IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

File Sharing System using Multi-keyword Ranked search with Data Cache

Nevil Raju Philip[1], R. Anitha[2]

M Tech Student, Department of PG Studies, National Institute of Engineering, Mysuru, India [1]

Associate Professor, Department of CSE, National Institute of Engineering, Mysuru, India [2]

Abstract: With the advent of communication systems and high speed data networks, the amount of data and files an individual or an organization has to maintain and access have also increased multi fold. But with the limited storage space available in a mobile device or a laptop it is not feasible to store all the files in the personal systems. Hence we require a common file sharing platform. These systems provide functionality to data owners to upload their private data and users to search and retrieve the required files. But performing search and retrieval while preserving data privacy is a challenge as data is stored globally by a third party. To improve data security almost all file sharing platforms store the data in an encrypted manner. For enabling search and retrieval functionality, it is required to create a searchable index which contains the keywords present in the files and its respective weight or relevance in the file. The initial systems only supported single keyword search and the keywords were not encrypted as computations on encrypted data is complex. To enhance security we propose a system that implements two-round search with multi-keyword search support. The keywords are also encrypted to enhance the security. To improve efficiency and system performance a user cache is implemented which helps is removing the duplicate files.

Keywords: Homomorphic encryption, Multi-Keyword based file retrieval, relevance search

I. INTRODUCTION

For efficient data management, individuals as well as organizations are resorting to outsourcing the data as it is becoming more flexible and cost effective. One of the concerns which prevent organizations to resort to such file sharing system is the concern about data security. To enhance security, the initial file sharing systems encrypt the files before storing them. But such systems used single keyword search and also the keywords were not encrypted. The goal of file sharing systems over encrypted data is to enable efficient search over encrypted data in the server without decrypting the files or the keywords. Recent advancements have resulted in more and more people resorting to file sharing platforms to store the files. Such a system will have large number of users and documents and hence the system should support multi-keyword based ranked search. Ranked search will allow users to retrieve only the most relevant documents with respect to the search query.

In order to attain privacy the files should be encrypted before storing it. Encrypting the data makes the search and retrieval challenging as the number of data files are large. Also in a file sharing system, the data owners will be sharing the files with large number of users who want to retrieve the required files only. One of the most commonly used method for search and retrieval is keyword based search, which allows users to selectively retrieve the files. This method was widely used in plaintext scenario where the keywords are stored as plaintext. Storing the keyword as plaintext compromises on data security as it leads to security concerns. So it is necessary to store the keywords also in an encrypted manner. In the initial systems the server used to build the searchable index, but this imposed additional complexity at the server. Also the files will be encrypted using either AES or DES encryption schemes. Hence server has to decrypt the files and then build the index. To remove this computational overhead, the encryption and search keyword generation is done at the data owner side[7] and then is stored in the server .

In this paper, in an attempt to increase the security the keywords are also encrypted. We implement an encryption scheme to encrypt the keywords which enables comparison to be performed on encrypted keywords and hence enables ranked multi-keyword search. The file sharing system does search on encrypted data and since the search query is also encrypted, this system provides high degree of data security.

The below architecture diagram shows a basic file sharing scenario. There are two types of users, data owner and data user. Data owner has the files to be uploaded to the server. The data owner creates the index from the files to be uploaded. The index contains the keywords and the weight of the keyword in the file. The files are encrypted using AES encryption scheme and the index is encrypted using partial homomorphic encryption. The data users can register into the system and the authorized users can search using multi-keyword search query. The search query is encrypted in similar homomorphic manner and is sent to the server. The server computes the score and sent

DOI 10.17148/IJARCCE.2016.5761

the result back to the user. The user request for the top-k files and it is then retrieved back from the server.

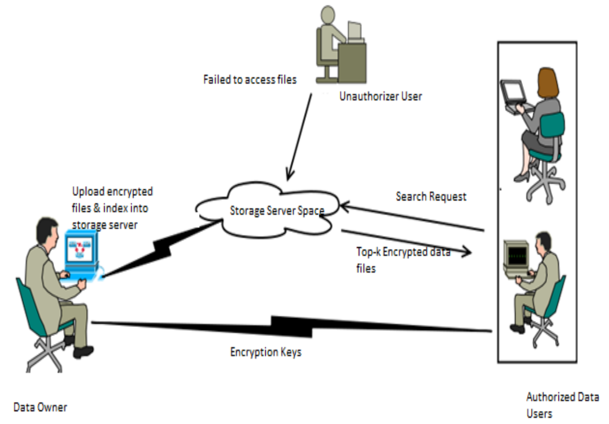

Fig 1 Basic Architecture of a search and retrieval system

II. RELATED WORK

Searchable Symmetric Encryption schemes enables search operations over encrypted data. Initial advancements in these fields[2],[4] provided ability to search over encrypted files without decrypting it, these schemes supported Boolean keyword search, ie all the files which has the keyword was retrieved. This makes the system highly inefficient.

As an improvement to single keyword based search, ranked search is introduced. In ranked search, the server gives the result for the search query based on some rank and relevance criteria like term or keyword frequency .One of the initial implementation of ranked search is Order Preserving symmetric Encryption (OPE) [3]. OPE is a deterministic encryption scheme where the numerical ordering is preserved even after encryption. But OPE still have the concern of statistical leakage

A new encryption scheme is proposed which allows computations to be performed on encrypted data. Such schemes are called Homomorphic encryption schemes [5]. Towards secure multi-keyword based ranked searched on encrypted cloud data [1] deals with the implementing the homomorphic encryption to enable efficient search and retrieval.

III.PROPOSED APPROACH

We implemented the top-k multi keyword based ranked search functionality to a file sharing systems. The files are encrypted with AES encryption scheme. The index is encrypted using partial homomorphic encryption. This enables the system to perform search on encrypted data without decrypting it first.

The data owner creates the index from the files before uploading it to the server. The index contains the keywords present and the term frequency. The index is encrypted with partial homomorphic encryption scheme.

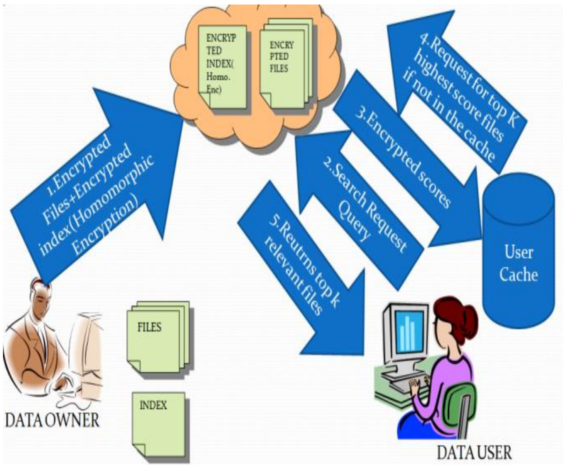
ENCRYPTED INDEX(Homo. Enc)
ENCRYPTED FILES
1.Encrypted Files+Encrypted index(Homomorphic Encryption)
4.Request for top K highest score files if not in the cache
3.Encrypted scores
2.Search Request Query
5.Returns top k relevant files
User Cache
FILES
INDEX
DATA OWNER
DATA USER

Fig 2 Basic Architecture of the proposed approach

For retrieving the files, the data user sents the multi-keyword search request query. The search query is also encrypted with similar partial homomorphic scheme before sending it to server. This ensures data security and prevents information leakage.

The server upon receiving the request process the request and find the score based on the multi-keyword search query. The server sent back the file id and score. The score denotes the weight of the keywords in the corresponding files. The user does the ranking and request for the top-k files that he require. The server responds by sending the corresponding encrypted files to the user who will decrypt and access the file.

To improve the system performance a user cache is introduces at the user side. The cache archives the previous search results. When the server sent the encrypted scores and file id. The user checks if the files are already present in the cache and only those files which are not present in the cache are requested to be downloaded. This prevents duplicate file accumulation in the user system and also improves the system performance and network bandwidth consumption.

IV. CONCLUSION

Here we deal with implementing a file sharing platform with multi-keyword based search facility, the system does ranked search and retrieves the top-k files the user require. Partial homomorphic encryption is used for encrypting the index and search query and hence the system security is enhanced. We introduced a user cache which will prevent duplicate file downloads and improves the system performance.

The system can be implemented for mobile devices so that we can access the data on the go and in a convenient manner. Another improvement that can be made is to incorporate attribute based encryption so that the data owner can selectively give data users permission to access files.

DOI 10.17148/IJARCCE.2016.5761

REFERENCES

[1] Jiadi Yu, Member IEEE, Peng Lu, Yanmin Zhu, Member IEEE, Guangtao Xue, Member IEEE Computer Society, and Minglu Li "Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, Vol. 10, NO. 4, July/August 2013.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data", Proc. IEEE 30th Intl Conf. Distributed Computing Systems (ICDCS), 2010.

[3] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant Yirong Xu , "Order Preserving Encryption for Numeric Data", IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120

[4] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions", Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[5] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers", Proc. 29th Ann. Intl Conf. Theory and Applications of Cryptographic Techniques, H. Gilbert, pp. 24-43, 2010.

[6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[7] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, 2011

[8] N. Smart and F. Vercauteren, "Fully Homomorphic Encryption with Relatively Small Key and Ciphertext Sizes," Proc. 13th Int'l Conf. Practice and Theory in Public Key Cryptography (PKC), 2010.

BIOGRAPHIES

Nevil Raju Philip is final year M.Tech Student of National Institute of Engineering, Mysuru. He has received his B.Tech from Cochin University of Science and Technology (CUSAT). His interested subjects include Data Security and Embedded Systems.

R. Anitha is Associate Professor in the Department of Computer Science & Engineering at National Institute of Engineering, Mysuru. She has received her M.Tech from VTU, and B.E. from University of Mysuru. She is pursuing her Ph.D. Her teaching and research interests are in the field of Networking & Cloud Computing.

DOI 10.17148/IJARCCE.2016.5761