ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

An Efficient and Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data: A Survey

Archana Galshetwar, Prof. Vaishali Nandedkar,

Dept. of Computer Engineering, PVPIT College of Engineering, Bavdhan Pune, India

ABSTRACT: The efficient search and privacy search is important objective in real time search engines, basically system present a scheme that discusses secure rank based keyword search over an encrypted cloud data. The data that has to be outsourced is encrypted using symmetric encryption algorithm for data confidentiality. The index file of the keyword set that has to be searched is outsourced to the local trusted server where the keyword set that is generated from the data files is also stored. This is done so that any untrusted server cannot learn about the data with the help of the index formed. The index is created with the help of Vector base cosine similarity (VCS) multiple strings matching algorithm which matches the pre-defined set of keywords with information in the data files to index them and store relevant data in B+ trees. Whenever the user searches for a keyword, the request is sent to the local trusted server and the indexed data is referred. The files are listed based on the certain relevance criteria. User requests for the required files to the un-trusted server. The parameters required for ranking is got from the data stored while indexing. Based on the ranking, the files are retrieved from the un-trusted server and displayed to the user. The proposed system can be extended to support Boolean search and multi keyword search techniques.

KEYWORDS: Symmetric Encryption algorithm, Rank based search, multiple string matching, relevance scoring, privacy preserving, and cloud computing

I. INTRODUCTION

Cloud Computing is the evolving technology that has changed the way of computing in IT Enterprise. It brings the software and data to the centralized data centers from where a large community of users can access information on pay per use basis. This poses security threats over the data stored. Data confidentiality may be compromised which has to be taken care of. So it becomes necessary to encrypt the data before outsourcing it to the cloud server. This makes data utilization a challenging task. Traditional searching mechanisms provide Boolean search to search over encrypted data, which is not applicable when the number of users and the number of data files stored in the cloud is large. They also impose two major issues, one being the post-processing that has to be done by the users to find the relevant document in need and the other is the network traffic that is undesirable in present scenario when all the files matching with keywords is retrieved. But this paper proposes ranked keyword search that overcomes these issues.

II. LITERATURE SURVEY

1] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang proposed A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud

In [1], author describes a secure multi-keyword ranked search scheme over encrypted cloud data, which supports dynamic update operations like deletion and insertion of documents. The vector space model and the widely-used TFIDF model are combined in the index construction and query generation. A special tree-based index structure and introduces a Greedy Depth-first Search algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and to ensure accurate relevance score calculation between encrypted index and query vectors.

DOI: 10.15680/IJIRCCE.2016. 0412198

2] Cong Wang proposed Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data.
In [2], author proposed search which solves processing overhead, data and keyword privacy, minimum communication and computation overhead. The owner build index along with the keyword frequency-based relevance scores for files. User request 'w' to CS with optional 'k' as Tw using the private key. The CS searches the index with scores and sends encrypted file based on ranked sequence. The CS searches the index with scores and sends encrypted file based on ranked sequence.

3] Madane S.A. proposed Single Keyword Search over Encrypted data on cloud.
In [3], Obtainable searchable encryption scheme consent to a user to firmly look for over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large joint data outsourcing cloud environment, they go through next shortcoming. Support only conventional Boolean keyword search without decrypting it. Single-keyword search without ranking, Boolean- keyword search without ranking and Do not get relevant data.

5] Ning Cao proposed Privacy-Preserving Multi Keyword Ranked Search over Encrypted Cloud Data.
In [5], proposed this search for known cipher text model and background model over encrypted data providing low computation and communication overhead. An Efficient and privacy preserving in Multi-Keyword Ranked Search over encrypted cloud data the coordinate matching is chosen for multi-keyword search. They usedinner product similarity to quantitatively evaluate similarity for ranking files. The drawback is that MRSE have small standard deviation which weakens keyword privacy.

6] Li, S. Yu proposed Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data.
In [6], defined and solved the problem of effective but safe and sound rank keyword search over Encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus making one step closer towards sensible consumption of privacy preserving data hosting services in Cloud Computing. These papers has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a protected cloud data utilization system to become a reality. The proposed ranking method proves to be efficient to go back extremely relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our future system in order to enhance the security of information on Cloud Service Provider

III. PROPOSED SYSTEM

We investigate the problem of maintaining the close relationship between different plain documents over an encrypted domain and propose a clustering method to solve this problem. We proposed the MRSE-HCI architecture to speed up server-side searching phase. Accompanying with the exponential growth of document collection, the search time is reduced to a linear time instead of exponential time. We design a search strategy to improve the rank privacy. This search strategy adopts the backtracking algorithm upon the above clustering method. With the growing of the data volume, the advantage of the proposed method in rank privacy tends to be more apparent. By applying the Merkle hash tree and cryptographic signature to authenticated tree structure, we provide a verification mechanism to assure the correctness and completeness of search results.

Secure Search Algorithm According to different data structures, search over encrypted data schemes may use different secure search algorithm to do the match. The inverted index structure allows fast direct intended file retrieval, so the search complexity is constant there. For example, the indexed keywords can be hashed and then store the associated file list at a table with its address being the hash value . When a user wants to search a keyword of interest, he/she first hashes it and submits the hash value to the server. Therefore, the server is able to find out the intended files efficiently. For schemes with index built from each document, the most efficient search algorithm merely enables linear search, i.e., the time for search is linear to the number of documents in the dataset, since the returned search results could not be determined until the search process goes through all the indexes within the document set. This is not desirable when a huge amount of data are present on the server. By utilizing tree-based structures to construct indexes for encrypted data search schemes, the corresponding secure search algorithm could be devised to achieve more efficient search than the linear search schemes. At the meantime, the same expressive queries as the schemes with index built per document could be realised under this index structure, such as range queries in database scenario and multi keyword text search with similarity-based ranking.

IV. CONCLUSION

In this paper, we investigated cipher text search in the scenario of cloud storage. We explore the problem of maintaining the semantic relationship between different plain documents over the related encrypted documents and give the design method to enhance the performance of the semantic search. We also propose the MRSE-HCI architecture to adapt to the requirements of data explosion, online information retrieval and semantic search. At the same time, a verifiable mechanism is also proposed to guarantee the correctness and completeness of search results. In addition, we analyze the search efficiency and security under two popular threat models. An experimental platform is built to evaluate the search efficiency, accuracy, and rank security. The experiment result proves that the proposed architecture not only properly solves the multi-keyword ranked search problem, but also brings an improvement in search efficiency rank security, and the relevance between retrieved documents.

FUTURE WORK

For the future enhancement system can use hadoop framework in multi cloud environment using load balancing approaches as well load rebalancing approach, because of our data is very, so need to balancing as well distribution.

REFERENCES

[1] Zhihua Xia et al,"Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL., N*O. 1.

[2] Cong Wang et al.,"Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE Transactions on parallel and distributed systems, vol. 23, no. 8, August 2012.

[3] Madane S.A, "Comparison of Privacy Preserving Single- Keyword Search and Multi-Keyword Ranked Search Techniques over Encrypted Cloud Data", 2014 International Jour nal of Computer Applications (0975 - 8887) Volume 126 - No.14, September 2015.

[4]Wenhai Sun et al., "Privacy-Preserving Multikeyword Text Search in the Cloud Supporting Similarity-based Ranking", the 8th ACM Symposium on Information, Computer and Communications Security, Hangzhou, China, May 2013.

[5]Ning Cao et al.,"Privacy-Preserving MultiKeyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, jan 2014.

[6]Ming Li et al,"Authorized Private Keyword Search over Encrypted Data in Cloud Computing", IEEE proc. international conference on distributed computing systems,June 2011, pages 383-392.

[7]A. Singhal "Modern Information Retrieval: A Brief Overview", IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

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

[9] Shih-Ting Hsu et al.,"A Study of Public Key Encryption with Keyword Search", International Journal of Network Security, Vol.15, No.2, PP.71-79, Mar. 2013.

[10] KuiRen et al.,"Towards Secure And Effective Data utilization in Public Cloud" IEEE Transactions on Network, volume 26, Issue 6, November / December 2012.