Review on Different Searchable Encryption Schemes for Cloud Computing

Mr. Hanumant B. Raut Mali
Department of Computer Engineering
RMD Sinhagad Institute of technology, Warje Campus, Pune
rauthanumant@gmail.com

Mr. Shrikant Nagure
Department of Computer Engineering
RMD Sinhagad Institute of technology, Warje Campus, Pune
shrikantnagure.rmdssoe@sinhgad.edu

***Abstract*—** Heavily available online data and its day to day expansion is need to be focus to store and retrieve it properly. This enforces the data owners tend to store their data into the cloud. This also suggest to handle the data properly and so release the burden of data storage and maintenance. But as the data owner and user, cloud server are not belong to same trusted domain, this may cause the outsourced to the risk. This enforce us to set the policy to avoid such risk factor. This gives us study scope to fine the different techniques to overcome such issue observed by different author. In this paper we try to underline the different solution, its limitation and results they achieved for retrieval of data securely and within less time. Definitely from this we will be able to propose our own solution.

Keywords- *cloud computing, cryptography, data integrity, network coding.*

*****

I. INTRODUCTION

The traditional information retrieval (IR) has already provided multi-keyword ranked search for the data user. In the same way, the cloud server needs provide the data user with the similar function, while protecting data and search privacy. It is meaningful storing it into the cloud server only when data can be easily searched and utilized.

The research done on TF IDF to calculate total relevance score is studied [1]. Wang C. has [2] presented one-to-many orders preserve mapping for avoiding file relevance score from the server for single keyword search. After that research presented by Wenhai Sun [3] on preserving privacy for multi keyword text search scheme with similarity based ranking. Then a novel multi keyword fuzzy search scheme by exploiting the locality-sensitive hashing technique is proposed by Bing Wang [4]. The schemes to deal with secure ranked multi-keyword search in a multi-owner model presented by Wei Zhang [5]. David Cash [6] have implemented the four several factors ignored by earlier coarse-grained theoretical performance analyses, including low-level space utilization, I/O parallelism and good put.

II. LITERATURE REVIEW

In the first paper The Gengiz and Savas [1] have produced the TF and IDF for calculating total relevance score is not efficient score ranking function for multi-keywords and is unable to implement the multi- keyword disjunctive Boolean operation. Produced the TF and IDF for calculating total relevance score is not efficient score ranking function for multi-keywords and is unable to implement the multi-keyword disjunctive Boolean operation. The index also contains the single keyword file score. For every file one new index is required so to search the index need to all the indexes that requires the more time. So need to improve the index structure and the relevance ranking function.

Wang C. [2] the need is to prevent the sum of the relevance score of the files for the multiple keywords. So we improve the indexing and ranking function to prevent the sum of score of multiple keywords. We improve the system for multi-

keyword search in terms of the time to search and the advanced relevance score. It uses the one to many orders preserving mapping to prevent the file relevance score from the server for single keyword search. If we need to use the multiple keyword search then the sum of the relevance score of the files need to be calculated to retrieve the files. The need is to prevent the sum of the relevance score of the files for the multiple keywords. So we improve the indexing and ranking function to prevent the sum of score of multiple keywords. We improve the system for multi-keyword search in terms of the time to search and the advanced relevance score.

In the paper, Wenhai Sun [3] have implementation of the proposed tree-based search algorithm on a real-world document set: the recent ten years' INFOCOM publications. The document set is built from the recent ten years' IEEE INFOCOM publications, including about 3600 publications, from which we extract about 9000 keywords. Wenhai have build the search index based on term frequency along with the vector space model by means of cosine similarity measure for achieving higher search result accuracy. For improving the search efficiency, he had proposed a tree-based index structure. Various adaption methods for multi-dimensional (MD) algorithm so that the practical search efficiency are much better than that of linear search. This author presents a

privacy-preserving multi-keyword text search (MTS) scheme with similarity-based ranking to address this problem. To support multi-keyword search and search result ranking, we propose to build the search index based on term frequency and the vector space model with cosine similarity measure to achieve higher search result accuracy. Proposed algorithm practical search efficiency is much better than that of linear search. To further enhance the search privacy, we propose two secure index schemes to meet the stringent privacy requirements under strong threat models, i.e., known Ciphertext model and known background model. Finally, we demonstrate the effectiveness and efficiency of the proposed schemes through extensive experimental evaluation.

Bing Wang [4] used the recent 10 years' IEEE INFOCOM publication as our experiment dataset which contains more than 3600 files. We extract 5734 keywords in total, and the average number of the keywords in a paper is 147 while the minimum and the maximum are 112 and 175 respectively. This scheme achieves fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need of a predefined dictionary and effectively supports multiple keyword fuzzy searches without increasing the index or search complexity. Extensive analysis and experiments on real-world data show that our proposed scheme is secure, efficient and accurate. This paper, we propose a novel multi keyword fuzzy search scheme by exploiting the locality-sensitive hashing technique. Our proposed scheme achieves fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need of a predefined dictionary and effectively supports multiple keyword fuzzy searches without increasing the index or search complexity. Extensive analysis and experiments on real-world data show that our proposed scheme is secure, efficient and accurate. To the best of our knowledge, this is the first work that achieves multi-keyword fuzzy search over encrypted cloud data. The need of a pre-defined dictionary is a limiting factor that makes dynamic data operations, such as dataset/index update, very difficult.

Wei Zhang [5] has conducted the experiments on a real data set, the Internet Request for Comments dataset (RFC). This dataset has 6870 plain text files with a total size about 349MB. The average size of each file is 52KB. The file size of this data set is demonstrated. Author used Hermetic Word Frequency Counter for extracting keywords from each RFC file. The keyword frequency of this data set is also shown in the paper. Wei Zhang has implemented Additive Order and Privacy Preserving Function family technique for ranking the search results as well as preserving the privacy of relevance scores between keywords and files. Also Zhang has implemented the secure search protocol for searching without knowing the actual data of both keywords and trapdoors. We propose schemes to deal with secure ranked multi-keyword search in a multi-owner model. To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, we systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel Additive Order and Privacy Preserving Function family. Extensive experiments on real-world datasets confirm the efficacy and efficiency of our proposed schemes. This scheme is limited to the single-owner model. We show that our approach is computationally efficient even for large data set and keyword set.

David Cash [6] used databases derived from the ClueWeb Collection or synthetically generated by an engine trained on US-census data. The key attributes of these databases and derived encrypted indices are summarized. Both database families contain atomic type and text columns. The ClueWeb databases were encrypted for a multi-client setting supporting conjunctions (OXT) and the census database where processed for single keyword search (SKS), also in multi-client settings. This implementation effort brought to the fore several factors ignored by earlier coarse-grained theoretical performance analyses, including low-level space utilization, I/O parallelism and good put. We accordingly introduce several optimizations to our theoretically optimal construction that model the prototype's characteristics designed to overcome these factors. We evaluate the performance of our prototype using two very large datasets: a synthesized census database with 100 million records and hundreds of keywords per record and a multi-million webpage collection that includes Wikipedia as a subset. Moreover, we report on an implementation that uses the dynamic SSE schemes developed here as the basis for supporting recent SSE advances, including complex search queries (e.g., Boolean queries) and richer operational settings (e.g., query delegation), in the above terabyte-scale databases. We found its performance to be limited by its database access patterns. Future work will shed more light on the best ways to design such masking techniques.

Qin Liu [7] proposed in this paper that the search that provides keyword privacy, data privacy and semantic secure by public key encryption. CSP is involved in partial decipherment by reducing the communication and computational aerial in decryption process for end users. The user submits the keyword trapdoor encrypted by users‟ private key to CS (Cloud Server) securely and retrieves the encrypted documents. Limitation: - The communication and computational cost for encryption and decryption is more.

IJCA [8] Obtainable searchable encryption scheme consent to a user to firmly look for over encrypted data through keywords without first applying decryption on it, the

proposed techniques support only conventional Boolean keyword search, without capturing any applicability of the files in the search result. When directly applied in large joint data outsourcing cloud environment, they go through next shortcoming. Limitations: - Single-keyword search without ranking. Boolean- keyword search without ranking. Do not get relevant data.

Jiadi [9] proposed this search using two round searchable encryption (TRSE). In 1st round, users submits multiple keyword 'REQ' 'W‟ as encrypted query for accomplishing data, keyword privacy and create trapdoor (REQ, PK) as Tw and sends to cloud server. Then cloud server calculates the score from encrypted index for files and returns the encrypted score result vector to user. In second round, user decrypt N with secret key and calculates the file ranking and then request files with Top k scores. The ranking of file is done on client side and scoring is done on server side. Limitation: - The contraction and confining is used to reduce cipher text size, still the key size is too large. The communication aerial will be very high, if the encrypted trapdoor's size is too large. It does not make effective searchable index update.

Ning [10] proposed this search for known cipher text model and background model over encrypted data providing low computation and communication overhead. The coordinate matching is chosen for multi-keyword search. They used inner product similarity to quantitatively evaluate similarity for ranking files. The drawback is that MRSE have small standard deviation σ which weakens keyword privacy. Limitation: - Multi-keyword ranked search (MRSE) for known cipher text model may produce two different trapdoor which vague the privacy leakage problem of trapdoor unlink ability which may weaken the keyword privacy. MRSE has small standard deviation σ which in turn weakens the keyword privacy. The integrity of the rank order is not checked in MRSE.

III. RESEARCH GAP

This proposed method has defined and solved the problem of effective but safe and sound rank keyword search over Encrypted cloud data [11]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain important criteria (e.g. keyword frequency) thus making one step closer towards sensible consumption of secure data hosting services in Cloud Computing. These papers has defined and solved the challenging problem of privacy preserving and efficient multi keyword ranked search over encrypted cloud data storage (MRSE), and establish a set of strict privacy requirements for such a protected cloud data utilization system to become a reality. Limitation: - Dynamic updating and deletion of the document from the cloud is not possible.

IV. CONCLUSION

From this literature we are able to conclude that privacy of data is most challenging issue. Different proposed solution have their own limitation. We observed that search using indexing on encrypted data provide efficient results. Also we will be able to obtain more security as it provides the search on secured data. From this someone get the research scope to enhance the better mechanism over indexing for less time required for search. In the literature, searchable encryption techniques are able to provide secure search over encrypted data for users.

V. FUTURE ENHANCEMENTS

Numerous researches have been done on the multiple keywords searching with preserving its privacy which are very helpful for study or make another research on it. System for multi-keyword search in terms of the time to search and the advanced relevance score

REFERENCES

[1] Cengiz and Savas “Efficient and Secure Ranked Multi-Keyword Search on Encrypted Cloud Data” PAIS 2012, March 30, 2012, Berlin, Germany Copyright 2012 ACM 978-1-4503-1143-4/12/03

[2] Cong Wang, IEEE, Ning Cao, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data" IEEE transactions on parallel and distributed systems vol.23 NO.8 YEAR 2012

[3] Wenhai Sun et.al, “Privacy-preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking”, ASIA CCS’13, May 8–10, 2013, Hangzhou, China. Copyright 2013 ACM 978-1-4503-1767-2/13/05

[4] Bing Wang et.al, “Privacy-Preserving Multi-Keyword Fuzzy Search over Encrypted Data in the Cloud”, IEEE INFOCOM 2014 - IEEE Conference on Computer Communications

[5] Wei Zhang et.al, “Secure Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing”, 2014, 978-1-4799-2233-8/14 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks

[6] David Cash et.al, “Dynamic Searchable Encryption in Very-Large Databases: Data Structures and Implementation”, NDSS’14, 23-26 February-2014,San Diego, CA, USA Copyright 2014 Internet Society, ISBN 1-891562-35-5, http://dx.doi.org/10.14722/ndss. 2014.23264

[7] Qin Liuy, Guojun Wangyz, and Jie Wuz, ”Secure and privacy preserving keyword searching for cloud storage services”, ELSEVIER Journal of Network and computer Applications, March 2011

[8] International Journal of Computer Applications (0975 – 8887) Volume 126 – No.14, September 2015

[9] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Member, IEEE Computer Society, and Minglu Li,”Toward Secure Multikeyword Top k Retrieval over Encrypted Cloud

Data", IEEE Journal of Theoretical and Applied Information Technology 10th August 2014. Vol. 66 No.1 © 2005 - 2014 JATIT & LLS. All rights reserved. ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195 64 Transactions on dependable and secure computing, vol. 10, no. 4, July/August 2013

[10] 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

[11] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang, "A Secure and Dynamic Multi-Keyword Ranked SearchScheme over Encrypted Cloud Data", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 2, FEBRUARY 2016