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 5, May 2015

Searching on Encrypted Cloud Documents using Keywords

S.Keerthiga, S.Savitha Karpagam, T.M.Sathish Kumar

P.G Scholar, Department of CSE, Velalar College of Engineering and Technology, Anna University Chennai, Thindal, India

Asst. Prof, Department of CSE, Velalar College of Engineering and Technology, Anna University Chennai, Thindal, India

Asst. Prof, Department of ECE, K.S.R College of Engineering, Anna University Chennai, Tiruchengode, India

ABSTRACT- Sensitive cloud data have to be encrypted to protect information safety measures, before outsourcing. The encryption technique makes effective data utilization service a very challenging job. Traditional searchable encryption technique allows users to securely search over encrypted data through keywords. The safety allows information searching scheme provides solution for secure ranked keyword search over encrypted obscure information. Ordered search greatly developed technique usability by enabling search result relevance ranking instead of sending undifferentiated results and further ensures the accuracy of folder recovery. The numerical determine process, i.e., importance gain, from information retrieval is explored to build a secure searchable index. One-to-many order-preserving mapping technique is developed to properly protect those sensitive score information. The system facilitates server-side ranking without losing keyword security. The method is improved to importance gain active development. Investigate outcome verification is also provided in the system. One-to-many order-preserving mapping method is also enhanced in reversible manner. The similarity analysis technique is used to identify the query results under the cloud data storage.

KEYWORDS: Cloud Computing; Document Retrieval; Keyword based Search Engine; Ranking; Relevance Score

I. INTRODUCTION

Cloud Computing is the long dreamed vision of computing as a check, where cloud users can slightly accumulate their information into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing properties [5]. The profits brought by this new computing model include but are not limited to: relief of the burden for storage space organization, universal information access with autonomous environmental locations and avoidance of capital cost on hardware, software and personnel preserve, etc

As Cloud Computing becomes widespread, more and more open information are being centralized into the cloud, such as e-mails, special health documents, company finance information and government files, etc. The fact that information owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted information at risk [4] the cloud server may leak information to unauthorized entities [10] or even be hacked [6]. It follows that responsive information has to be encrypted prior to outsourcing for information privacy and combating unsolicited contact. Information encryption makes efficient information utilization a very challenging task given that there could be a large amount of outsourced information files. Besides, in Cloud Computing, information owners may share their outsourced information with a large number of customers, who might want to only get back certain specific information files they are interested in during a given assembly. One of the most accepted ways to do so is through keyword-based search. Such keyword search method allows users to selectively retrieve files of interest and has been widely applied in plaintext search states. Unfortunately, information encryption, which limits customer's capability to perform keyword search and further demands the protection of keyword privacy, makes the usual plaintext search methods fail for encrypted cloud data.

DOI: 10.15680/ijircce.2015.0305054

Our involvement can be reviewed as follows:

1. For the first time, we define the problem of secure ranked keyword search over encrypted cloud data and provide such an efficient set of rules, which fulfills the secure positioned search functionality with little weight gain information leakage against keyword privacy.
2. Thorough security analysis shows that our ranked searchable symmetric encryption scheme indeed enjoys “as-strong-as-possible” security guarantee compared to previous searchable symmetric encryption (SSE) schemes.
3. We investigate the practical considerations and enhancements of our ranked search method, including the efficient support of weight gain dynamics, the confirmation of ranked search results and the reversibility of our proposed one-to- many order-preserving mapping techniques.
4. Extensive experimental results demonstrate the effectiveness and efficiency of the proposed solution.

II. RELATED WORK

Traditional searchable encryption has been widely studied as a cryptographic primitive, with a focus on security definition formalizations and efficiency improvements. Song et al. 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 developed a Bloom filter-based per-file index, reducing the workload for each search request proportional to the number of files in the collection. Chang and Mitzenmacher also developed a similar per-file index scheme. To further enhance search efficiency, Curtmola et al. 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. Aiming at tolerance of both minor typos and format inconsistencies in the user search input, fuzzy keyword search over encrypted cloud data has been proposed by Li et al. in [9]. Very recently, a privacy-assured similarity search mechanism over outsourced cloud data has been explored by Wang et al. in [2]. Note that all these schemes support only Boolean keyword search and none of them support the ranked search problem which we are focusing on in this paper. Following our research on secure ranked search over encrypted data, very recently, Cao et al. [1] propose a privacy-preserving multi keyword ranked search scheme, which extends our previous work in [1] 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 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. 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. Zerr et al. use a different order-preserving mapping based on presampling and training of the relevance scores to be outsourced, which is not as efficient as our proposed schemes. 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. This can be observed from the Binary Search (.). 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.

III. PROBLEM STATEMENT

Sensitive cloud information have to be encrypted to protect information security, before outsourced to the commercial public cloud. The encryption process makes effective information utilization service a very challenging job. Traditional searchable encryption techniques allow users to securely search over encrypted information through keywords.

Searchable encryption technique supports only Boolean search process. Large amount of users and information files are not efficiently handled by the searchable encryption model. The privacy enabled information searching scheme provides solution for secure ranked keyword search over encrypted cloud information. Ranked search enhances system usability by enabling search result relevance ranking.

Relevance score is a statistical measure approach is used in information retrieval. Relevance score is used in secure searchable index preparation process. One-to-many order-preserving mapping technique is used to properly protect those sensitive score information. The system facilitates server-side ranking without losing keyword privacy. Ranked Searchable Symmetric Encryption (RSSE) scheme is used to perform secured data retrieval process.

The following drawbacks are identified in the existing system.

- Static relevance score model
- Complex reversible operation under order preserving scheme
- Result authentication is not provided
- Retrieval latency is high

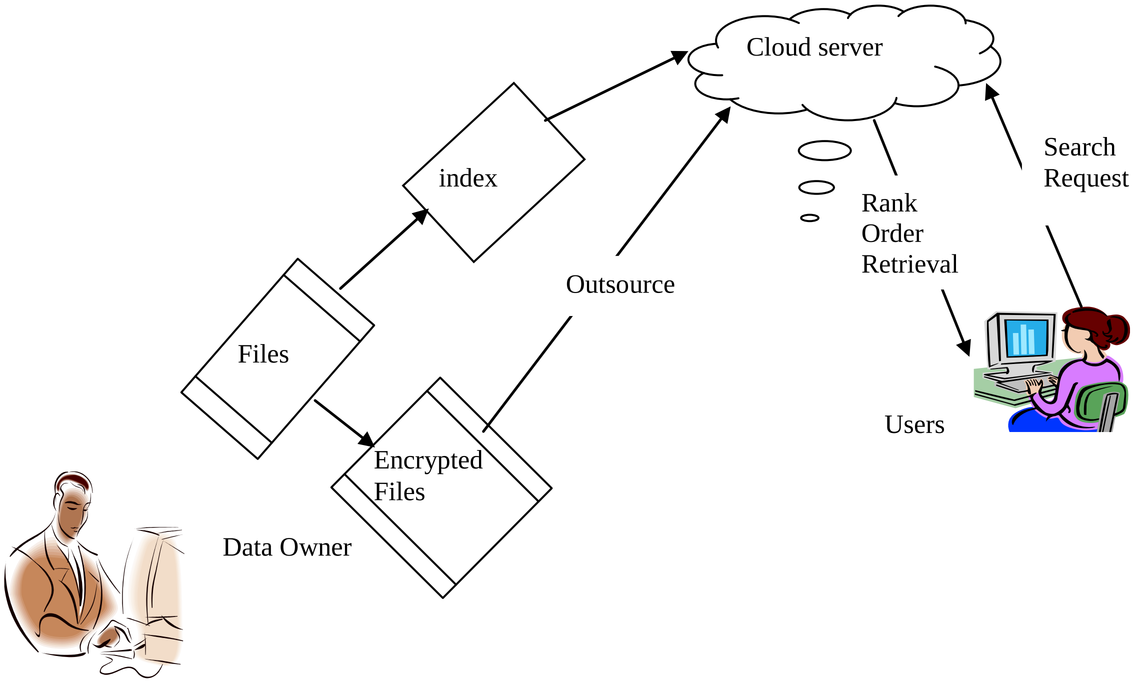

Fig: 1 Architecture of Search over Encrypted Cloud Data

IV. PROPOSED METHODOLOGY

The system is improved to support relevance score dynamics process. Search result authentication is also provided in the system. One-to-many order-preserving mapping technique is also enhanced in reversible manner. The similarity analysis scheme is used to identify the query results under the cloud data storage.

The cloud information center manages the transactional information ethics. The system is designed to provide data security and privacy for the transactional information over the cloud surroundings. The order preserving mapping model is used for the encryption process. The score operations are used to obtain the information ethics in a ranked manner. The dynamic scoring mechanism is used in the system.

The system is divided into two functions. They are information resource and customer function. The information resource manages the transactional information values. The customer application issues the query value

DOI: 10.15680/ijircce.2015.0305054

and collects the information from the information resource. The information ethics are updated in the information resource in an encrypted standard. The information retrieval and ranking operations are carried out on the encrypted information standard only. The scheme secures the information under the storage and query transmission process.

The system is divided into five most important modules. They are information resource, storage space organisation, score assignment, client and query process. The data source module is designed to manage the data values. The storage management module is designed to perform the data encryption and changed procedure. The score assignment module is used to assign the relevance score the for the transactional information values. The customer application is used to obtain the data value from the data source. The query process module is designed to submit and collect the data values.

Data Source

The data source application is designed to manage the transactional and user information. The user information are updated with their right to use in order. All the question record is keep under the data source application. The transactional data values are maintained for dissimilar domains. The data values are updated in encrypted format. The data retrieval is performed under the data source application.

Storage Management

The storage management is designed to handle data encryption and update operations. The order preserving mapping technique is used to encrypt the record standards. The method includes the reversible sort preserving map model for the encryption process. The information update function can be dynamically performed on the system. The data values are updated and stored in the encrypted design. The transactional information and its encryption process are carried out under the data source environment.

Score Assignment

The score assignment module is designed to assign the score values for the transactions. The similarity value is estimated to assign the achieved standards. The weight achieved is used to position the transaction data values. The data retrieval is carried out with the score operations. The incremental information update initiates the dynamic score task process. The dynamic score task process updates the score values based on the new transaction data values.

Client

The client application is designed to perform the data retrieval operations. The data values are collected from the server and updated into the user border. Each user is true with sole identification value. The customer collects the data values with query keywords.

Query Process

The query process module is designed to fetch the transactional data values. Query keyword is collected from the client. The query keyword is encrypted and transferred to the data source. The data source performs the searching process. The transactional data values are compared and similarity values are estimated. The results are prepared using the similarity value and entry levels. The client application decrypts the transactional data values and produces the results in a ranked way.

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 5, May 2015

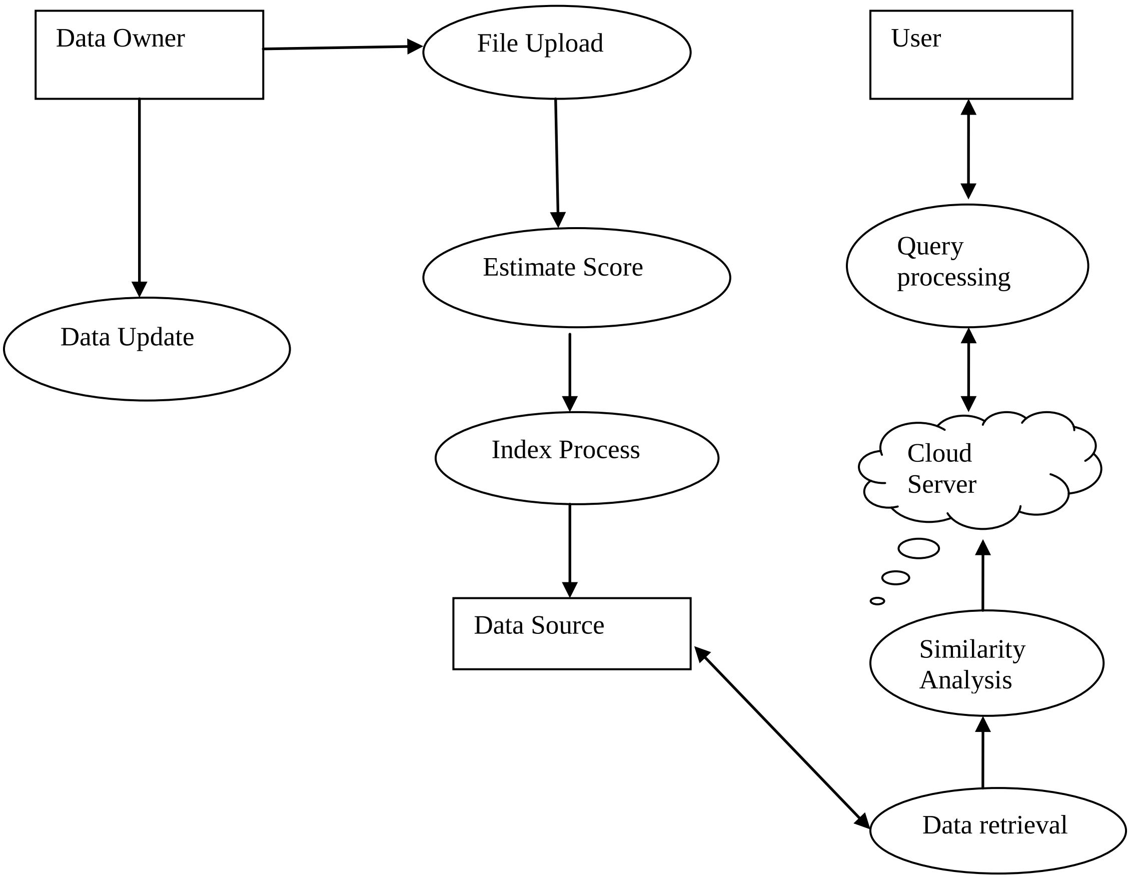

Fig: 2 Data Flow Diagram

V. EXPERIMENTAL RESULTS

Figure 3 present overall graphical representation of mean precision ratio of search engine for first 20 documents. While figure 3 and 4 represents graphical representation for queries (number 1 to number 10).

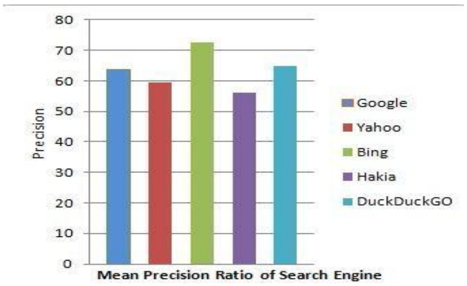

Fig. 3. Precision ratio of search engines for first 20 documents

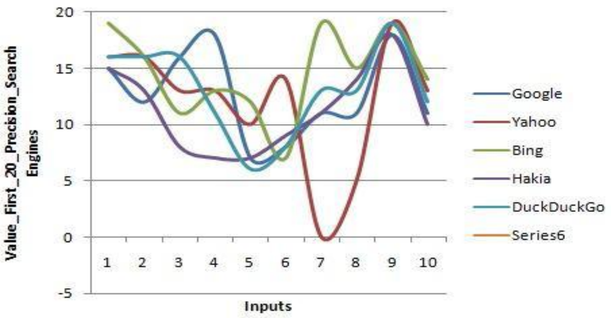

Fig. 4 Visual Representation of Search Engines for first 20 Precision

It is clear from the Fig 4 that semantic search engine like Bing and DuckDuckGo retrieved more relevant documents than keyword based search engine like Google and Yahoo. However, search performance of Hakia, which is a semantic search engine is lowest. Table 1 shows the results whether a search engine able to provide correct answer or not for natural language queries asked.

TABLE 1 CORRECT ANSWER FOUND SEARCH ENGINES

Query Number	Google	Yahoo	Bing	Hakia	DuckDuckGo
Q1	×	×	×	✓	×
Q2	✓	×	×	✓	✓
Q3	✓	✓	✓	✓	✓
Q4	✓	✓	×	×	×
Q5	✓	✓	✓	✓	✓

VI. CONCLUSION AND FUTURE WORK

Cloud customers can remotely store their data on a shared pool of configurable computing capital in cloud. Searchable Symmetric Encryption scheme is used to provide storage and retrieval security. Order Preserving Symmetric Encryption scheme is enhanced in reversible method. The scheme is get better result authentication and similarity based ordered representation. The information storage space and investigate method is carried out with encrypted query model. The system performs index operations on encrypted information ethics. The scheme also secure the investigate outcome. The system supports incremental information update scheme.

REFERENCES

1. N. Cao and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE Infocom ’11, 2011.
2. C. Wang, K. Ren, S. Yu, K. Mahendra and R. Urs, “Achieving Usable and Privacy-Assured Similarity Search over Outsourced Cloud Data,” Proc. IEEE INFOCOM, 2012.
3. M. Armbrust, I. Stoica and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing,” Technical Report UCB-EECS-2009-28, Univ. of California, Feb. 2009.
4. Cloud Security Alliance “Security Guidance for Critical Areas of Focus in Cloud Computing,” http://www.cloudsecurityalliance.org, 2009.
5. C. Wang, N. Cao, J. Li and Lou, “Secure Ranked Keyword Search over Encrypted Cloud Data,” Proc. IEEE 30th Int’l Conf. Distributed Computing Systems, 2010.
6. B.Kerbs, “Payment Processor Breach May Be Largest Ever,’’ http://voices.washingonpost.com/securityfix/2009/01/payment_processor_breach_may_b.html, Jan. 2009.
7. Ning Cao, Ming Li, Kui Ren and Wenjing Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 1, January 2014
8. S. Zerr, D. Olmedilla, W. Nejdl and W. Siberski, “Zerber+r: Top-k Retrieval from a Confidential Index,” Proc. Int’l Conf. Extending Database Technology: Advances in Database Technology (EDBT ’09), 2009.
9. J. Li, Q. Wang, K. Ren and W. Lou, “Fuzzy Keyword Search over Encrypted Data in Cloud Computing,” Proc. IEEE Infocom ’10, 2010.
10. Z. Slocum, “Your Google Docs: Soon in Search Results?” http://news.cnet.com/8301-17939_109-10357137-2.html, 2009.