iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3010

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Achieving Efficient Information Retrieval Query over Encrypted Cloud Data

Rachana Mudholkar [1], Chetan Patil[2], Himanshu Kumar[3], Kishlay Kishan[4], Shubhang Magotra[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, Dr. D. Y. Patil Institute of Engineering, Management & Research, Akurdi, Pune, India*

Abstract: Cloud computing is an emerging technology which reshapes the advance in information technology. Cloud computing faces two type of issues i.e. privacy efficiency based on the aggregation and distribution layer we provide an extension to the Ostrovsky scheme by OIRQ i.e. (optimized information retrieval query) OIRQ is the technology that provides optimized information query through matrix construct. To further reduce querying cost incurred on the cloud. Query are classified into multiple ranked where higher ranked query can retrieve the higher percentage of matched files.
Keywords: Multiple-keyword ranked search, dynamic update, cloud computing, searchable encryption.

I. INTRODUCTION

Cloud computing as an emerging technology is expected to reshape the information technology processes in the near future. As there are lot of benefits of cloud computing e.g. cost-effectiveness, flexibility and scalability, more and more organizations choose to outsource their data for sharing in the cloud. In a typical cloud usage scenario, an organization subscribes the cloud services and authorizes its staff to share files in the cloud. All the files on cloud are 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 where keywords are used, how to protect *user privacy* from the cloud, which is a third party outside the security boundary of the organization, becomes a key problem. Other than this there are various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers brings privacy concerns[1]. The service providers (CSPs) that keeps the data for users may access users' sensitive information without authorization. The approach that is used to protect the data confidentiality is to encrypt the data before outsourcing. This will have a huge impact on cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data that cannot be directly applied on the encrypted data. Downloading all the data from the cloud data and decrypt locally is obviously impractical However, these methods are not practical due to their high computational overhead for both the cloud sever and user [2]. On the opposite hand there area unit more sensible special-purpose solutions, like searchable encoding (SE) schemes have created specific contributions in terms of potency, practicality and security.

Searchable encoding schemes change the consumer to store the encrypted knowledge to the cloud and execute keyword search over cipher text domain. Abundant works are dead underneath totally different threat models to
realize varied search practicality, like single keyword search, similarity search, multi-keyword Boolean search, graded search, multi-keyword graded searched, multi keyword ranked search achieves more and more attention for its practical applicability. Some dynamic schemes have been executed to support inserting and deleting operations on document collections [3]. These area unit vital works as it's extremely potential that knowledge homeowners ought to update their data on the cloud server few of the dynamic schemes support economical multi-keyword hierarchic search. User privacy may be classified into search privacy and access privacy.

Search privacy means the cloud is aware of nothing regarding what the user is checking out, and access privacy means the cloud is aware of nothing regarding that files area unit came
to the user. When the files area unit keep within the clear forms, a naive answer to safeguard 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. [4]. While this will give the required privacy, the communication price is high. 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 therefore the cloud. The ADL deployed within a corporation has 2 main functionalities: aggregating user queries and distributing search results.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

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[5][9].
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 significantly, by employing a series of secure functions, COPS will defend user privacy from the ADL, the cloud, and different users.

II. RELATED WORK

A number of ways are planned in recent years to produce user privacy and additionally concerning non-public looking schemes on streaming information (2005).In this paper, R. Ostrovsky and W. Skeith[1] thought-about the matter of personal looking on streaming information. He showed that during this model we will with efficiency implement looking for information or documents beneath a secret criteria (such as presence or absence of a hidden combination of hidden keywords) beneath numerous cryptologic assumptions. The results are often viewed in an exceedingly form of ways: as a generalization of the notion of a non-public data Retrieval as positive results on privacy-preserving datamining; and as a delegation of hidden program computation. In searchable regular encoding (2006) permits a celebration to outsources the storage of his information or documents to a different party in an exceedingly non-public manner, whereas maintaining the power to by selection search over it. This drawback has been the main target of active analysis and a number of other security definitions and constructions are planned. during this paper we start by reviewing existing notions of security and propose new and stronger security definitions. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky[2] given 2 constructions that show secure beneath new definitions. curiously, additionally to satisfying stronger security guarantees, the new constructions square measure a lot of economical than all previous constructions. Further, previous work on sou'-sou'-east solely thought-about the setting wherever solely the owner of the information or files is capable of submitting search queries. They additionally take into account the natural extension wherever AN arbitrary cluster of parties apart from the owner will submit search queries. The sou'-sou'-east is formally outlined during this multi-user setting, ANd presents an economical construction. economical data retrieval for hierarchical queries in cost-efficient cloud environments (2012).Cloud computing as AN rising technology trend is anticipated to reshape the advances in data technology. during this paper, it addresses 2 basic problems in an exceedingly cloud environment: privacy and potency. Here initial review a non-public keyword-based file retrieval theme planned by Ostrovsky et.[3]. Then, supported AN aggregation and distribution layer (ADL), given a theme, termed economical data retrieval for hierarchical question (EIRQ), to any cut back querying prices incurred within the cloud. Queries square measure classified into multiple ranks, wherever the next hierarchical queries will retrieve the next proportion of matched files. in depth evaluations are conducted on AN analytical model to look at the effectiveness of this theme. Searchable Encryption: ancient searchable encoding has been wide studied as a cryptologic primitive, with a spotlight on security definition formalizations and potency enhancements. Songet al. [4] initial introduced the notion of searchable encoding. They planned a theme within the regular key setting, wherever every word within the file is encrypted severally beneath a special two-layered encoding construction. Thus, a looking overhead is linear to the complete file assortment length. Goh [5] developed a Bloom filter primarily based per-file index, reducing the work load for every search request proportional to the amount of files within the assortment. Chang Jiang et al. [6] additionally developed an analogous per-file index theme. To any enhance search potency, Curtmola et al. [7] planned a per-keyword primarily based approach, wherever one encrypted hash table index is constructed for the whole file assortment, with every entry consisting of the trapdoor of a keyword ANd an encrypted set of connected file identifiers. Searchable encoding has additionally been thought-about within the public key setting. Boneh et al. [8] given the primary public-key primarily based searchable encoding theme, with a similar state of affairs to it of [8]. In their construction, anyone with the general public key will write to the information hold on on the server however solely approved users with the non-public key will search. Recently, aiming at tolerance of both minor types and format in consistencies in the user search input, fuzzy keyword search over encrypted cloud data Note that all these schemes support only Boolean keyword search, and none of them support the ranked search problem which we are focusing in this paper propose a privacy-preserving multi-keyword ranked search scheme, which extends our previous work in [9] with support of multi-keyword query. They choose the principle of “coordinate matching”, i.e. as many matches as possible, to capture the similarity between a multi keyword search query and data document, and later quantitatively formalize the principle by a secure inner product computation mechanism. One disadvantage of the scheme is that cloud server has to linearly traverse the whole index of all the documents for each search request.

III. COMPARISON TABLE

A. Existing System

SR.NO	NAME OF KEYWORD	PRECISION BEFORE UPLOADING	PRECISION AFTER UPLOADING
1	JAVA	88%	100%
2	PAYTHON	79%	100%
3	.NET	95%	100%
4	JEE	65%	100%

B. Comparison Of Existing System With Proposed System

	Searchable encryption	Ranked searchable encryption
Efficiency	Low	High
Accuracy	Low	High
Data Search	Returns non relevant Information	returns more relevant information
Traffic	High	Low
Security	Low	High

IV. METHODLOGY

Cloud computing as an emerging technology is expected to reshape the information technology processes in the near future. Using cloud service i.e. SAAS (software as a service) we proposed the solution for the existing system. Existing system proposes issue such as major privacy issue, loss of data accuracy that in increase computation time and reducing random data in the proposed system we provide methodology that give secure and multi key word ranked search over encrypted data this is provide by the matrix construct by Optimized Information Retrieval Query. This method overcomes the drawback of the existing system. ORIQ (optimized information retrieval query) mainly perform three major function i.e. aggregation, filtration, and distribution. The methodologies that are being used to overcome the drawback of the existing systems are Admin panel, Ostrovsky's scheme and the Optimized Information Retrieval Query (OIRQ) i.e. proposed. Via using this technology there is a possibility to avoid the disadvantages o the existing system.

A. Disadvantages Of Existing System

As per the survey conducted it is found that the existing system proposes the following disadvantages.

Existing system is less accurate. The data that is asked by the user has accuracy issue. Along with the accurate data the user is also provided with some inaccurate material. The existing system also faces the difficulty like loss of data. In the scenario of providing accurate information to the user the existing system loses some accurate data too. Another issue that faced by the existing system is Privacy. This is one of the major factors that is important while retrieving information from the cloud data. Privacy of the user and the information that is to be retrieved is to maintained. The existing system also faces the problem of Realizing of Random Data. When the user requests for a specific amount of data he/she is provided with a random data. For example if the user requests for 88% of data from the cloud data the user is provided with 100% of the data.

B. Solution For Existing System

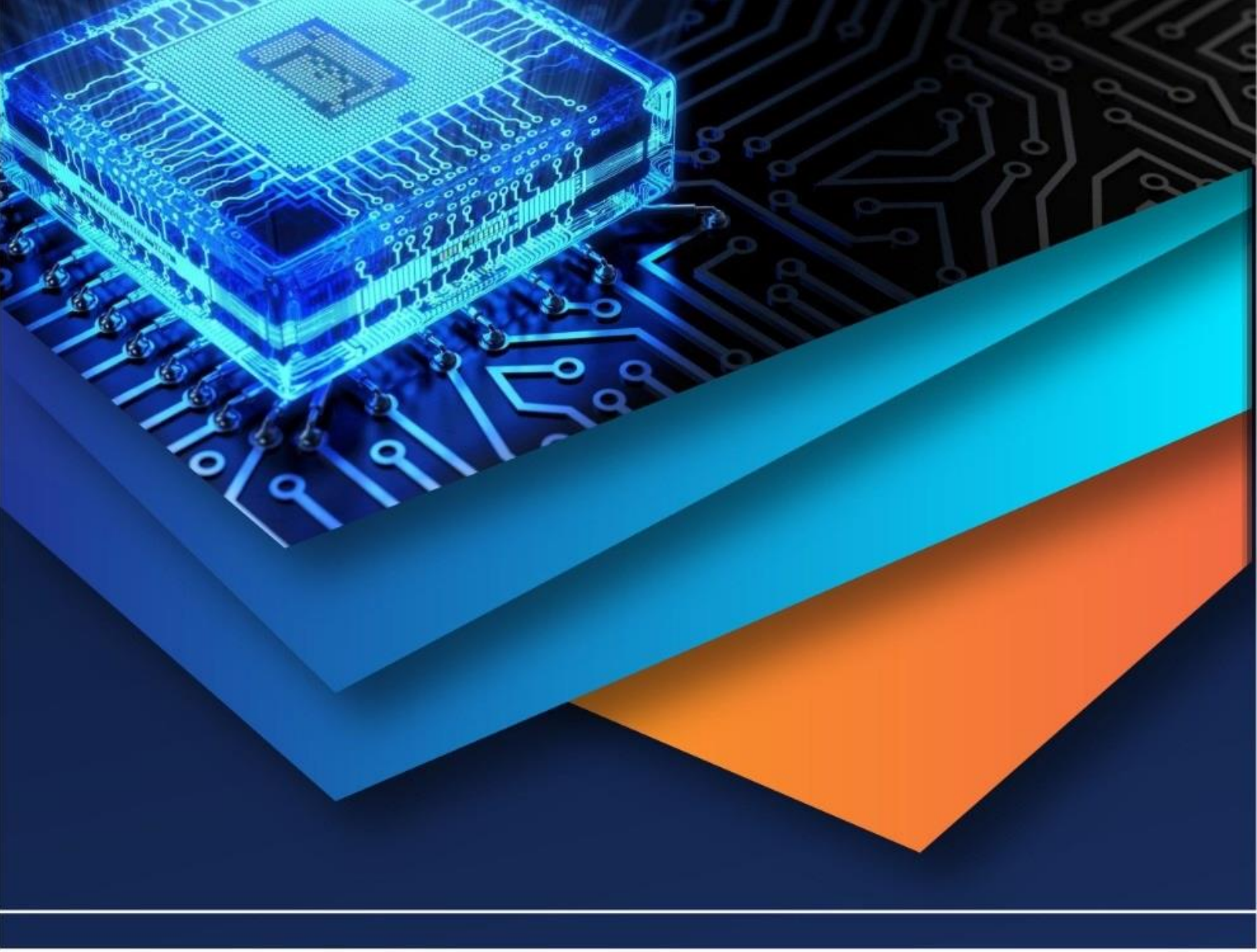
Here the new proposed scheme is called Optimized Information Retrieval Query is introduced. This new system uses the mechanism of flexible ranking mechanism which allows user to provide a rank and can personally decide how many matched files will cloud return. The basic idea is to construct a matrix that allows the cloud to filter out certain percentage of matched files. This new scheme reduces the querying overhead and also computational cost. The OIRQ system protects user privacy which allows each user to retrieve matched file on demand. This is not an easy work because the cloud needs to correctly filter up files according to rank of the query without knowing anything about user privacy. It has two extensions. The first extension shows the least amount of modification from the Ostrovsky's scheme and the second extension provide privacy by leaking the least amount of information to the client.

V. CONCLUSIONS

The survey done it can be concluded that the OIRQ scheme based on ADL to provide differential query services while protecting user privacy by using our scheme user can retrieve different percentage of matched file by specifying query of different ranks. By further reducing the communication cost incurred on the cloud, the OIRQ makes the private searching techniques more applicable to a cost efficient environment. However OIRQ scheme we simply determine the rank of each file by the highest rank of queries matches. For our future work, we will try to design a flexible ranking mechanism for the OIRQ schemes.

REFERENCES

[1] R. Ostrovsky and W. Skeith, "Private searching on streaming data," in Proc. of CRYPTOLOGY 2005.
[2] R. Curtmola, J. Garay, S. Kamara, andR.Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. ofACM CCS, 2006.
[3] 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.
[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000.
[5] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, Report 2003/216, 2003.
[6] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS'06, 2006.
[7] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. of INFOCOM'11, 2011
[8] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY