ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Survey on Keyword Search over Outsourced Cloud Data Using Ranked Searchable Symmetric Encryption

Deepa R, Reeja S L

M Tech Student, Dept of CSE, Marian Engineering College, Trivandrum, Kerala, India

Asst. Professor, Dept of CSE, Marian Engineering College, Trivandrum, Kerala, India

ABSTRACT: Cloud computing provide service over the internet. For the protection of data privacy sensitive information are encrypted before it out sourced to the public cloud. Searchable encryption scheme allow users to securely search over the encrypted data through keyword. This scheme does not capture any relevant information. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results. It improves the file retrieval accuracy. Secure searchable index are used for information retrieval. Ranked search over outsourced cloud data improves the efficiency of searching encrypted cloud data.

KEYWORDS: Ranked search, searchable encryption, confidential data, cloud computing

I. INTRODUCTION

Cloud computing provide reliable services delivered through data centres that are built on virtualized compute and storage technologies ,it becomes prevalent more sensitive information like government documents,emails,business information are centralized into cloud. Cloud storage is a service model in which data is maintained, managed and made available to users. For data privacy sensitive information are encrypted before it out sourced to the public cloud. In cloud computing data owners share their out sourced data to a large number of users. Keyword based search allows selectively retrieve the file of interest[3].Traditional searchable encryption schemes allow users to securely search over encrypted data through keywords, without capturing any relevance of data files[4].

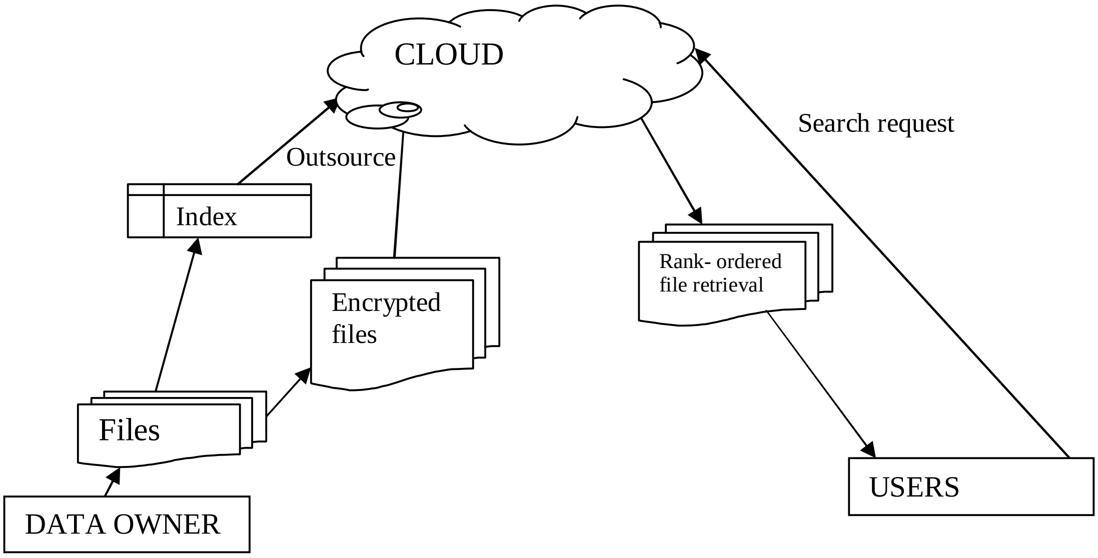

Fig: Architecture for search over encrypted cloud data

DOI: 10.15680/ijircce.2015.0306079

Figure shows architecture for search over encrypted cloud data. The architecture has three entities data owner, cloud server, end users. Data owner having the collection of files he out sourced the files in an encrypted format. User search the required file with help of keyword in secure manner, cloud server search the index file and return the corresponding list of files. Before outsourcing, data owner first create a secure searchable index from a set of *distinct* keywords extracted from the file collection, and store both the index *and* the encrypted file collection on the cloud server. Authentication between user and data owner is done. For the searching operation user create a secure search request in the form of trapdoor to the server. On receiving the search request server search on the index file and issue the corresponding files to the user. Ranked search greatly enhances system usability by returning the matching files in a ranked order with the help of some relevance criteria.

II. RELATED WORK

Searchable Encryption: Traditional searchable encryption has been widely studied as a cryptographic primitive, with a focus on security definition formalizations and efficiency improvements. Songet al. [5] first introduced the notion of searchable encryption. They proposed a scheme in the symmetric key setting, where each word in the file is encrypted independently under a special two-layered encryption construction. Thus, a searching overhead is linear to the whole file collection length. Goh [6] developed a Bloom filter based per-file index, reducing the work load for each search request proportional to the number of files in the collection. Chang et al. [10] also developed a similar per-file index scheme. To further enhance search efficiency, Curtmola et al. [7] proposed a per-keyword based approach, where a single encrypted hash table index is built for the entire file collection, with each entry consisting of the trapdoor of a keyword and an encrypted set of related file identifiers. Searchable encryption has also been considered in the public-key setting. Boneh et al. [8] presented the first public-key based searchable encryption scheme, with an analogous scenario to that of [8]. In their construction, anyone with the public key can write to the data stored on the server but only authorized users with the private key can 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 [11] 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 documents, 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, while ours is as efficient as existing SSE schemes with only constant search cost on cloud server .Secure top-k retrieval from Database Community [12] from database community are the most related work to our proposed RSSE.

The idea of uniformly distributing posting elements using an order-preserving cryptographic function. However, the order-preserving mapping function proposed does not support score dynamics, i.e., any insertion and updates of the scores in the index will result in the posting list completely rebuilt. Besides, when scores following different distributions need to be inserted, their score transformation function still needs to be rebuilt. On the contrary, in our scheme the score dynamics can be gracefully handled, which is an important benefit inherited from the original OPSE[13]. It uses a different order-preserving mapping based on pre-sampling and training of the relevance scores to be outsourced, which is not as efficient as our proposed schemes. This can be observed from the Linear Search procedure in Algorithm, where the same score will always be mapped to the same random-sized non-overlapping bucket, given the same encryption key. In other words, the newly changed scores will not affect previous mapped values. We note that supporting score dynamics, which can save quite a lot of computation overhead when file collection changes, is a significant advantage in our scheme. Moreover, both works above do not exhibit thorough security analysis which we do in the paper. Other Related Techniques Allowing range queries over encrypted data in the public key settings where advanced privacy preserving schemes were proposed to allow more sophisticated multi-attribute search over encrypted files while preserving the attributes’ privacy. Moreover, the two schemes do not support the ordered result listing on the server side. Thus, they cannot be effectively utilized in our scheme since the user still does not know which retrieved files would be the most relevant. Though these two schemes provide provably strong security, they are generally not efficient in our settings, as for a single search request, a full scan and expensive computation over the whole encrypted scores corresponding to the keyword posting list are required.

III. RECOMMENTATION TECHNIQUE

A. *Practical techniques for searches on encrypted data*

The scheme is based on sequential scan method. It is encryption algorithm provides provable secrecy. Then they show how the scheme can be extended to handle controlled searching and hidden searches. After considering a few criteria and scheme modifications, the authors come out with a complete version of PTSED scheme. PTSED consists of several steps: Pre-encryption, searching, and decryption. The purpose of the pre encryption first step is to hide the actual searching keyword and to prevent any unauthorized party which can excess the remote server using cryptanalysis to break the whole encrypted message after a few keyword searches. Before starting the searching algorithm, the user has to provide some information since the server will not learn anything more than what is provided by the user. After the server gathers the required information from the user, the searching algorithm will run based on the information gathered. In this case, the server may return the file to the end user if the keyword is match. Otherwise, it will continue to search until the end of the file. After the user search and retrieve the encrypted file containing the specific keyword, the final step is to decrypt the retrieved file back to plaintext [5]

B. *Secure Indexes*

The SI scheme was proposed by Goh. The scheme builds a secure index for documents. This secure index allows a user to search for an encrypted document that is containing a keyword without decrypting the document. A Bloom Filter (BF) is used as a per document index to keep track of each of the unique words. Before each of the unique keywords is indexed and stored into bloom filter objects, those unique keywords have to go through a pseudorandom function twice. The purpose of doing so is to make sure that for each two or more documents, if they contain the same keyword the codeword will represent it differently [6].

C. *Public key encryption with keyword search*

This paper describes problem of searching on data that is encrypted using a public key system. Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants ls to test whether the email contains the keyword “urgent" so that it could route the email accordingly. Alice, on the other hand does not wish to give the gateway the ability to decrypt all her messages. They define and construct a mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word “urgent" is a keyword in the email without learning anything else about the email. This mechanism as Public Key Encryption with keyword Search [8].

D. *Searchable symmetric encryption*

Searchable symmetric encryption (SSE) allows a party to outsource the storage of its data to another party (a server) in a private manner, while maintaining the ability to selectively search over it. Private-key storage outsourcing allows clients with either limited resources or limited expertise to store and distribute large amounts of symmetrically encrypted data at low cost. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. To address this, several techniques have been proposed for provisioning symmetric encryption with search capabilities the resulting construct is typically called searchable encryption [7]. This leads to following drawbacks,
1) Non relevant data search result
2) Large unnecessary network traffic, which is absolutely undesirable in today’s pay-as-you-use cloud paradigm. 3) Decrease the efficiency and File retrieval accuracy.

E. *Ranked Searchable Symmetric Encryption*

Cloud Computing ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our design goals on both system security and usability, we propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption (RSSE) scheme, in the spirit of “as-strong-as possible” security guarantee. Specifically, we explore the statistical measure approach from IR and text mining to embed weight information (i.e., relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection.

In information retrieval, inverted index is a widely-used indexing structure that stores a list of mappings from keywords to the corresponding set of files that contain this keyword, allowing full text search. For ranked search purposes, the task of determining which files are most relevant is typically done by assigning a numerical score, which can be precomputed. In information retrieval, a ranking function is used to calculate relevance scores of matching files to a given search request. The most widely used statistical measurement for evaluating relevance score in the information retrieval community uses the TF× IDF rule, where TF (term frequency) is simply the number of times a given term or keyword (we will use them interchangeably hereafter) appears within a file (to measure the importance of the term within the particular file), and IDF (inverse document frequency) is obtained by dividing the number of files in the whole collection by the number of files containing the term[1].

To enable ranked searchable symmetric encryption for effective utilization of outsourced and encrypted cloud data under the aforementioned model, our system design should achieve the following security and performance guarantee.
The following goals:

Ranked keyword search: To explore different mechanisms for designing effective ranked search schemes based on the existing searchable encryption framework;

Security guarantee: to prevent cloud server from learning the plaintext of either the data files or the searched keywords, and achieve the "as strong- as-possible" security strength compared to existing searchable encryption schemes

Comparison Table

	Searchable symmetric Encryption	Ranked Searchable symmetric Encryption
Data search	Returns non relevant information	Returns more relevant information
Efficiency	Low	High
Traffic	High	Low
Security &Accuracy	Low	High

IV. CONCLUSION

Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and ensures the file retrieval accuracy. System use statistical measure approach, i.e. relevance score, from information retrieval to build a secure searchable index. Ranking is done with the help of score value it helps to retrieve more relevant information. Ranked Searchable Symmetric Encryption is the best method for searching the outsourced cloud data.

REFERENCES

1. Cong Wang, Ning Cao, Kui Ren , Wenjing Lou, Senio(2012), "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data*", IEEE Transactions on Parallel and Distributed systems, VOL.23,NO.8.*
2. 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.
3. P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on Jan. 23rd, 2010 Online at http://csrc. nist.gov/groups/SNS/cloud-computing/index.html, 2010.
4. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep. UCBEECS- 2009-28, Feb 2009.
5. 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.

6. E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, Report 2003/216, 2003
7. 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.
8. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proc. of EUROCRYP'04, volume 3027 of LNCS*. Springer, 2004.
9. Ms. M. R. Girme, 2Prof. G.M. Bhandar" Efficient Ranked Keyword Search For Achieving Effective Utilization Of Remotely Stored Encrypted Data In Cloud" Volume 3, Issue 6, June 2014
10. Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.
11. 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.
12. S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in Proc. of EDBT'09, 2009.
13. A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Orderpreserving symmetric encryption," in Proc. of Eurocrypt'09, volume 5479 of LNCS. Springer, 2009.

BIOGRAPHY

Deepa R is a M Tech computer science student in Marian Engineering college Kazhakootam Trivandrum Kerala. She completed her B Tech with first class in Computer Science from Cochin University.

Reeja S L is working as an Assistant professor at Computer Science and Engineering department of Marian Engineering College Kazhakootam Trivandrum Kerala.