Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 2, February 2017, pg.131 – 136

Enabling Personalized Search and Integrity Check over Encrypted Outsourced Data with Efficiency Improvement

Sunil Chakravarthi Kaippada[1], Deepa S[2], Ambarish A[3]

[1]Department of Computer Science and Engineering, Malabar CET, India

[2]Department of Computer Science and Engineering, Malabar CET, India

[3]Department of Computer Science and Engineering, Malabar CET, India

[1] sunilchakravarthikuku@gmail.com; [2] dpsnkr@gmail.com; [3] ambruappat@gmail.com

Abstract— These days in cloud computing search over encrypted outsourced data is an important topic. Current model follows a "one size fits all" or "exact keyword" search and ignore personalized search intention, So in this paper it propose a new scheme called personalized multi-keyword ranked search over encrypted data(PRSE) which helps to provide better user search experience and cloud integrity check to find the reliable cloud service providers(CSPs).We build a user interest model based on the user's search history and semantic ontology WordNet and it will helps to enable personalized search over the cloud data and it proposes two PRSE schemes to improve the search efficiency.

Keywords— cloud security, outsourcing security, personalized search, user interest model, cloud integrity.

I. Introduction

Cloud computing has been doing a very impotent role in these days in the whole world. It allows more and more companies to outsource their data and applications to the public clouds. But still it also vulnerable to security attacks since data owners no longer has direct control over data. So, to provide data privacy and security it recommended a practise for data owners to encrypt their data before uploading to cloud. This will helps the data owners to protect their data from illegal use and untrusted cloud service providers. But it also becomes too difficult for search purposes since it have to implement a searchable scheme for encrypted outsourced data.

Searchable encryption [1] is a popular search scheme over encrypted data, but it does not fully satisfy the user's needs and search intention. Most existing systems follow a "one size fits all" or "exact keyword" search for encrypted data, and ignores user's search experience due to their different behaviour towards the search. Enabling search with personalized intention will helps to increase the search efficiency by improving the results of search. So most of the schemes support exact keyword search, so we have to reformulate the search query using an interest model built based on user's search history. By reformulating search query based on user's search history in information retrieval (IR) we can improve the search efficiency and increase the user's search experience and this method is known as query extension.

Before outsourcing the data to public clouds, they have to create an index vector based on the document and then it needed to encrypt both index vector and document collection independently using symmetric key encryption. It also save their query terms in each search to update their user interest model to support query extension. During search operation it first uses query extension mechanism and then it encrypt the queryusing the same method used for index vector and then it broadcast to cloud which will use this request to search over encrypted index vector to provide the search results

This system slows the user to check the integrity of the cloud service provider using integrity check. It allows the user to check whether any of their uploaded file is modified or misused by cloud service providers by checking our original file and uploaded file tags. A tag will use to provide cloud integrity checking which is generated based on the contents of file during the uploading time.

II. BACKGROUND

A. Cloud computing

Unlimited virtualized storage resources and computational environment can be provided by cloud computing platforms as services through the internet. Today's cloud service providers guarantee highly available and parallel computing resources with low costs. But the most important problem faced by the cloud computing platform is the security issues, which will also includes illegal use by cloud service providers. To provide privacy and security for data, data owners have to encrypt their data before uploading.

B. Personalized ranked search

Based on the stored data we have to perform a search operation. Since data is encrypted we have to use new search scheme which also includes personalized query terms to provide better user's search experience. It should support personalized ranking of results considering different user's interests as keyword priority or preference.

C. Query semantic extension

Based on user's search history our scheme reformulates the search query inputted by user to generate a better and improved search query. Here it extend the search query with terms from user interest model based on the term frequency before submitting it into public cloud to perform search operation , which will helps to solve the limitation of keyword exact search.

D. Privacy-preserving and cloud integrity

We have to prevent cloud from knowing additional information from document collection, index vector, and search request. So system must provide keyword privacy, index confidentiality, query confidentiality and by providing integrity check system can make sure that users can check the reliability of the cloud service providers.

III.RELATED WORK

The first searchable encryption scheme in symmetric key is proposed by Song et al. [1], in which each word is encrypted independently and the user has to go through the whole document to search a certain keyword. And then some security definitions and many improvements or constructions have been proposed by Goh [2], Chang and Mitzenmacher [3] and Curtmola et al. [4]. Boneh et al. [5] propose the first public key-based searchable encryption scheme, here anyone owning the private key can search data encrypted by the public key. Among them, to address the spelling mistake, import issue of fuzzy keyword search scheme is proposed by Li et al. [6] and improved in [7]. And Wang et al. [8] and Cao et al. [9] do research on secure ranking on single keyword and multi-keyword respectively. After that, Sun et al. [11] improve the efficiency of multi-keyword search by adopting MDB-tree. However, most of existing searchable encryption schemes support only "exact keyword search", which affects user's experience and data usability. Fu et al. [12] propose a semantic keyword search scheme based on stemming algorithm, which helps users find relevant documents which contains semantically close keywords related to the query word. Furthermore, personalized search is also missed or ignored. Shen et al. [10] proposed a preferred keyword search scheme over encrypted data, but how to measure keyword preference is ignored. The artificial method of measuring keyword preference is time consuming and imposes a burden on the user. Moreover, it fails to consider different users' search histories and thus has great randomness.

IV.PRELIMINARIES

This section introduces some necessary background knowledge for our proposed scheme.

A. Keyword weight

Keywords are practical tools to summarize document content. In order to express keyword's significance to the document, we adopt the most widely statistical measurement "TF_IDF", where term frequency (TF) is the occurrence of the term appearing in the document, and inverse document frequency (IDF) is usually obtained by dividing the total number of document collection by the number of documents containing the term. Specially, TF represents the importance of the term within a document and IDF indicates the importance or degree of distinction within the whole document collection.

B. Keyword priority

Through the well-trained user interest model, we can get the access frequency of each keyword. The higher the access frequency of a keyword is, the more important the keyword is from viewpoint of the user. The importance of the keyword usually means its rank priority.

C. Relevance score

It is used to measure the score of the query to a document. We can divide the whole relevance score into many sub-scores to represent the connection of file to the keywords in the query. The product of the keyword weight and the keyword priority is regarded as the sub-score, and thus the accumulated sub-scores form the relevance score of the query to a document. We can express the keyword weight and the keyword priority as a vector, and so the inner product of these two vectors can achieve.

D. Secure inner product

When calculating the relevance score of the document to the query, we should use two vectors: the document vector and the query vector. However, it is not advisable to directly outsource two vectors onto the cloud at the risk of leaking index privacy and query privacy.

E. WordNet[20]

It is a lexical database available online, which offers a large repository of English lexical items. It is organized hierarchically by a collection of synset, the smallest unit, which represents a specific meaning of a word. Synsets are connected to one another through some conceptual-semantic and lexical relations, including synonym, hypernym, hyponym, meronym, holonym and so on. These relations play an important role in computational linguistics and natural language processing. For brevity, only three kinds of relations: Synonymous relation, Hypernym/Hyponym relation, Meronym/Holonym relation are used in building of the user interest model.

V. SYSTEM COMPONENTS

The system uses cloud architecture which involves a data user, a data owner and public cloud.

A. *Data owner*

The user uses the public cloud server for efficiently store their data in the cloud and access the data later. But before it uploads their data into cloud it create an index vector which is encrypted using a symmetric key and also it encrypt the document independently then it sends both index vector and document collection to cloud. During file access request it sends key for decrypting the document to data user. It also can check integrity by using tag generated based on the content of the original document and uploaded document.

B. *Data user*

The data user input their search query into cloud. During each search operation system updates user search history and built a user interest model to reformulates the original query and the extended query used for search.

C. *Cloud*

Cloud is the cloud service provider which provides us the online data storage facility. It stores encrypted data uploaded by data owners and perform search operation using the query done by using the search query which undergo query extension and return the results. It stores the encrypted index vector on which it performs the search rather than searching the all document collection.

VI. SYSTEM ARCHITECTURE

The system model in cloud computing have three different entities: the data owner, the data user and the cloud server, there exists a user interest model stored in the user side. The user interest model is built upon the user's long-term search history. It records access frequency of both query keywords and their related keywords with the help of WordNet [11].

Different access frequency of keywords, as keyword priority can reflect their different importance in viewpoint of the data user. To search for files of interest, the data user should firstly produce a search request. And then query reformulation that achieves keyword priority of query terms will be carried out through the user interest model.

The encrypted search query through search control mechanism, e.g., broadcast encryption[4], will be sent to the cloud. Upon receiving the search request from the authorized user, the cloud server will conduct designated search operation over the index and send back relevant encrypted documents, which have been well ranked by the cloud server according to some ranking criteria

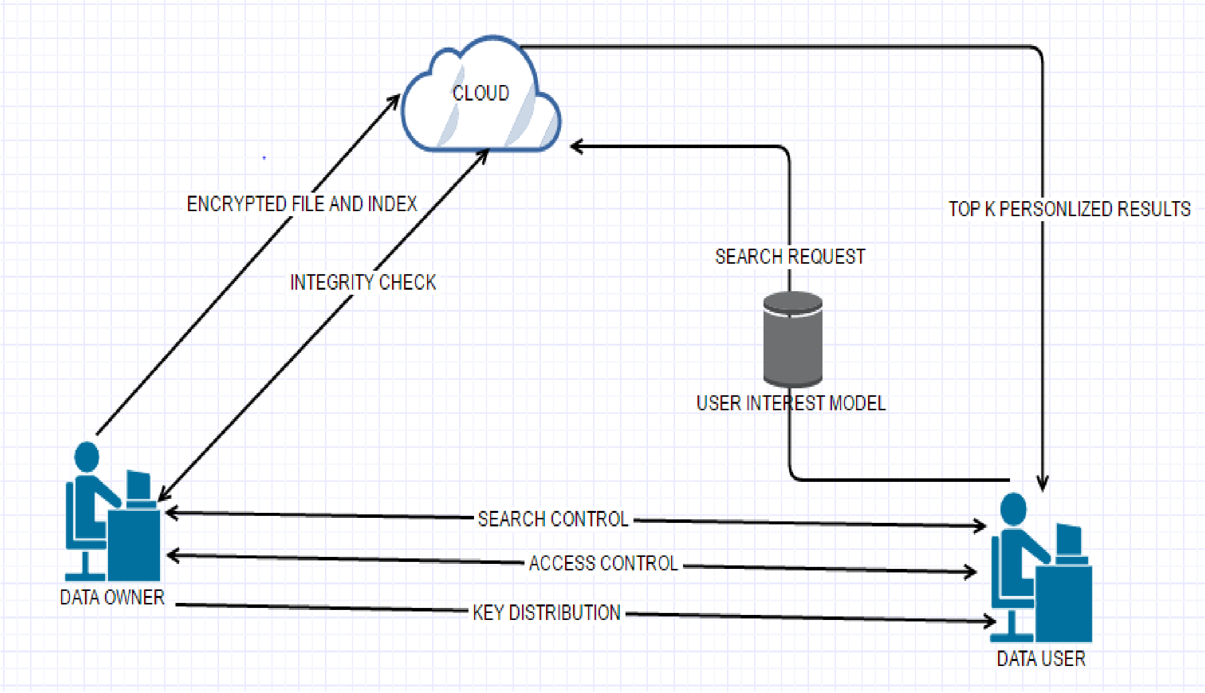

Fig 1: Architecture

The PRSE system consists of the following algorithms

A. Setup:

In this initialization phase, the data owner produces a symmetric key as SK.

B. BuildIndex(D,SK):

Based on the keyword set W extracted from the document collection D, the data owner can build a searchable index I, which is encrypted by the symmetric key SK, and encrypt the document collection independently. After that, the data owner outsources the index I and encrypted documents C onto the cloud.

C. BuildUIM:

The system collects the user's long-term search history, records the access frequency of query keywords, and then builds the user interest model.

D. GenQuery(W,U,SK):

With the query words W, the data user conducts query reformulation with the user interest model and then produces a corresponding trapdoor T encrypted by the same symmetric key SK. At last, T, together with parameter k, will be submitted to the cloud.

E. SearchIndex (I,T,K):

Upon receiving the trapdoor T, the cloud server conducts designated search operation over the index I, and returns RT; k sorted by the relevance score between documents and T.

VII. EVALUATION

The proposed approach used the improved search query but it uses more time than MRSE [9] to generate query because it uses query extension method .but document results are lesser in number because it only includes user intended documents.

In The following graph (fig 2), explains the advantage of using this proposed approach in the cloud computing environment.

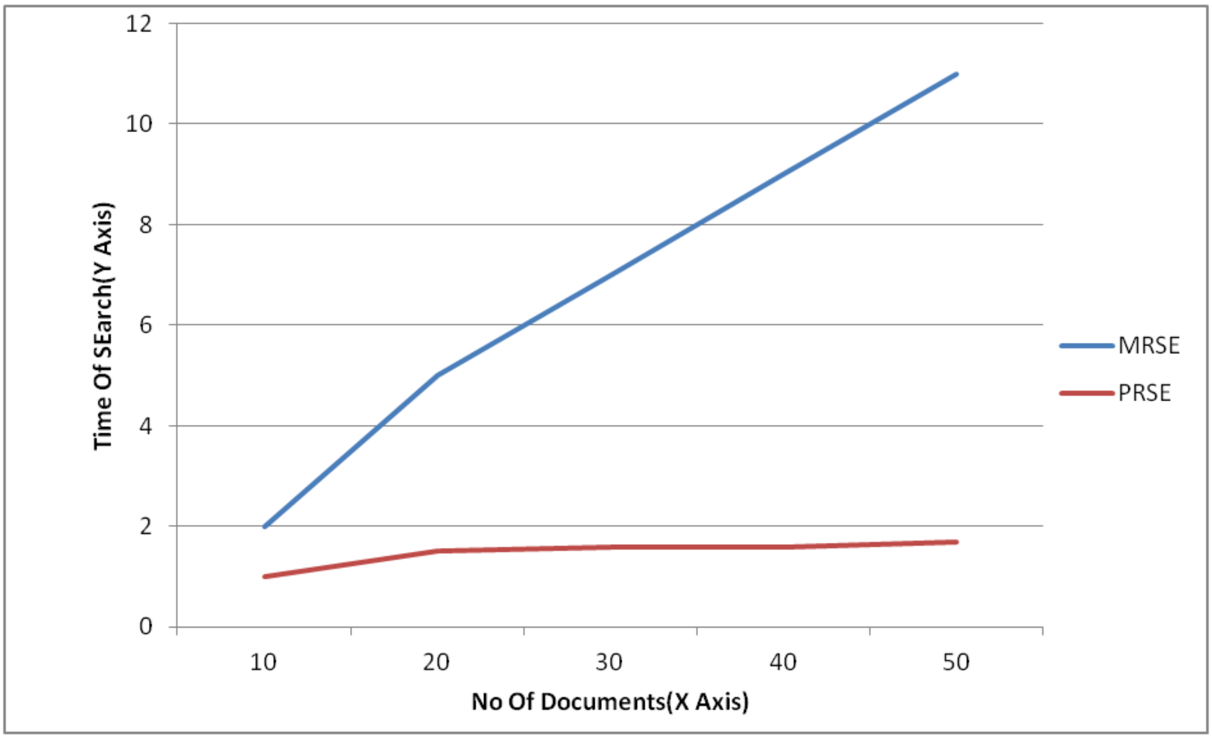

Fig 2: Time for Search

VIII. FUTURE WORKS

The current search engines, like GOOGLE, have launched personalized search service, where the user can indicate his interests or preferences explicitly, or preferences may be automatically collected. However, they have not been widely adopted yet since users worry about their privacy.. Moreover, most of existing personalized search schemes is inapplicable to cipher text. So we have to implement a trust towards the user for allowing us to use their search interests for future use and also have to develop a scheme to ease this process

IX. CONCLUSIONS

The In this paper, we address the problem of personalized multi-keyword ranked search over encrypted cloud data. Considering the user search history, we build a user interest model for individual user with the help of semantic ontology WordNet [11]. Through the model, we have realized automatic evaluation of the keyword priority and solved the limitation of the artificial method of measuring. Moreover, we propose two PRSE schemes to solve two limitations (the model of “one size fit all” and keyword exact search) in most existing searchable encryption schemes. In addition, thorough privacy analysis and performance analysis demonstrates that our scheme is practicable.

REFERENCES

[1] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symp. Security Privacy, 2000, pp. 44–55

[2] E.-J. Goh. (2003). Secure indexes. Cryptology ePrint Archive, Rep. 2003/216 [Online]. Available: http://eprint.iacr.org/

[3] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in Proc. 3rd Int. Conf. Appl. Cryptography Netw. Security, 2005, pp. 442–455.

[4] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky,"Searchable symmetric encryption: Improved definitions and efficien constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 79–88.

[5] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. Int. Conf. TheoryAppl. Cryptographic Techn., 2004, pp. 506–522.

[6] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. J. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. IEEE INFOCOM, San Diego, CA, USA,2010, pp. 1–5.

[7] C. Liu, L. H. Zhu, L. Li, and Y. Tan, "Fuzzy keyword search on encrypted cloud storage data with small index," in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., 2011, pp. 269–273.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE30th Int. Conf. Distrib. Comput. Syst. Workshop, 2010, pp. 253–262.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM, 2011, pp. 829–837.

[10] Z. Shen, J. Shu, and W. Xue, "Preferred keyword search over encrypted data in cloud computing," in Proc. IEEE 21st Int. Symp.Quality Service, 2013, pp. 1–6.

[11] G. A. Miller, "WORDNET: A lexical database for English," Commune. ACM, vol. 2, no. 11, pp. 39–41, 1995.

[12] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting Similarity-based ranking," in Proc. ACM 8th SIGSAC Symp. Inf., Comput. Commune. Security, 2013, pp. 71–82.

[13] Z. Fu, J. Shu, X. Sun, and D. Zhang, "Semantic keyword search based on trie over encrypted cloud data," in Proc. Proc. 2nd Int. Workshop Security Cloud Compute., 2014, pp. 59–62.