ISSN: 0976-3104
SPECIAL ISSUE
Jayanthi and Prabadevi

THE IIOAB JOURNAL

RESEARCH OPEN ACCESS

MULTI - KEYWORD RANKED SEARCH OVER ENCRYPTED DATA SUPPORTING SYNONYM QUERY

Jayanthi M.* and Prabadevi

School of Information Technology and Engineering, VIT University, Vellore, TN, INDIA

ABSTRACT

With the advantage of storage as a service many enterprises are moving their valuable data to the cloud, since it costs less, easily scalable and can be accessed from anywhere any time. The trust between cloud user and provider is paramount. We use security as a parameter to establish trust. Cryptography is one way of establishing trust. Searchable encryption is a cryptographic method to provide security. In literature many researchers have been working on developing efficient searchable encryption schemes. In this paper we explore some of the effective cryptographic techniques based on vector space model (VSM), which manages increased control complexity of the data centre, and a less efficient cloud storage system.

Received on: 6th-May-2015
Revised on: 2nd-Sept-2015
Accepted on: 17th-Sept-2015
Published on: 5th-Jan-2016

KEY WORDS

Vector space model, cloud service provider, term frequency, inverse document frequency

***Corresponding author: Email:** jayanthimurali93@gmail.com **Tel:** +91-9894716850

INTRODUCTION

Cloud Computing paradigm provides a variety of service to the consumers. Many consumer electronic devices (e.g. Smartphone) with support of high speed computing combined with the emerging cloud. A cloud computing middleware Media Cloud for set top boxes for classifying, searching, and delivering media inside home network and across the cloud [1].

One hand, consumer-centric cloud computing a new model of enterprise-level IT infrastructure that provides on demand high quality applications and services from a shared pool of configuration computing resources for consumers. On the other hand, some problems may be caused in this circumstance since the Cloud Service Provider (CSP) possesses full control of the outsourced data. Sensitive data are encrypted before outsourcing to the cloud.

However, encrypted data make the traditional data utilization services based on plaintext keyword search useless [2]. The simple and embarrassed method of downloading all the data and decrypting locally is obviously impractical, because the authorized cloud consumers must hope to search their interested data rather than all the data

MATERIALS AND METHODS

Multi keyword ranked search

The existing systems like exact or fuzzy keyword search, supports only single keyword search. These schemes doesn't retrieve the relevant data to users query therefore multi-keyword ranked search over encrypted cloud data remains a very challenging problem. To meet this challenge of effective search system, an effective and flexible searchable scheme is proposed that supports multi-keyword ranked search [3]. To address multi-keyword search and result ranking, Vector Space Model (VSM) is used to build document index, each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. A new vector is also generated in the query phase. The vector has the same dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure can be used to compute similarity of one document to the search query. To improve search efficiency, a tree-based index structure used which is a balance binary tree is. The searchable index tree is constructed with the document index vectors. So the related documents can be found by traversing the tree.

THE IIOAB JOURNAL
ISSN: 0976-3104

Synonym search

While user searching the data on cloud server it might be possible that the user is unaware of the exact words to search, i.e. there is no tolerance of synonym substitution or syntactic variation which are the typical user searching behaviors and happen very frequently [4]. To solve this problem semantic based search method is used. To improve the search for information it is necessary that search engines can understand what the user wants so they are able to answer objectively. To achieve that, one of the necessary things is that the resources have information that can be helpful to searches.

The Semantic Web proposed to clarify the meaning of resources by annotating them with metadata data over data [5]. . By associating metadata to resources, semantic searches can be significantly improved when compared to traditional searches. It allows users the use of natural language to express what he wants to find [6]. Here the enhanced VSM algorithm is proposed for improving documental searches optimized for specific scenarios where user want to find a document but don´t remember the exact words used, if plural or singular words were used or if a synonym was used.
The defined algorithm takes into consideration:
1) The number of direct words of the search expression that are in the document;
2) The number of word variation of the search expression that are in the document;
3) The number of synonyms of the words in the search expression that are in the document.

Vector space model algorithm

Vector space model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers, for example, index terms. It is used in information filtering, information retrieval, indexing and relevancy rankings. To address multi-keyword search and result ranking, Vector Space Model (VSM) is used to build document index, each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. A new vector is also generated in the query phase [7].

RESULTS

Our system consists of 3 entities viz., data owner, data user and the cloud server as shown in **Figure– 1**.

Data owner

1. Encrypts the data files for securing the data in cloud before uploading into the cloud.
2. They define the access rights for the user who want to access those documents.
3. The access right is a 2-state variable: permission granted or permission denied.

Cloud server

1. Stores the encrypted data files and encrypted index tree.
2. It accepts the encrypted keywords and returns the matching data file based on their relevance.

Data user

1. User can search for encrypted data files in cloud with encrypted keywords.
2. The purpose of using encrypted keywords is that even the cloud server must not be able to infer the contents of data files.

This system is implemented in **ASP. NET** framework using C# and the process includes the following:

1. The data owner uploads the text files (documents) in cloud storage. Once the files get uploaded the owner is generating the encryption key through which the uploaded files are encrypted and are available only for the authorized users.
2. The registered data users can search the interesting data and can request the key from the data owner to download the document. The system matches the partial substring or various synonyms of the actual document name in the storage. For example, consider a document named Model.doc is uploaded in the cloud storage. The user query strings such as form, plan, and mod will be matched for the string "Model" and the document Model.doc will be retrieved. Through which Multi-Keyword searching is achieved.
3. Data owner forwards the decryption key to the secured user to access the documents; else the pop up error message will arise when the document is accessed without decryption key request made by the respective user.

File uploading

THE IIOAB JOURNAL
ISSN: 0976-3104

In the new system we build a secure index and outsource it along with the encrypted data items. Each index is mapped to a data bucket. Data bucket contain id of all the documents which have the bucket index as one of its index. At time of document upload the client send a request to the server for a unique password. Then the server generate the password form the features of document and how uploading the document and send to the client/data owner. Then using that password the data owner encrypt the document and uploaded the document along with the secure indexes [8].

At the time of file upload the server check for corresponding bucket in the data base based on the index word given by the user and selects the corresponding bucket to a data item. If there is no such a bucket then the server creates a new bucket for the documents index and adds the document to the newly created bucket. The password generation module which generates the unique password for the document provides extra security. The password generation consists of message digest creation using SHA-1 and converts the 160 bit message digest to 128 bit key for encryption [9].

The data owner's unique identifier, the file's unique identifier and the file name are used to generate the message digest. The AES is the encryption scheme used to encrypt the files. Based on the words inside the document the system itself able to predict the index words, but the final submission is by the data owner. He can select from the predicted words by the system or can add manually and the figure for this is shown in **figure–2** File uploading.

Files searching

Search is performed based on the all the synonyms and search words across bucket indexes and return all the authorized documents corresponding to from the selected buckets. The results are ranked based on the history and the number of times the document id is present in the buckets. That is it support for multi-key word search and then returns the best result as the first document and the figure for this is shown in **figure–2** File searching

Files downloading

The input to the SHA-1 is the data owner's unique identifier, the file's unique identifier and the file name. Then the 160 bit is converted to 128 bit and the 128 bit key is used for the AES decryption. The file and the key are given to the client and the decryption is performed in the client side and the figure for this is shown in **figure–2** File Downloading.

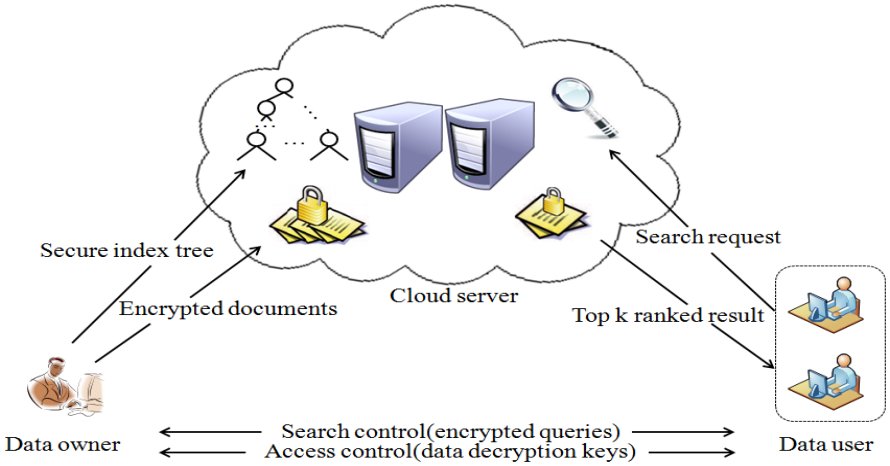

Fig: 1. Searchable encryption architecture using vsm

A **B**

THE IIOAB JOURNAL
ISSN: 0976-3104

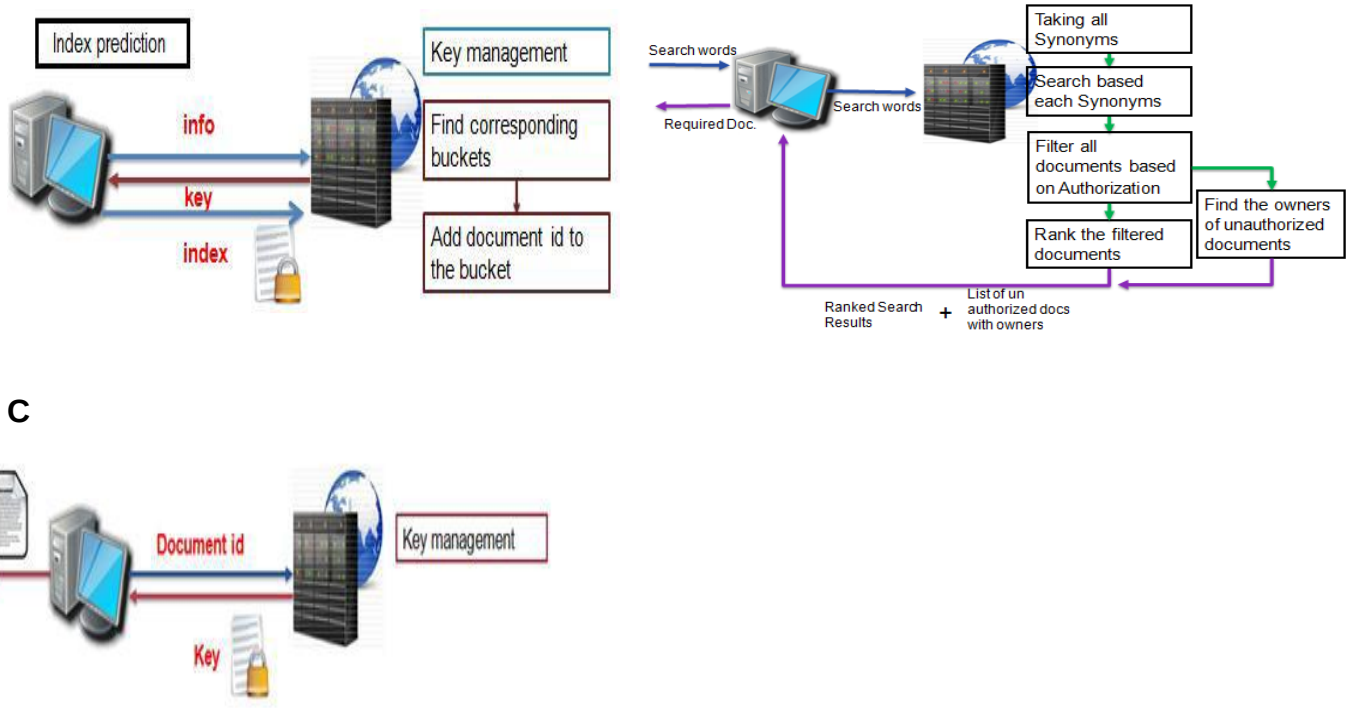

Fig: 2. File uploading, searching, downloading

DISCUSSION

Here we present the experimental evaluation of the Synonym based Ranked Secure Search over Encrypted Data. We investigated the success of our proposed scheme in the context of user unaware key management, automatic index prediction; synonym based searching, multi-key word search, user authentication and document ranking. For the purpose of testing we constructed a database of 200 documents entries.

User unaware key management

To verify the user unaware key management we want to evaluate mainly 2 things. First one is the uniqueness of the key. The second is the same key generation at the time of encryption and decryption. First one is evaluated using uploading same document with same indexes by same and different data owners [10]. And the result is checked by performing search on the data base for any two similar encrypted documents. 10 data owners are uploaded the same document 5 tines but cannot find the similar encrypted files in the data base. The rest is evaluated by downloading random 100 documents and check they are perfectly decrypted or not. We observed that all are decrypted correctly.

Automatic index prediction

The automatic index prediction is evaluated by checking the predicted indexes and keywords of the documents. Testing is done with 20 document files, 20 text files, 20 PDF file. The most frequent words, which may consider as the candidate for the index, are successfully predicted by the system [11].

THE IIOAB JOURNAL
ISSN: 0976-3104

Search Performance based on number of documents in each bucket

The performance of the system is checked by testing the execution time of the search with different indexes which have various numbers of documents inside each index bucket [11]. The number of documents in each bucket is denoted. At the starting the search time is increased with increasing number of documents inside the bucket. But later it become almost constant.

CONCLUSION

The energy efficiency is the important key Wireless sensor networks. With data transmission is the major part of energy consumption, chaos theory based time series prediction method to enhance energy efficiency. The proposed Chaos Theory based Data Aggregation (CTAg) based approach reduces redundant data, communication overhead and number of packet transmission between aggregator and sink node by using adaptive thresholds. The time series prediction using CTAg method was energy efficient and performed less computation to obtain the forecasted data. The experiments also show CTAg achieves better performance compared to other prediction approaches like Kalman Filter [KF] based prediction.
.

CONFLICT OF INTEREST

The author declares having no competing interests.

ACKNOWLEDGEMENT

We are thankful to VIT University for providing necessary resources for successfully implementation of this system

FINANCIAL DISCLOSURE

No financial support was received for this implementation.

REFERENCES

[1] Almearas. [2014] Achieving Effective Cloud Search Services Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query- *IEEE Transactions on Consumer Electronics- 2014*

[2] R Sanchez, P Arias, D Diaz-Sanchez, and A Marin. [2012] Enhancing privacy and dynamic federation in IdM for consumer cloud computing, *IEEE Trans. Consumer Electron* 58(1): 95–103.

[3] Chai and G Gong. [2012]Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers, Proceedings of (ICC'12), *IEEE International Conference on Communications* pp. 917–922.

[4] PA Cabarcos, FA Mendoza, RS Guerrero, AM Lopez, D Diaz-Sanchez. [2012] SuSSo: seamless and ubiquitous single sign-on for cloud service continuity across devices, *IEEE Trans. Consumer Electron* 58(4): 1425–1433.

[5] D Diaz-Sanchez, F Almenarez, A Marin, D Presario, PA Cabarcos. [2011] Media cloud: an open cloud computing middleware for content management, *IEEE Trans. Consumer Electron*, 57(2): 970–978.

[6] S Grzonkowski, PM Corcoran. [2011] Sharing cloud services: user authentication for social enhancement of home networking, *IEEE Trans. Consumer Electron*, 57(3): 1424–1432.

[7] J Li, Q Wang, C Wang, N Cao, K Ren, W Lou. [2010] Fuzzy keyword search over encrypted data in cloud computing,*Proceedings of IEEE INFOCOM'10 Mini-Conference, San Diego*, CA, USA, pp. 1–5, Mar.

[8] C Wang, N Cao, J Li, K Ren, and W Lou. [2010] Secure ranked keyword search over encrypted cloud data, Proceedings of *IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, 253–262.

[9] SG Lee, D Lee, and S Lee. [2010] Personalized DTV program recommendation system under a cloud computing environment," *IEEE Trans. Consumer Electron.*, 56(2):1034–1042.

[10] N Cao, C Wang, M Li, K Ren, and W Lou. [2011] Privacy-preserving multi keyword ranked search over encrypted cloud data," Proceedings of *IEEE INFOCOM* 2011, 829–837.

[11] Chai, G Gong. [2012] Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers, Proceedings of IEEE *International Conference on Communications* (ICC'12), 917–922.

[12] R Sanchez, P Arias, D Diaz-Sanchez, A Marin. [2012] Enhancing privacy and dynamic federation in IdM for consumer cloud computing," *IEEE Trans. Consumer Electron*., 58(1): 95–103.

THE IIOAB JOURNAL
ISSN: 0976-3104

Jayanthi M. *is currently associated with School of Information and Technology, VIT, Vellore, India. She is now pursuing M.S. in Software Engineering. Her area of interest is Cloud Computing, Data Mining*.

Prof. Prabadevi *is working as an Assistant Professor in the School of Information and Technology, VIT University, Vellore. She has completed her Undergraduate and post graduation under Anna University, Chennai and currenty pursuing her Ph.D at VIT University, in the area of Security in Cloud Computing.*