International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 08
July 2017

Privacy Preserving and Fast Record Search and Retrieval Scheme over Encrypted Cloud Data

T.Charan Singh ; B.Swarna sri ; V.Navya sree
[1]Assistant Professor in Department Of CSE.Sri Indu College Of Engineering and Technology
[2]PG Scholar in Department Of CSE.Sri Indu College Of Engineering and Technology
[3]PG Scholar in Department Of CSE.Sri Indu College Of Engineering and Technology

Abstract:
Nowadays, cloud computing becomes efficient and flexible with reduced cost and utility of on-demand high qualityapplications and services, so internet usage strongly relies on cloud for privacy preserving and fast Records retrieval. For consumers, they want to find the most relevant products or Records , which is highly desirable in the "pay-as-you use" cloud computing paradigm. Sensitive Records is encrypted before outsourced to cloud. Although Substitutable encryption scheme has been developed to conduct retrieval over encrypted Records , these schemes only support exact or fuzzy keyword Substitute, mainly evaluate the similarity of keywords from the structure but the semantic relatedness is not considered. This work focuses on realizing secure semantic Substitute through query keyword semantic extension based on the co-occurrence probability of terms, the semantic relationship library is constructed to record the semantic similarity between keywords. To achieve Proficiency of the Substitute method we enhance the TFIDF algorithm by extending the keyword set with semantic words or natural language words for the keywords. This will ultimately support Records retrieval on querying semantic query. Even when user doesn't know exact or synonym of keywords of encrypted Records , he can try Substituting it by its meaning in natural language. WordNet method makes the Substitute scheme even more reliable and better.

Keywords: Cloud computing, multi-keyword Substitute, semantic based Substitute, TFIDF, anaphora resolution, WorldNet Ontology.

1. INTRODUCTION

Today, consumer centric cloud computing is a new model of enterprise-level in IT infrastructure providing the on-demand high quality applications and services from a shared pool of computing resources. The Cloud Service Provider (CSP) has full control of the outsourced Records ; it may learn some additional information from that Records therefore some problems arise in the circumstance. So, sensitive Records is encrypted before outsourcing to the cloud. However the encrypted Records make the traditional plaintext Substitute methods useless. The simple and awkward method is downloading all Records and decrypt it locally is obviously impractical, because the consumers want to Substitute only the interested Records rather all the Records . Therefore it is essential to explore an efficient and effective Substitute service over encrypted outsourced Records.

The existing Substitute approaches like Classified Substitute, multi-keyword Substitute that enables the cloud customers to find the most relevant Records quickly. It also reduces the network traffic by sending the most relevant Records to user request. But In real Substitute scenario it might be possible that user Substitutes with the synonyms of the predefined keywords not the exact or fuzzy matching keywords, due to lack of the user's exact knowledge about the Records .These approaches supports only exact or fuzzy keyword Substitute. That is there is no tolerance of synonym substitution and/or syntactic variation which are the typical user Substituting behaviors happens very frequently. Therefore synonym based multi-keyword Classified Substitute over encrypted cloud Records remains a challenging problem.

To overcome this problem of effective Substitute system this paper proposes an efficient and flexible Substituteable scheme that supports both multi-keyword Classified Substitute and Semantic based Substitute. The Vector Space Model is used to address multi-keyword Substitute and result ranking. By using VSM document index is build i.e. each

document is expressed as vector where each dimension value is the Term Frequency (TF) weight of each corresponding keyword. Another vector is generated in query phase. It has same dimension as that of document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure is used to calculate the similarity between the document and the Substitute query.

To enhance the Proficiency of the Substitute method we use the extended keyword set with semantic words or natural language words for the keywords. This will ultimately support Records retrieval on querying semantic query. Even when user doesn't know exact or synonym of keywords of encrypted Records , he can try Substituteing it by its meaning in natural language. WordNet ontology is used to solve the problem of anaphora resolution. This makes the Semantic Substitute more efficient and User doesn't need to worry about the keyword generated for each particular word on the cloud by adapting this method Records will be retrieved from the cloud in well secure manner and also cost can be minimized by employing these scheme into the structure and also we are incorporating WordNet method which makes the Substitute scheme even more reliable and Better.

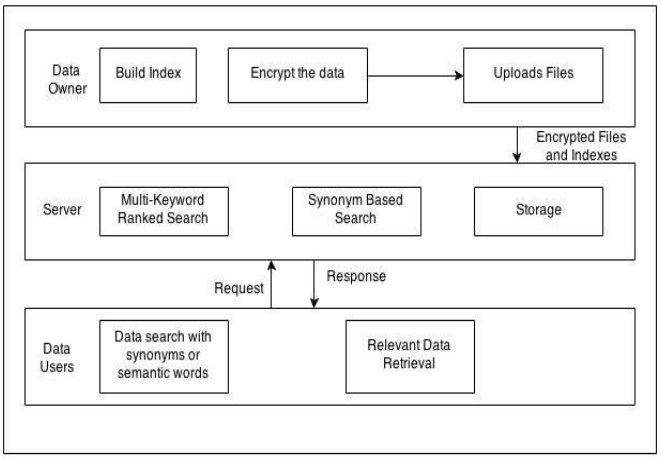

Figure 1: System Architecture

2. LITERATURE SURVEY

J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, uses the Fuzzy keyword Substitute method that enhances system usability by returning the matching files containing exact match of the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. They exploit edit distance to quantifykeywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads [2].

C. Wang, N. Cao, J. Li, K. Ren, and W. Lou proposes the Classified Substitute that enhances system usability by returning the matching files in a Classified order regarding to certain relevance criteria (e.g., keyword frequency). It gives a straightforward yet ideal construction of Classified keywordSubstitute under the state-of-the-art Substituteable symmetric encryption (SSE) security definition, and demonstrates its inProficiency. To achieve more practical performance, they propose a definition for Classified Substituteable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE) [3].

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou designed a system that solves the challenging problem of privacy-preserving multi-keyword Classified Substitute over encrypted cloud Records (MRSE), and establish a set of strict privacy requirements for such a secure cloud Records utilization system to become a reality. Among various multi-keyword semantics, they choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity

between Substitute query and Records documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement [4].

W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou present a privacy-preserving multi-keyword text Substitute (MTS) scheme with similarity-based ranking to address this problem. To further enhance the Substitute privacy, they propose two secure index schemes to meet the stringent privacy requirements under strong threat models. In particular, to support multi-keyword queries and Substitute result ranking functionalities, they proposes to build the Substitute index based on the vector space model, i.e., cosine measure, and incorporate the TF × IDF weight to achieve high Substitute result accuracy[6].

Zhangjie Fu, Xingming Sun, Nigel Linge and Lu Zhou proposes an effective approach to solve the problem of multi-keyword Classified Substitute over encrypted cloud Records supporting synonym queries. To address multi-keyword Substitute 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 Substitute query. To improve SubstituteProficiency , a tree-based index structure which is a balance binary tree is used

3. METHODOLOGY

3.1 Multi-Keyword Classified Substitute:

The existing systems like exact or fuzzy keyword Substitute, supports only single keyword Substitute. These schemes doesn't retrieve the relevant Records to users query therefore multi-keyword Classified Substitute over encrypted cloud Records remains a very challenging problem. To meet this challenge of effective Substitute system, an effective and flexible Substituteable scheme is proposed that supports multi-keyword Classified Substitute. To address multi-keyword Substitute 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 Substitute query [1].

To improve SubstituteProficiency , a tree-based index structure used which is a balance binary tree is. The Substituteable index tree is constructed with the document index vectors. So the related documents can be found by traversing the tree.

3.2 Semantic Based Substitute:

While user Substituteing the Records on cloud server it might be possible that the user is unaware of the exact words to Substitute, i.e. there is no tolerance of synonym substitution or syntactic variation which are the typical user Substituteing behaviors and happen very frequently. To solve this problem semantic based Substitute method is used. To improve the Substitute for information it is necessary that Substitute 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 Substitutees.

The Semantic Web proposed to clarify the meaning of resources by annotating them with metaRecordsRecords over Records . By associating metaRecords to resources, semantic Substitutes can be significantly improved when compared to traditional Substitutees. It allows users the use of natural language to express what he wants to find. Here the enhanced E-TFIDF algorithm is proposed for improving documental Substitutees 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 Substitute expression that are in the document; 2) the number of word variation (plural/singular or different verbs conjugation) of the Substitute expression that are in the document; 3) the number of synonyms of the words in the Substitute expression that are in the document; weights to each one of this components as the fuzziness part of the algorithm [7].

4. CONCLUSION

The proposed Semantic Substitute with WordNet methodology makes the Substitute process more efficient. The proposed scheme could return not only the exactly matched files, but also the files including the terms semantically related to the query keyword. The concept of co-occurrence probability of terms is used to get the semantic relationship of keywords in the Records set. It offers appropriate semantic distance between terms to accomplish the query keyword extension. To guarantee the security and Proficiency , the Records is encrypted before outsourced to cloud, and provides security to Records sets, indexes and keywords also. Then the Records owner groups the indexes and forms the ontology based on the documents which is having syntactically and semantically similar words.

The overall performance evaluation of this scheme includes the cost of metaRecords construction, the time necessary to build index and ontology construction as well as the Proficiency of Substitute and WordNet methodology which makes the Substitute scheme still more efficient to the user and by employing this technique keyword that we used for Substituteing will also protected and better Substitute mechanism can be achieved.

References

[1] Zhangjie Fu, Xingming Sun, Nigel Linge and Lu Zhou,Attaining Effective Cloud Substitute Services: Multi-keyword Classified Substitute over Encrypted Cloud Records Supporting Synonym Query", IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014.

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

[3] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Classified keyword Substitute over encrypted cloud Records ,"

Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 253-262,2010.

[4] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword Classified Substitute over encrypted cloud Records ," Proceedings of IEEE INFOCOM 2011, pp. 829-837, 2011.

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

[6] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou, "Privacy preserving multi-keyword text Substitute in the cloud supporting similarity based ranking," ASIACCS2013, Hangzhou, China, May 2013, pp. 71-82, 2013.

[7] Sara Paiva, "A Fuzzy Algorithm for Optimizing Semantic Documental Substitutees", International Conference on Project Management / HCIST 2013.

[8] Automatic Pronominal Anaphora Resolution in English Texts, Tyne Liang and Dian-Song Wu.