



ANALYSIS ON EFFICIENT INFORMATION RETRIEVAL FOR RANKED QUERY IN A COST EFFECTIVE OF CLOUD

Nagarapu Madhuri¹, Gude Suchitra²

¹M.Tech Student, ²Asst.Professor, Dept of CSE

V.S.Lakshmi Engg College for Womens, Matlapalem, Kakinada, (India)

ABSTRACT

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. In this paper, we present a scheme, termed efficient information retrieval for ranked query (EIRQ), based on an aggregation and distribution layer (ADL), to reduce querying overhead incurred on the cloud. In EIRQ, queries are classified into multiple ranks, where a higher ranked query can retrieve a higher percentage of matched files. A user can retrieve files on demand by choosing queries of different ranks. This feature is useful when there are a large number of matched files, but the user only needs a small subset of them. Under different parameter settings, extensive evaluations have been conducted on both analytical models and on a real cloud environment, in order to examine the effectiveness of our schemes.

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

I.INTRODUCTION

Cloud computing technology is a most necessary technology for information technology. Cloud computing as an emerging technology is expected to reshape information technology processes in the near future. Due to the overwhelming merits of cloud computing, e.g., cost-effectiveness, flexibility and scalability, more and more organizations choose to outsource their data for sharing in the cloud. As a typical cloud application, an organization subscribes the cloud services and authorizes its staff to share files in the cloud. Each file is described by a set of keywords, and the staff, as authorized users, can retrieve files of their interests by querying the cloud with certain keywords. In such an environment, how to protect user privacy from the cloud, which is a third party outside the security boundary of the organization, becomes a key problem. User privacy can be classified into search privacy and access privacy. Search privacy means that the cloud knows nothing about what the user is searching for, and access privacy means that the cloud knows nothing about which files are returned to the user. When the files are stored in the clear forms, a naive solution to protect user privacy is for the user to request all of the files from the cloud; this way, the cloud cannot know which files the user is really interested in. While this does provide the necessary privacy, the communication cost is high. Private searching was proposed by Ostrovsky. which allows a user to retrieve files of interest from an untrusted server without leaking any information. However, the Ostrovsky scheme has a high computational cost, since it requires the cloud to process the query (perform homomorphism encryption) on every file in a collection. Otherwise, the cloud will learn that certain files, without processing, are of no interest to the user. It will quickly become a performance bottleneck when the cloud needs to process thousands of queries over a collection of hundreds of



thousands of files. We argue that subsequently proposed improvements also have the same drawback. Commercial clouds follow a pay-as-you-go model, where the customer is billed for different operations such as bandwidth, CPU time, and so on. Solutions that incur excessive computation and communication costs are unacceptable to customers. To make private searching applicable in a cloud environment, our previous work designed a cooperate private searching protocol (COPS), where a proxy server, called the aggregation and distribution layer (ADL), is introduced between the users and the cloud. The ADL deployed inside an organization has two main functionalities: aggregating user queries and distributing search results. Under the ADL, the computation cost incurred on the cloud can be largely reduced, since the cloud only needs to execute a combined query once, no matter how many users are executing queries. Furthermore, the communication cost incurred on the cloud will also be reduced, since files shared by the users need to be returned only once. Most importantly, by using a series of secure functions, COPS can protect user privacy from the ADL, the cloud, and other users. In this paper, we introduce a novel concept, differential query services, to COPS, where the users are allowed to personally decide how many matched files will be returned. This is motivated by the fact that under certain cases, there are a lot of files matching a user's query, but the user is interested in only a certain percentage of matched files. In this paper we propose a scheme, termed Efficient Information retrieval for Ranked Query (EIRQ), in which each user can choose the rank of his query to determine the percentage of matched files to be returned. The basic idea of EIRQ is to construct a privacy-preserving mask matrix that allows the cloud to filter out a certain percentage of matched files before returning to the ADL. This is not a trivial work, since the cloud needs to correctly filter out files according to the rank of queries without knowing anything about user privacy. Focusing on different design goals, we provide two extensions: the first extension emphasizes simplicity by requiring the least amount of modifications from the Ostrovsky scheme, and the second extension emphasizes privacy by leaking the least amount of information to the cloud.

Our key contributions are as follows:

1. We propose three EIRQ schemes based on the ADL to provide a cost-efficient solution for private searching in cloud computing.
2. The EIRQ schemes can protect user privacy while providing a differential query service that allows each user to retrieve matched files on demand.
3. We provide two solutions to adjust related parameters; one is based on the Ostrovsky scheme, and the other is based on Bloom filters.
4. Extensive experiments were performed using a combination of simulations and real cloud deployments to validate our schemes.

II. RELATED WORK

Many searching techniques over encrypted cloud data have proposed. S. Deshpande [7] suggested a technique searching over encrypted cloud data using fuzzy keywords. They used Edit distance to quantify keyword similarity and developed two techniques on constructing fuzzy keyword sets to achieve optimized storage and representation overheads. Cong wang et al. [1] Has proposed a method ranked keyword search over encrypted cloud data using keyword frequency and order preserving encryption. It supports only single keywords at a time. Is the keyword frequency deciding document file score. Rank given to every file based on the relevance score of that file. Top ranked files have sent to users instead all files. To enrich search functionality N. Cao et al. [2] Have proposed a scheme supporting conjunctive keywords search. It is privacy – preserving multi-keyword ranked search technique using symmetric encryption. M. Chou et al. [6] proposed a solution for fuzzy multi-keyword search over encrypted cloud data using privacy aware Bed Tree. They used a co-occurrence probability approach to identify useful multi-keywords for publishing data, documents and relevant fuzzy keyword sets constructed using edit distance. They constructed index tree for all data, documents, where each leaf node having the hash value of a keyword, one or two data vectors that represents n- gram of that keyword and bloom filters for each edit distance value. C. Wang et al. [3] Suggested a technique to build storage efficient similarity keyword set with a given document collection, edit distance as a similarity metric. Based on that, they built a private trie traverse-searching the index to achieve similarity search functionality with constant time complexity.

C. Yang et al. [4] Designed a scheme adopting three sparse matrices instead dense matrix pair in MRSE to encrypt index, they combined their scheme with a bloom filter to gain the ability for index updating. D.X. Song, D et al. [8] Proposed a method, remotely searching over encrypted data using an untrusted server and provided proof for the resulting crypto system. They supported controlled, hidden search and query isolation. C. Wang et al. [10] Proposed an efficient Ranked Searchable Symmetric Encryption scheme (RSSE). It supports ranked keyword search, security guarantee and efficiency with minimum communication cost. B. Wang et al. [11] suggested a novel multi-keyword fuzzy search scheme by exploiting the locality sensitive hashing techniques. They achieved fuzzy matching through algorithmic design rather than expanding index file and it eliminates the need of predefined dictionary. N. Cao et al. [12] proposed two Multi - keyword Ranked Search over Encrypted cloud data (MRSE) schemes based on a similarity measure coordinate matching while meeting different privacy requirements in two different threat models. They enhanced ranked search mechanism to support dynamic data operations. In no above techniques supported semantic search and synonym-based search. In our work, we designed an efficient, secure multi-keyword synonym ranked searching technique over encrypted cloud data by designing BMS tree and DFST algorithm.

III. ARCHITECTURE

Co-operate searching protocol (cops) is like a proxy server called as aggregation and distribution layer (ADL) is placed inside an organization. This ADL is act as a mediator between the cloud and an organization. The functioning of ADL is the aggregation and distribution. The ADL only reduces the computation cost

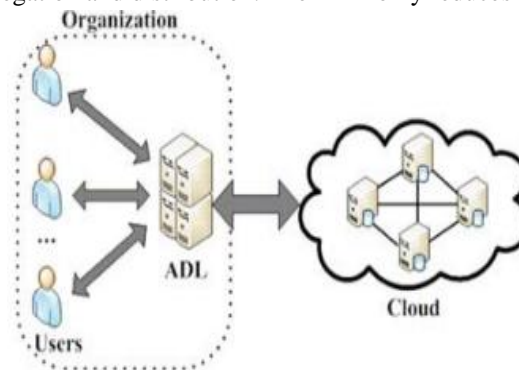


Fig.1. Architecture of EIRQ

The working of an ADL is the many users can send many queries to ADL. Then adl can aggregate the different user" s queries makes into a single query and then sends to cloud. The cloud will process the query sends response to ADL. Then the adl will distribute the results to particular users. Because of this process to reduce the communication cost and query overhead.

Here introduce a major concept differential query services. Where users are sends the queries to the cloud and process the query sends results to users. Lot of files is matched users query. But the user doesn" t want that files, only they interested on certain percentage of files. In the proposed model have the cloud, organization and ADL. ADL is placed inside the organization based on requirement of number users. In this model used only single ADL inside an organization. Assume an organization have two users. They are Jack and Jan. They want files from the cloud. The Jack and Jan 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. We achieve these goals by using Bloom Filters.

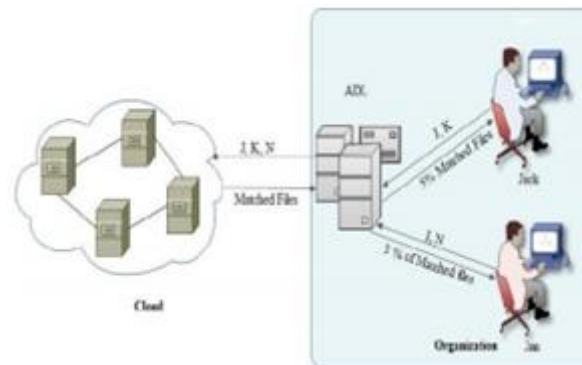


Fig.2. EIRQ Model

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 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 send request from the user to cloud for establishment of a connection from the cloud. Then authorized user should have their own login name and passwords.
3. After login to user generate a query. 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 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.

IV.SYSTEM MODEL

In this paper, we considered a cloud computing system model involves three different entities. Those are Data Owner, Cloud Service Provider and Data user as illustrated in Fig. 1. The responsibility of each entity is as follows: Data Owner (DO): DO has a collection data documents $DC = \{d1, d2, \dots, dm\}$ with sensitive information to be outsourced to the cloud server. To provide data privacy, the documents are encrypted before outsourcing. DO creates a dictionary based on keywords extracted from the all m documents based on Term Frequency Inverted Document Frequency (TFIDF) [13] which is described in section 4. The dictionary includes synonyms of each keyword from the thesaurus [14]. The dictionary is having n keywords, and for each keyword may have t synonyms, so that the dictionary size is $n \times t$. DO creates an index vector for each document based on the keywords extracted from the document. The size of the index vector is equal to the number of keywords in the dictionary that is n . Each dimension in the index vector stores sum of the frequency of keyword and corresponding synonyms in the dictionary is denoted as term frequency (TF) in our system. Index vectors of all documents are encrypted before outsource to the cloud. DO create query vector based on keywords entered by Data user. To provide user privacy, query vector encrypted, as Trapdoor and send to Data user. The data owner sends search access control to the authorized data user.

Data users: Data users are the users who accessing sensitive data from the cloud. The cloud server searches keywords or synonyms related to documents, which are interested to data user and sends to the data owner. The data user receives trapdoor and searches access control of data owner and sends trapdoor and access control to the cloud server to retrieve required documents from the cloud.

Cloud Service Provider (CSP): Cloud server receives encrypted documents and encrypted index vectors from data owner and stores into data owner's cloud storage. Cloud server having the capability to take the data request from user and check the search access control of the user. It will retrieve the documents from cloud storage depending upon the privileges to access number of documents. To increase the document retrieval accuracy from cloud server, the top scored (ranked) documents return to data user from the cloud server. Fig. 1 shows the architecture of multi-keyword synonym query over encrypted cloud data.

Threat model: The cloud server is measured as “honest-but-curious” [15] in our proposed scheme. The cloud server follows the proposed method specification and also examines data in its cloud storage and data which are received from data user during the processing to learn additional information. We consider one threat model for our system with different attack capabilities that is as follows: Known ciphertext model: In this model, the cloud server knows only encrypted documents and encrypted index vectors, which are outsourced from data owner.

V. CONCLUSION



We propose three EIRQ schemes (EIRQ Simple, EIRQ Privacy, and EIRQ Efficient) are worked through ADL. It offers differential query services, which will also protect the user privacy. These schemes are provide, clients are recovered certain percentage of matched records by particular queries of various ranks. Private searching technique is used to cost efficient cloud environments. In our EIRQ scheme assign ranks for each query, then highest rank files are matched and user recovered certain percentage of matched files. However, in the EIRQ schemes, we simply determine the rank of each file by the highest rank of queries it matches. For our future work, we will try to design a flexible ranking mechanism for the EIRQ schemes.

REFERENCES

- [1] Qin Liu, Chiu C. Tan, Member, IEEE, Jie Wu, Fellow, IEEE, and Guojun Wang, Member, IEEE, “Towards Differential Query Services in Cost-Efficient Clouds”, IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 6, June 2014.
- [2] P. Mell and T. Grance, “The NIST Definition of Cloud Computing (Draft)”, in NIST Special Publication. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2011.
- [3] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions”, in Proc. ACM CCS, 2006, pp. 79-88. [4] R. Ostrovsky and W. Skeith, “Private Searching on Streaming Data”, in Proc. CRYPTO, 2005, pp. 233-240.
- [5] R. Ostrovsky and W. Skeith, “Private Searching on Streaming Data”, J. Cryptol., vol. 20, no. 4, pp. 397-430, Oct. 2007.
- [6] J. Bethencourt, D. Song, and B. Waters, “New Constructions and Practical Applications for Private Stream Searching”, in Proc. IEEE SP, 2006, pp. 1-6.
- [7] J. Bethencourt, D. Song, and B. Waters, “New Techniques for Private Stream Searching”, ACM Trans. Inf. Syst. Security, vol. 12, no. 3, p. 16, Jan. 2009.
- [8] 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.
- [9] G. Danezis and C. Diaz, “Improving the Decoding Efficiency of Private Search”, Int’l Assoc. Cryptol. Res., IACR Eprint Archive No. 024, Schloss Dagstuhl, Germany, 2006.
- [10] G. Danezis and C. Diaz, “Space-Efficient Private Search with Applications to Rate less Codes”, in Proc. Financial Cryptogr. Data Security, 2007, pp. 148-162.
- [11] M. Finiasz and K. Ramchandran, “Private Stream Search at the Same Communication Cost as a Regular Search: Role of LDPC Codes”, in Proc. IEEE ISIT, 2012, pp. 2556- 2560.
- [12] X. Yi and E. Bertino, “Private Searching for Single and Conjunctive Keywords on Streaming Data”, in Proc. ACM Workshop Privacy Electron. Soc., 2011, pp. 153-158.
- [13] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes”, In Proc. of EUROCRYPT, 1999.
- [14] Q. Liu, C. C. Tan, J. Wu, G. Wang, “Efficient information retrieval for ranked queries in cost-effective cloud environments”, in Proc. of IEEE INFOCOM, 2012.
- [15] S. Yu, C. Wang, K. Ren, W. Lou, “Achieving secure, scalable, and fine-grained data access control in cloud computing”, in Proc. of IEEE INFOCOM, 2010.



AUTHORS PROFILE:

	<p>NAGARAPU MADHURI is a student of V.S.LAKSHMI ENGINEERING COLLEGE FOR WOMENS. Presently he is pursuing M.Tech [Computer Science and Engineering] from this college and he also completed his B.Tech .</p>
	<p>Mrs. G.SUCHITRA, working as a Asst.Professor in the Dept.of Computer Science and Engineering from V.S.Lakshmi Engineering College, Matlapalem, Kakinada.</p>