct and complete results in response to its client's queries. Later research focused on outsourcing encrypted data with efficient querying over encrypted domain.

Issues in this scheme are

- This searching leads to collision in the
- Network.
- Encrypted files are largely processed after main searching.
- Largest encrypted files are post processed.

3) Searchable symmetric encryption: improved definitions an efficient constructions

Searchable symmetric encryption: improved definitions and efficient constructions
Publication: ACM CCS
Authors: R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky

This paper presents the Searching and retrieving of the outsourced data by using multi user Searchable Symmetric Encryption (SSE). This method focuses active research, several security definitions and constructions, which is achieved by Non-adaptive setting and Adaptive adversary [6]. This technique is used because it is more secure and efficient than other constructions, it Supports multi user setting and authentication is not require in this method.

Issues in this method

- This method is not suitable for large scale cloud data and cannot accommodate high level requirements.

4) Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data

Privacy-preserving multikeyword ranked search over encrypted cloud data
Publication: IEEE INFOCOM
Authors: Ning Cao, Cong Wang , Li, Ming , Kui Ren, Wenjing Lou.

This paper presents the searching of Encrypted cloud data using Privacy-Preserving Multi-keyword Ranked Search [10] (MRSE) method. In this paper co-ordinate matching technique is used. Coordinate matching is used to find the similarity between search query and data documents. Another technique that is Inner product Similarity, also used to describe the Multi-keyword Ranked Search over Encrypted Cloud Data (MRSE). Here four modules of searching that are Encrypt Module, Client Module, Multi-Keyword Module, and Admin Module are performed over encrypted cloud data [10].

The advantages of this method are

- High Efficiency, Multi-keyword Ranked Search, Privacy-Preserving.
- It is Eliminate unnecessary traffic and Improve Search accuracy.
- Similarity measurements also easily searched [10].

Issues in this method

- The Disadvantages of this is Single Keyword search with ranking and Boolean keyword search with ranking are not possible.
- This is not suitable for large scale cloud data.
- It provide much less semantics and this schemes are developed as crypto primitives.

B. Secure search protocols

We review three keyword-based private and secure search protocols, namely Ostrovsky, COPS and EIRQ.

1) Ostrovsky protocol

Private searching on streaming data
Publication: ACM CRYPTO
Author : R. Ostrovsky and W. Skeith III

A key privacy search solution was proposed by Ostrovsky et al. [1], which allows a user to retrieve files but this scheme is of high computational cost since it requires the cloud to process the query using homomorphic encryption on every file in a collection, allows a client to provide an untrusted server with an encrypted search query. The server uses the query on a stream of documents and returns the matching documents to the client. New scheme for conducting private keyword search on streaming data which requires server to client communication is been implemented and returns the content of the matching documents. The previous best scheme for private stream searching was shown to have communication and storage complexity. This technique requires a small amount of metadata to be returned in addition to the documents. Paper also gives an alternative method for returning the necessary metadata based on a unique encrypted download.

Ostrovsky Scheme: The Ostrovsky scheme is a process of accessing the files from cloud to clients. This process has the following steps:

1) Ostrovsky Scheme having the user and cloud. The users are only authorized [1] from the cloud network, and then only accessing is possible otherwise it is not possible.

2)First send request from the user to cloud for establishment of a connection form the cloud. Then authorized user should have their own login name and passwords.

3) After login to user Generate a query [2]. This query is encrypted into 0's and 1's and then sends to cloud. At the cloud side Private Search has been done. So those find out the matched files.

4) Cloud sends the matched files to encrypted [1] buffer. Then Files are recovered at the user side. This scheme is very query overhead as well as every time accesses the broadband connection. This process is more costly to accessing files at every query.

Issues in Ostrovsky Protocol

- Ostrovsky protocol suffers from the problem of lack of aggregation of queries.
- Although it ensures privacy by using Paillier cryptosystem, it does not lower the costs incurred by the customers of the cloud.

2) COPS protocol

Cooperative Private Searching in Clouds
Publication: IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS
Author : Q. Liu, C. C. Tan, J. Wu, and G. Wang

COPS (Cooperative private searching) protocol was proposed by Qin Liu et al. in [3]. It introduced an aggregation and distribution layer (ADL) between the users and the cloud. ADL responsibilities include aggregation of queries from the users and distribution of results from the cloud to the users.

If the user's queries contain common keywords, it leads to lowered cost since the queries are aggregated by the ADL. Even in the scenario of no common keywords among users" queries, the merging of queries helps in considerably lowered

number of round trips to the cloud, thereby lowering the overall costs.

Following Stapes Shows working of COPS protocol

Step 1. Individual users generate a query using QueryGen algorithm. The query might look as {0,1,1,0,0,0…} where 1 and 0 have the same meaning as in Ostrovsky protocol. Note that the individual user queries are not encrypted before they are submitted to the ADL.

Step 2. The ADL runs the QueryMerge algorithm to merge all the user queries and sends a combined encrypted request to the cloud. Encryption is performed using the public key of the organization.

Step 3. The cloud runs the PrivateSearch algorithm to find files matching the combined query. The file survival rate in this algorithm is based on the number of mapping times γ and the buffer size β. It sends two encrypted buffers to the ADL, namely file pseudonym buffer and file content buffer. The pseudonym buffer consists of file names replaced with file pseudonyms.

Step 4. The ADL runs the ResultDivide algorithm to distribute appropriate files to each user.

Step 5. Individual users run FileRecover algorithm to retrieve matched files.

Issues in COPS Protocol

- COPS protocol can send too many results leading to excessive CPU consumption on the cloud. A lot of network bandwidth is required for transferring the response buffers from the cloud.
  ADL introduces delays and the response time is impacted. For users with lower tolerance for delays, multiple ADLs could be deployed within the organizational boundary.

3) EIRQ protocol

Towards Differential Query Services in Cost-Efficient Clouds
Publication: IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS
Authors: Qin Liu, Chiu C. Tan, Jie Wu

There are different query services are introduced in[4]. Where users send the queries to the cloud and cloud processes queries generates results process and sends it to users. In this case lot of files is matched users query. But the users are interested on certain percentage of files.

In the EIRQ model contains the cloud, organization and ADL. ADL is placed inside the organization based on requirement of users. Assume an organization have two users. They are Alice and Bob. They want files from the cloud. The Alice and Bob want files which are starts with the letters J, K and J, N respectively. The design goals of this scheme are Cost Efficiency and User Privacy. These goals are achieved by using Bloom Filters.

EIRQ Scheme: The EIRQ scheme is a process of recover the files from cloud to clients. This process has the following steps:

1) The EIRQ Scheme having the user and cloud [4]. The users are only authorized from the cloud network, and then only accessing is possible otherwise it is not possible.
2) This process is going on both wired network and wireless network also. First user sends request to ADL for establishment of a connection form the ADL. Then authorized user should have own login name and passwords.
3) After login user generates a query. This query is encrypted into 0's and 1's and then sends to ADL. the ADL runs Matrix Construct Algorithm [4] based on that Keywords and Ranks. This process is called as Aggregation.
4) After the aggregation process, ADL sends the Mask Matrix to Cloud. the cloud runs File Filter Algorithm. This algorithm filter out the files based on the Ranks and keywords. And sends encrypted buffer to ADL
5) The ADL runs the ResultDivide algorithm to distribute appropriate files to each user.
6) Individual users run FileRecover algorithm to retrieve matched files.

Issues in EIRQ Protocol

- Different users might have different tolerance to delays introduced by the ADL. A potential solution mentioned in the supplemental file pertaining to [4] is the provision of a timer value along with the user query. User i could send a timer value of Ti along with the query. ADL would wait no longer than the shortest time Ti of all the queries. At the expiry of shortest time Ti, ADL generates the mask matrix of all queries received so far and sends it to the cloud.

4) Comparison

There are following differences in the Ostrovsky scheme, COPS protocol and EIRQ scheme with respect to various parameters such as security, computational and communicational cost.

TABLE I. COMPARISON

Protocol	Parameter		
	Security and privacy	***Computational Cost***	***Bandwidth Cost***
Otrovsky	Yes	No	No
COPS protocol	Yes	Yes, to some extent	Yes, to some extent
EIRQ	Yes	Yes	Yes

IV. CONCLUSIONS

Cloud computing is used for sharing and retrieving information. In this paper we present different Techniques for searching over outsourced encrypted data. This study concludes that rank based retrieval is most efficient for searching on encrypted data because it is more secure, fast search access and does not leak information to untrusted authorities. However, while retrieving information from cloud environment it is necessary to get desired information with optimal communication and computation cost. In this paper, we have analyzed various algorithms which is used for efficient information retrieval in cloud environment. We have also shown the comparison of these algorithms which is useful for better understanding of these algorithms in terms of different parameters.

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," ACM Transactions on Information and System Security, 2009.

[3] Q. Liu, C. Tan, J. Wu, and G. Wang, ''Cooperative Private Searching in Clouds,'' J. Parallel Distrib. Comput. , vol. 72, no. 8, pp. 1019-1031, Aug. 2012.

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

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

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

[7] G. Danezis and C. Diaz, "Improving the decoding efficiency of private search," in IACR Eprint archive number 024, 2006.

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

[9] A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill, "Order-preserving symmetric encryption," Advances in Cryptology-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 Proceedings IEEE April 2011..

[11] W. Wong, D. Cheung, B. Kao, and N.Mamoulis, "Secure knn computation on encrypted databases," in Proc. of 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.

[13] P. Paillier, ''Public-Key Cryptosystems Based on Composite Degree Residuosity Classes,'' in Proc. EUROCRYPT, 1999, pp. 223-238.

[14] V. Anand, Ahmed Abdul Moiz Qyser, " A comparative study of secure search protocols in pay-as-you-go clouds", International Journal of Research in Engineering and Technology, Volume: 03 Special Issue: 05 , May-2014

[15] Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou,"Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, SYSTEMS, VOL. 23, NO. 8.