IJCSEC- International Journal of Computer Science and Engineering Communications.
Vol.3, Issue 2, 2015, Page.701-704, ISSN: 2347–8586

Effective Cloud Search Based on Multi Keyword Ranked Over Encrypted Cloud Data

S.SASIKALA[1] , A.JOSEPH SELVA KUMAR[2]
[1]*PG Student* [2]*Asst. Prof Dept. of IT,*
[1&2]*Affiliated to Anna University Chennai, Dept. of Computer Science and Engineering,*
Idhaya Engineering College for Women,
sasik.be@gmail.com.

ABSTRACT

In recent years, consumer-centric cloud computing paradigm has emerged as the development of smart electronic devices combined with the emerging cloud computing technologies. A variety of cloud services are delivered to the consumers with the premise that an effective and efficient cloud search service is achieved. For consumers, they want to find the most relevant products or data, which is highly desirable in the "pay-as-you use" cloud computing paradigm. As sensitive data (such as photo albums, emails, personal health records, financial records, etc.) are encrypted before outsourcing to cloud, traditional keyword search techniques are useless. Meanwhile, existing search approaches over encrypted cloud data support only exact or fuzzy keyword search, but not semantics-based multi-keyword ranked search. Therefore, how to enable an effective searchable system with support of ranked search remains a very challenging problem. This paper proposes an effective approach to solve the problem of multi-keyword ranked search over encrypted cloud data supporting synonym queries. The main contribution of this paper is summarized in two aspects: multi-keyword ranked search to achieve more accurate search results and synonym-based search to support synonym queries.

The ranked search enables cloud customers to find the most relevant information quickly. Ranked search can also reduce network traffic as the cloud server sends back only the most relevant data. Multi-keyword search is also very important to improve search result accuracy as single keyword search often return coarse search results. To meet the challenge of effective search system, this paper proposes a practically efficient and flexible searchable scheme which supports both multi-keyword ranked search and synonym-based search. To address multi-keyword search and result ranking, Vector Space Model (VSM) (says database,) is used to build document index, that is to say, each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. The contributions of this paper are summarized as follows: For the first time, a semantics-based multi-keyword ranked search technology over encrypted cloud data which supports synonym queries is proposed. The search results can be achieved when authorized cloud customers input the synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. Extensive experiments on the real-world dataset further show the effectiveness and efficiency of proposed solution.

Index Terms: Cloud computing, Consumer centric cloud, Multi-keyword ranked search, synonym based search.

INTRODUCTION

In recent years, many consumer electronic devices (e.g. Smartphone) with support of high speed computing combined with the emerging cloud computing paradigm provide a variety of service to the consumers. A novel middleware architecture that allows sessions initiated from one device to be seamlessly transferred to a second one under a cloud computing environment. Cloud computing middleware Media Cloud for set top boxes for classifying, searching, and delivering media inside home network and across the cloud. A personalized DTV program recommendation system under a cloud computing environment. The system can analyze and use the viewing pattern of consumers to personalize the program recommendations. However, all these services are likely to be available to consumers only with the premise that an effective and efficient cloud search service is achieved. Consumers want to find the most relevant products or data, which is highly desirable in the "pay-as-you use" cloud computing paradigm.

One hand, consumer-centric cloud computing is 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. There may exist unauthorized operation on the outsourced data on account of curiosity or profit. So sensitive data are encrypted before outsourcing to the cloud. However, encrypted data make the traditional data utilization services based on plaintext keyword search useless. The simple and awkward 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. Hence, it is an especially important thing to explore an effective search service over encrypted outsourced data. Existing search approaches cannot accommodate such requirements like ranked search, multi-keywords search, semantics-based search etc. The ranked search enables cloud customers to find the most relevant information quickly. Ranked search can also reduce network traffic as the cloud server sends back only the most relevant data.

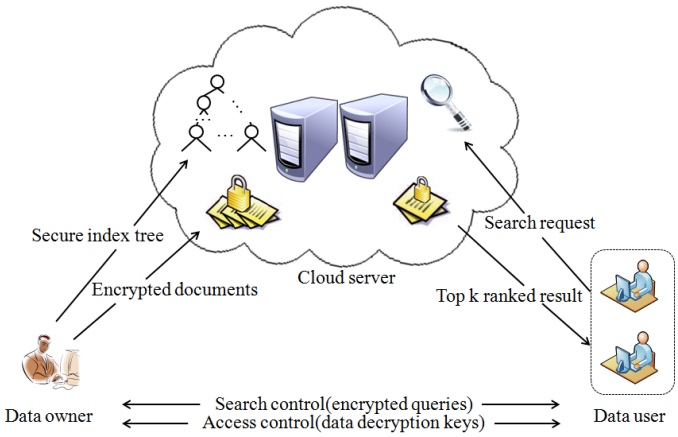
Multi-keyword search is also very important to improve search result accuracy as single keyword search often return coarse search results. In the real search scenario, it is quite common that cloud customers' searching input might be the synonyms of the predefined keywords, not the exact or fuzzy matching keywords due to the possible synonym substitution (reproduction of information content), such as commodity and goods, and/or her/his lack of exact knowledge about the data. The existing searchable encryption schemes support only exact or fuzzy keyword search. That is, there is no tolerance of synonym substitution and/or syntactic variation which, on the other hand, are typical user searching behaviors and happen very frequently. Therefore, synonym-based multi-keyword ranked search over encrypted cloud data remains a very challenging problem. To meet the challenge of effective search system, this paper proposes a practically efficient and flexible searchable scheme which supports both multi-keyword ranked search and synonym based search. To address multi-keyword search and result ranking, Vector Space Model (VSM) is used to build document index, that is to say, 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 which is a balance binary tree is used. The searchable index tree is constructed with the document index vectors. So the related documents can be found by traversing the tree.

The contributions of this paper are summarized as follows: For the first time, a semantics-based multi-keyword ranked search technology over encrypted cloud data which supports synonym queries is proposed. The search results can be achieved when authorized cloud customers input the synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. By incorporating the state-of-art text feature extraction technique TFIDF (term frequency-inverse document frequency), an enhanced semantic feature extraction method E-TFIDF is proposed. The E-TFIDF algorithm, which can extract the most representative keywords from outsourced text documents, improves the accuracy of search results. Extensive experiments on the real-world dataset further show the effectiveness and efficiency of proposed solution.

CLOUD AUTHORIZATION

The owner will register his/her details to the cloud server provider. Then the registered data owner will upload his/her data's to Cloud service provider. Upload can be done as private or public. Private file's will be uploaded in the data owner private folder and Public file's will be uploaded to common public folder. Private files will be accessed only to his/her search but the public files will be accessed to all the data users who accessing that cloud service provider. Here the files will be upload as it normal file without any encryption. Data owner will sometimes act as data user.

ARCHITECTURE



CONTENT ENCRYPTION

Data owner will create the index file from the source file which should be uploaded in the cloud service. And then source file will be encrypted with the encrypting algorithm and the encrypted file will be uploaded to the cloud service provider along with its index file. Encryption key is only known to data owner.

CONTENT INDEXING

To create index file we use Text miming Process. The Text mining process analyses the text word by word and also picks up the literal meaning behind the group of words that constitute the sentence. The Words are analyzed in WorldNetApi so that the related terms can be found for use in the index file. All the communication to cloud server will be done through web service.

EFFECTIVE CLOUD SEARCH OVER PROCESSED QUERY

Data owner or data user will try give query to the cloud server. The query also will be processed to find the related terms and then it will be searched in the index files. If the terms is present in more than one file, then will search results will be show according to its rank frequency. To which file, frequency is more that file will be listed first. User also will able to give rank to that particular file which will be updated in the cloud server database. Next time that according to frequency and user rank files will be listed to user. If the user selects the particular file then the Encrypted content for the particular file will be provided to the user. Then user has to decrypt that file to access using key which is used for Encryption.

RELATED WORK

A. Consumer-centric Cloud Services

A novel middleware architecture that allows sessions initiated from one device to be seamlessly transferred to a second one under a cloud computing environment. A cloud computing middleware Media Cloud for Set-top boxes for classifying, searching, and delivering media inside home network and across the cloud. A personalized DTV Program Recommendation system under a cloud computing environment. The system can analyze and use the viewing pattern of consumers to personalize the program recommendations. A user centric approach to authentication for home networks. This approach enables the sharing of personalized content and more sophisticated network-based services over a conventional TCP/IP infrastructure. An IDM architecture based on privacy and reputation extensions to enable the global scalability and usability for consumer cloud computing paradigm. However, all these services are likely to be available to consumers only with the premise that an effective and efficient cloud search service is achieved.

B. Searchable Encryption in Cloud

To apply the searchable encryption to cloud computing, some researchers have been studying further on how to search over encrypted cloud data efficiently. A fuzzy keyword search scheme over encrypted cloud data, which combines edit distance with wildcard-based technique to construct fuzzy keyword sets, to address

problems of minor typos and format inconsistence. A ranked search scheme, in which by giving each keyword a weight TF-IDF, the cloud server can rank relevant data files with no knowledge of a specific keyword weight. But this scheme supports only single keyword search. A ranked scheme supporting multi-keyword, which uses a vector space model and characteristics of matrix to realize trapdoor unlinkablility and there by preserves data privacy. A verifiable symmetric search encryption scheme, which can prove the correctness and completeness of results. A multi-keyword ranked search scheme based on vector space model (VSM). The VSM can measure the similarity between document index vector and query vector and hence support more accurate ranked search results. But this scheme cannot support semantics-based search.

CONCLUSION

For the first time, proposes an effective approach to solve the problem of synonym-based multi keyword ranked search over encrypted cloud data. The main contributions are summarized in two aspects: synonym-based search and similarity ranked search. The search results can be achieved when authorized cloud customers input the synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. The vector space model is adopted combined with cosine measure, which is popular in information retrieval field, to evaluate the similarity between search request and document. Finally, the performance of the proposed schemes is analyzed in detail, including search efficiency and search accuracy, by the experiment on real-world dataset. The results show that the proposed solution is very efficient and effective in supporting synonym-based searching.

FUTURE WORK

The next work is to research semantics-based search approaches over encrypted cloud data that support syntactic transformation, anaphora resolution and other natural language processing technology. The aim is that cloud consumers can search the most relevant products or data by using the designed system.

REFRENCES

1. Cao.N et al (2011), “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” Proceedings of IEEE INFOCOM 2011, pp. 829-837.

2. Grzonkowski.S, and Corcoran.P.M (2010) “Sharing cloud services: user authentication for social enhancement of home networking,” IEEE Trans.Consumer Electron., vol. 57, no. 3, pp. 1424-1432.

3. Lee.S.G et al (2010), “Personalized DTV program recommendation system under a kcloud computing environment,” IEEE Trans. Consumer Electron., vol. 56, no. 2, pp. 1034-1042,

4. Li.J et al (2010), “Fuzzy keyword search over encrypted data in cloud computing,” Proceedings of IEEE INFOCOM’10 Mini-Conference, San Diego, CA, USA, pp. 1-5.

5. Sun.W, et al (2013),“Privacypreserving multi-keyword text search in the cloud supporting similarity based ranking,” ASIACCS 2013, Hangzhou, China, May 2013, pp. 71-82.

6. Wang.C et al (2010), “Secure ranked keyword search over encrypted cloud data,” Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 253-262.

7. Zhangjie Fu (2014), “Achieving Effective Cloud Search Services:Multi-Keyword ranked search over encrypted cloud data supporting synonym query” ,“IEEE Transactions on Consumer Electronics”, Vol. 60, No. 1.