S.Amsaveni *et al*, International Journal of Computer Science and Mobile Applications,

Vol.3 Issue. 10, October- 2015, pg. 01-05 **ISSN: 2321-8363**

A REVIEW ON KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

G.T.Prabavathi
Assistant Professor of Computer Science
Gobi Arts & Science College
Gobichettipalayam
gtpraba@gmail.com

S.Amsaveni
Research Scholar in Computer Science
Gobi Arts & Science College
Gobichettipalayam
amssathy@gmail.com

Abstract

Due to the advancements in cloud computing users are allowed to store sensitive information in the centralized cloud. When data owners share their outsource data with a large number of users who are willing to retrieve only certain specific data, then keyword based search is essential. Keyword based search technique allows users to retrieve selective files of interest. The objective of privacy preserving algorithm is to extract relevant knowledge from large amount of data while protecting the sensitive information at the same time. Applying keyword based search in the encrypted cloud data is a challenging task due to security and privacy obstacles. There exists many research works to solve the problem of privacy preserving over encrypted cloud data using multi-keyword ranked search technique. This paper is a preliminary attempt to survey the algorithms that preserves privacy in cloud data using multi-keyword ranked searching.

Keywords: cloud computing, privacy preserving, keyword search.

1. Introduction

Privacy is an important issue for cloud computing both in terms of legal complains and user trust and needs to be considered at every phase of design [5]. Privacy preserving is used to preserve the security of fields. If a database has to be shared among several users and some data contained in the database should be prevented by using access control methods in order to guarantee that only authorized people are allowed to have access that sensible information, then the need of privacy and preserving the privacy emerges.

The cloud computing provides users with the ability to outsource their data to public cloud for economic savings and flexibility (10). Hence, Keyword based searches are essential in cloud data. Keyword searches are typically done so that users can actively search clouds to query a collection of data. It allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Multi-keyword ranked search algorithms are readily available in information retrieval process. Similar algorithms are essential for cloud server data while protecting the cloud data as well as to enhance the search privacy.

The paper is organized as follows. Chapter II deals with the concerns in privacy preserving and Chapter III lists the uses of ranked keyword searching. Chapter IV discusses the importance of multi-keyword ranked searching and a study on few algorithms that deal with multi-keyword ranked searching.

Vol.3 Issue. 10, October- 2015, pg. 01-05 **ISSN: 2321-8363**

2. Privacy Preserving

Privacy can be described as "Limited access to a person and to all the features related to the person" [Igor]. Privacy Preserving has originated as an important concern with reference to the success of the data mining. A class of Data Mining (DM) methods, known as privacy preserving data mining algorithms, has been developed by the research community working on security and knowledge discovery. It deals with protecting the privacy of individual data or sensitive knowledge without sacrificing the utility of the data.

The aim of privacy preserving algorithm is the extraction of relevant knowledge from large amount of data, while protecting at the same time sensitive information. Several DM techniques incorporating privacy protection mechanisms, have been developed that allow one to hide sensitive item sets or patterns, before the data mining process is executed(2). In information retrieval (IR) community, there exist state-of-the-art techniques to achieve good result ranking using multi-keyword queries on plain text. The success of privacy preserving data mining algorithm is measured in terms of its performance and data level of uncertainty. Next chapter deals with ranked keyword searching and the necessity of multi-keyword ranked searching.

3. Ranked Keyword Searching

As cloud computing has become an integral part of IT industry, data owners share their outsourced data. Due to these vast amounts of information available on WWW, large number of users attempts to retrieve certain specific data files they are interested in. One of the most popular ways to do so is through keyword based search. Keyword searches are done to utilize cloud data for a certain query. Such keyword search techniques allow users to selectively retrieve files of interest and have been widely applied in plain text search scenarios (C.wang). Great efforts have been made for facilitating users via keywords search. However, there are few researchers about entertaining the exact user query and presenting a ranked URL list according to it. Keywords searchers are typically done in such a way that users can utilize clouds to query a collection (7).

To eliminate unnecessarily network traffic by not sending back the irrelevant data, ranked keyword search is used. This technique is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation should not leak any keyword related information. To improve the search result accuracy as well as to enhance the user searching experience, it is necessary for such ranking system to support multi-keyword search, as single keyword search often yields far too coarse results (5). The information is retrieved from the matching files to calculate the relevance scores of given request.

If ranking system supports multiple keyword search then, it is possible to improve the search result accuracy as well as user searching experience can be enhanced. In all web search engines, users provide a set of keywords instead of only one keyword to indicate that they are interested in a particular area. Each keyword in the user query is used to narrow down the search process.

4. Multi-Keyword Search Algorithms

Ning Cao et al (6) established a set of strict privacy requirements for solving Privacy Preserving Multi-Keyword Ranked Search over Encrypted data in Cloud Computing (MRSE). They proposed a basic idea for the MRSE based on secure inner product computation. In their proposed technique "*Coordinate Matching*" is used as similarity measure to

capture relevant data for a search query. The "*Inner Product Similarity*" scheme is used to quantitatively evaluate the similarity measure. The algorithm is practical, flexible and has low overhead on both computation and communication. In the MRSE scheme proposed by the authors, a new random number *t* is assigned to the extended dimension of each query vector to increase the difficulty for the cloud server to learn the relationships among received trapdoors. In addition a dummy keyword is inserted in each data vector and a random value is assigned to it. The authors also proposed a more advanced MRSE scheme to achieve various privacy requirements in two different threat models.

Li Chen et. al. (3) proposed a Semantic Multi-keyword Ranked Search Scheme over the encrypted cloud data. It utilizes the "*Latent Semantic Analysis (LSA)*" to reveal relationship between terms and documents. LSA takes the advantages of implicit higher-order structure in the association of terms with documents and it adopts reduced-dimension vector space to indicate words and documents. The authors used a vector consisting of term frequency (TF) values as indexes to the documents. From this matrix, LSA between terms and documents where analysed. A security splitting KNN technique was employed to encrypt the index and queried vector to obtain the accurate ranked results. Through this approach the authors extracted not only the exact matching files, but also the files including the terms that are latent semantically associated to the query keyword. When compared to the original MRSE scheme, the proposed technique attains higher score in F-measure.

Shieba et al (8) solves the challenging problem of privacy preserving MRSE over encrypted cloud data based on secure inner product computation and efficient similarity measure of coordinate matching, i.e., as many matches as possible in order to capture the relevance of data documents to the search query. The authors proposed significantly improved MRSE scheme to achieve various string and privacy requirements in two different threat models. An algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the members of the group. In the proposed system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the cloud service provider (CSP) and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained.

There are various approaches on encrypted cloud data search that either focus on single keyword search or become inefficient when a large amount of documents are present. So, a support for the efficient multi-keyword search is proposed by Yanzhi et al (11). The author proposed a light weight search that supports efficient multi-keyword ranked search in cloud computing system. First a basic scheme using polynomial function is used to hide the encrypted keyword and search patterns for efficient multi-keyword ranked search .To enhance the search privacy, the authors proposed a privacy preserving scheme which utilizes the secure inner product method for protecting the privacy of the searched multi-keywords. In this work, the authors encrypted the keywords and then constructed a polynomial function to hide them in search trapdoor generation. The detailed scheme to achieve the ranked multi-keyword search over encrypted data are performed using the functions setup, build index and trapdoor .The authors used ranked multi-keyword search scheme to hide the keywords in the search query using polynomial functions which provides guaranted privacy over two threat models. Through the proposed scheme, the authors conducted extensive experiments based on the real-world dataset. The experimental results demonstrated that the scheme proposed by Yanzhi et al (11) enables the encrypted multi-keyword ranked search service is highly efficient in cloud computing.

C. Wang et al (3) proposed an effective and secured ranked keyword search over encrypted cloud data. Ranked search greatly enhanced system usability by returning the matching files in the ranked order regarding to certain relevance criteria. The author proposed a definition for Ranked searchable symmetric encryption (RSSE) and gave an efficient design by properly utilizing the existing cryptographic primitive, order preserving symmetric encryption. Order-preserving symmetric encryption (OPSE) is deterministic encryption scheme where the numerical ordering of the plain texts gets preserved by the encryption function. If this property is not related appropriately it will leak lot of information like other deterministic encryption schemes. So, the authors modified order-preserving symmetric encryption scheme (OPSE) to reduce the amount of information leakage from the deterministic property. Then a file ID is introduced as an additional seed in the final cipher text chosen process.

5. Conclusion

The advancements of cloud computing allow users to store sensitive information in the centralized cloud data. The data owners share their outsource data with a large number of users to retrieve only certain specific data for which for keyword based search is essential. Privacy preserving algorithms focus on the process of extracting relevant knowledge from cloud data while protecting the sensitive information. The aim of this paper is to study the researches in privacy preserving over encrypted cloud data using multi-keyword ranked search technique. This paper has reviewed few algorithms related to multi-keyword ranked search problems**.**

REFERENCES

[1] A. Swaminathan, Y. Mao, G.M.Su, H. Gou, A. L. Varna, S. He, M. Wu, and D.W. Oard, "Confidentiality-preserving rank-ordered search," in Proc. of the Workshop on Storage Security and Survivability, 2007.

[2] C Wang, N Cao, J Li, K Ren, and W Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," In Proceedings of ICDCS, Vol.2, pp.253-262,2010.

[3] Li Chen, Xingming Sun, Zhihua Xia , Qi Liu," An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data" International Journal of Security and Its Applications, Vol.8, No.2, pp.323-332,2014.

[4] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," In Proceedings of 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383-392, 2011.

[5] N. Cao, Z. Yang, C. Wang, K.Ren and W. Lou, "Privacy preserving Query over Encrypted Graph-Structured Data in Cloud Computing, In Proceedings of Distributed Computing Systems (ICDCS), pp. 393-402, 2011.

[6] Ning Cao, C Wang, M Li, K Ren, and W Lou, "Privacy-preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", In Proceedings of IEEE INFOCOM, Vol.25, No.1,pp. 829–837, 2014.

[7] R. Agrawal and R. Srikant, "Privacy-Preserving data mining", proceeding of the ACM SIGMOD Conference on Management of Data, pp. 439–450, May 2000.

[8] Shiba Sampat Kale et al, "Privacy-preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over Encrypted Cloud Data"International Journal of Computer Science and Information Technologies, Vol.5 (6), pp.7093-7096, 2014.

[9] Wenhai Sun and et al., "Privacy-Preserving Multi-Keyword Text Search in the cloud supporting similarity-based ranking," in Proceedings of ACMSIGSAC,.pp.253-262,2013.

[10] Y.C.Chang and M.Mitzenmacher." Privacy preserving keyword searches on remote encrypted data". In Proc. of ACNS'05, pp. 442–455, 2005.

[11] Y Ren, Y Chen, J Yang, and B Xie, "Privacy-preserving Ranked Multi-Keyword Search Leveraging Polynomial Function in Cloud Computing," Globeco.,IEEE, pp. 594 - 600, 2014.

[12] Z.Yang, S.Zhong., and R.N.Wright."Privacy-preserving Queries on Encrypted Data". In Proc. of ESORICS'06, pages 479–495, 2006.

