International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Survey of Rank Based Search Results over Encrypted Cloud Data

Mithuna .R[1], Suguna .M[2]

[1]PG Student, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore

[2]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore

Abstract: ***Migration of users into the cloud environment is increased with the high popularity of services provided by the cloud providers. When large number of users outsource their files into the cloud environment privacy becomes the most important issue. In order to assure privacy users encrypt their files before outsourcing it to the cloud server. The efficient searching and effective retrieval of files becomes most complex when the files are stored in the encrypted format. In the previous work Multi-Keyword Ranked Search over Encrypted (MRSE) cloud data is implemented to assure privacy enhanced searching mechanism. The ranking mechanism is used to retrieve the most similar files over the encrypted files. However one cannot assure that whether all retrieved results are having similar fields. This work focuses on implementing the rank test method to check the integrity of the rank order (i.e.) to find out the resulting files are having similar fields or not. Our proposed system is used to retrieve the files with the most similarity values and hence the privacy is improved.***

Keywords: Cloud computing, multi keyword search, privacy preserving, ranked search, encrypted file

1. Introduction

Cloud computing holds a promise to deliver large-scale utility computing services to a wide range of consumers. Cloud computing (Figure 1) can be defined as "a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned, and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers". Some of the examples for emerging Cloud computing infrastructures/platforms are Microsoft Azure, Amazon EC2, Google App Engine, and Aneka.

Cloud computing delivers infrastructure, platform, and software that are made available as subscription-based services in a pay-as-you-go model to consumers. These services are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) in industries. The importance of these services was highlighted in a recent report from the University of Berkeley as: "Cloud computing, the long-held a service".

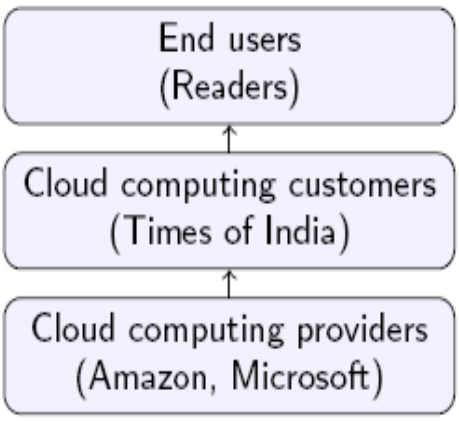

Figure 1: Cloud computing overview

In cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. To protect data privacy and unauthorized access data owners may encrypt their files before outsourcing to cloud server [1]. Downloading all the files and decrypting it locally is a complex task. Data stored in cloud should be efficiently searched and retrieved by the users. One of the most popular ways to do so is through keyword-based retrieval.

Keyword-based retrieval is a typical data service and widely applied in plaintext scenarios, in which users retrieve relevant files from the file set based on keywords [2]. However, it turns out to be a difficult task in cipher texts scenario due to limited operations on encrypted data. Besides, to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to users [3].

Cloud server must perform result relevance ranking to meet the effective data retrieval instead of performing undifferentiated results. Ranked search also eliminates unnecessary network traffic by ranking the search results and decreases the use of bandwidth consumed. In order to improve the search results accuracy and to enhance user searching experience such ranking system should support multiple keyword search as single keyword search offers only relevant files. Each keyword in the request helps the cloud server to narrow down the results. Many different types of search methods have been proposed and used.

1.1 Background and Related Work

Many organizations, companies store more valuable data on cloud in order to outsource it to their users. The benefits of

the new computing model include but are not limited to: relief of the trouble for storage administration, data access, and avoidance of high expenditure on hardware mechanism, software, etc.

Ranked search improves system usability by matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency).

As directly outsourcing relevance scores will leak a lot of sensitive information against the keyword privacy, asymmetric encryption with ranking result of queried data which will give only expected data.

1.2 Secure Ranked Keyword Search over Encrypted Cloud Data

Traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional *Boolean* keyword search1, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following two main drawbacks. On the one hand, for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead; On the other hand, invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. In short, lacking of effective mechanisms to ensure the file retrieval accuracy is a significant drawback of existing searchable encryption schemes in the context of Cloud Computing. Nonetheless, the state-of-the-art in information retrieval (IR) community has already been utilizing various scoring mechanisms to quantify and rank-order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be addressed in the context of encrypted data search.

2. Practical Techniques for Searches on Encrypted Data

This paper shows how to support searching functionality without any loss of data confidentiality. An example is where a mobile user with limited bandwidth wants to retrieve all email containing the word "Urgent" from an untrusted mail-storage server in the infrastructure. This is trivial to do when the server knows the content of the data, but how can we support search queries if we do not wish to reveal all our email to the server?

The answer is to present cryptographic schemes that enable searching on encrypted data without leaking any information to the untrusted server.

The techniques are provably secure. The techniques provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext given only the ciphertext. The techniques provide controlled searching, so that the untrusted server cannot search for a word without the user's authorization. The techniques support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The techniques also support query isolation, meaning that the untrusted server learns nothing more than the search result about the plaintext.

The schemes are efficient and practical. The algorithms we present are simple and fast. More specifically, for a document of length n, the encryption and search algorithms only need O(n) number of stream cipher and block cipher operations. The schemes introduce essentially no space and communication overhead. They are also flexible and can be easily extended to support more advanced searches.

2.1. Existing system

Existing searchable encryption schemes allows user to securely search through the encrypted cloud data by using keywords. By allowing user to decrypt the files which match only their interest security and privacy is provided. However these search mechanisms support Boolean keyword search, single keyword search, fuzzy keyword search, conjunctive keyword search [4][5].Users are limited to searching experience in these type of searches.

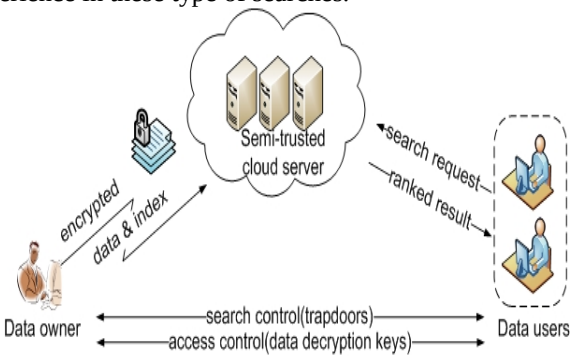

Figure 2 Architecture for Multi Keyword search

Multi keyword Ranked Search over Encrypted cloud data enables one to search through the cloud by providing multiple keywords of interest. This makes the user to get the ranked results that match their query [6]. However one cannot assure the correctness of the rank order. Ranked results are provided to users but the integrity of the rank order may not be known or checked to prove whether the files are having similar fields or not. The existing work cannot retrieve the most similar files which may violate the performance level.

3. Problem Formulation

3.1 Proposed System

For our system we choose the efficient similarity measure of coordinate matching i.e. as many matches as possible to capture the relevance of the data documents according to queried keywords. Specifically inner product similarity metrics is used to capture similarity scores between the query keywords and the words in the documents. Each document is linked with a binary vector as a sub index

where each bit represents whether corresponding keyword is contained in the document [9]. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by inner product of query vector with data vector. However, directly outsourcing data vector or query vector will violate index privacy or search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy issues Searchable Symmetric Encryption technique[7] is used.

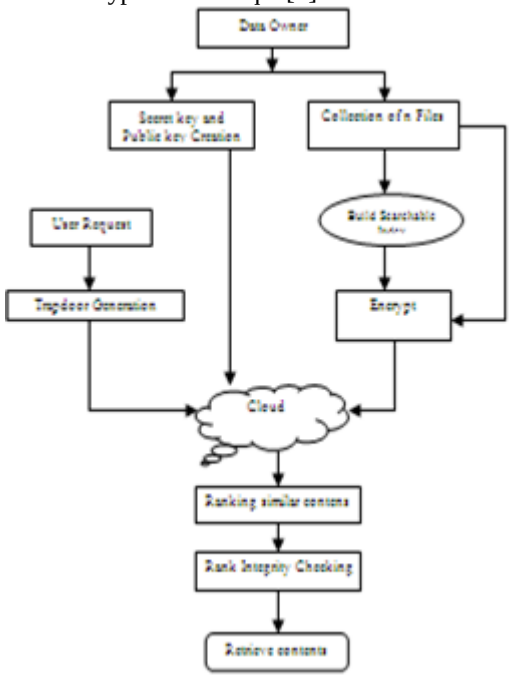

Figure 3: Architecture for Rank integrity checking over encrypted cloud data

MRSE framework is improved step by step to achieve various privacy requirements in two levels of threat models.
1) Showing the problem of Secured Multi-keyword search over encrypted cloud data
2) Propose two schemes following the principle of coordinate matching and inner product similarity

Consider three different entities, as illustrated in Fig3. Data owner, data user, and cloud server. Data owner has a collection of data documents to be outsourced to third party cloud server in the encrypted format. To activate the capability of searching over encrypted cloud data, data owner, before sending data, will first build an encrypted searchable index, and then outsource both the index and the encrypted document collection to cloud server. To search the document, an authorized user require a corresponding trapdoor through search mechanisms, upon receiving request from data users, cloud server is responsible to search the index and then returns the matched set of documents id to the cloud sever. Cloud server matches the documents id and then returns the set of documents. To improve document retrieval accuracy, cloud server must rank the search result according to some ranking criteria. In this work, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top-k multi keyword retrieval over encrypted cloud data with high security and practical efficiency.

3.2 Rank Test Method

Fried man rank test is implemented in this work to improve the search efficiency of the files. This mechanism is used to improve the search retrieval process by testing the rank fixed by the cloud servers. Whenever the user enters the multi keyword for file retrieval, the trapdoor will be generated. Through that trap door value, the cloud server will match the user's keyword query with the searchable index of the encrypted files and will retrieve the most similar files. In this process, in order to assure retrieved files are similar to the queries and the result is retrieved by checking all the files in the group, rank test mechanism is used. The Fried man rank test mechanism is implemented in this work which can ensure that all files from the group is ranked properly and only the similar files are retrieved. This is done by ranking each block in file and comparing with the other files.

Fried man rank test is calculated as follows:

$$F_R = \frac{12}{rc\,(c+1)} \sum_{j=1}^{c} R_j^2 - 3r(c+1)$$

Where,

R_j^2 =Square of the total of the ranks for group j (j=1, 2… c)

r = number of blocks
c = number of groups

3.3 Design Goals

To activate the integrity checking of the ranked search for effective utilization of outsourced cloud data under the aforementioned model, the system should be designed by considering the security considerations also. The system is expected to give the following security and performance guarantees as follows.

a) **Secured Multi-keyword Ranked Search** : To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning unmatched results.
b) **Privacy**: To prevent the cloud server from learning any information from the dataset and the index, and to meet the basic and improved privacy requirements by providing matched results.
c) **Effectiveness with high performance**: Ranked search provides low communication and computation overhead.

4. Conclusion

This study presents the survey of Multi keyword search for efficient file retrieval process**.** In this paper different types of search methods, their drawbacks and encryption techniques are studied. Various algorithms for ensuring privacy and efficiency have been studied. According to the available existing works multi keyword search greatly enhances security, privacy and efficiency. By using the rank test method integrity checking of the rank order is determined. It ensures files are ranked properly and only similar files are retrieved.

Reference

[1] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS Springer Heidelberg.
[2] Google, "Britney spears spelling correction," Referenced online at http: //www.google.com/jobs/britney.html, June 2009.
[3] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.
[4] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.
[5] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. of ACNS, pp. 31–45, 2004.
[6] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proc. of ICICS, 2005.
[7] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10, 2010.
[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in *Proc. of ACNS*, 2005.
[10] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

.