DOI: http://dx.doi.org/10.26483/ijarcs.v8i7.4266

ISSN No. 0976-5697

Volume 8, No. 7, July – August 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

SECURE SEARCH OVER ENCRYPTED DATA TECHNIQUES : SURVEY

Swati T Harane
Bharati Vidyapeeth Deemed University
College of Engineering, Pune

Prof. Gajanan Bhole
Bharati Vidyapeeth Deemed University
College of Engineering, Pune

Prof. Milind Gayakwad
Bharati Vidyapeeth Deemed University
College of Engineering, Pune

*Abstract***:** This paper presents the need of storing data on the cloud servers to overcome the load of the database and its maintenance The concept of customers of the storage and server of cloud feature is completely different they even don't lie in the same trusted domain ,the outsourced data is expected to be under the risk . The main difficulty that is arising in the multi watchword search is providing the relevant score for the considerations of file for the use of multiple watch words. But the required time for the same search is more as thought .Therefore multiple researches have been considered and are being compared to understand the concepts in a good manner.

Keywords: ENCRYPTED, information, search, dataset, algorithm

INTRODUCTION

The old form of retrieving information has a provision for the multi-watchword search dor the database user and its allocated server. Following the similarity, the server of the cloud needs to be allocating the user with the functions and all the similar facilities. If we want to search the data easily and due to the proper safety purposes it is mandatory to store the dataset in the cloud server only. The experiments performed on the two techniques called as TF and IDF to do calculations on total relevance score is studied [1].In the given scheme Wang C. [2]has represented one-to-many orders to show the working of preserving mapping facility for avoiding the file relevance score from the server for single watchword search. After the Wenhai Sun [3]theory of maintain the security for multi watchword text search , further calculations were made. A *multi watch word* search scheme is presented in this paper which is build by forbidding the sensitive technique know as hashing which was given by Bing Wang[4]. To deal with safe ranked multi watchword search , two new techniques are provided by Wei Zhang[5] The factors ignored by previous versions performances analyses including the space utilization , parallelism in input and output is implemented by. David Cash[6].

LITERATURE SURVEY

Here , the research proposed by two men Gensiz and Savas[1] who have build the two systems known as TF and IDF for making some calculations to show total relevance score is not efficient score ranking system for multi-watchwords and isn't able to watchwords for the disjuncture true and false commonly known as boolean operation. In this research paper the builded TF and IDF for working out the total relevance score isn't upto the mark score ranking function for multi-watchwords. The main index also has only one watchword score of the file and for every other file the other index is used to search the index which is needed for all the indexes that requires the more time. Therefore the main index structure and the ranking function needs to be modified and make it more efficient and improved in order to achieve high relevance score.

In the old system proposed by Wang C[2] explains to forbid the total relevance score of the files for the multiple watchwords. Therefore the main index structure and the ranking function needs to be modified and make it more efficient and improved in order to achieve high relevance score.

Hence to modify and improve the function for multi-wachwords search in terms of advanced relevance score, the system uses the one to many cardinality orders preserving mapping to forbid the file relevance score from the server for single keyword search. Therefore to use the multiple watchword search and to retrieve the files the sum of the relevance score of the files need to be calculated. There is mandatory need to forbid the total of the relevance score of the files for the multiple watchwords. Hence modifying and improving the main index and rank system to forbid the total of score of multiple watchwords is mandatory.

The current paper, explained by another great man Wenhai Sun[3] has executed one of the already explained tree-based search algorithm on real world objects which was taken from the the last ten years' INFOCOM publications. The INFOCOM publications real world document set is created from the past ten years' IEEE INFOCOM publications, which includes about 3600 publications, from only about 9000 watchwords were extracted. To achieve greater search result accuracy Wenhai[4]has build one more of the index known as the search index. This is done by the use of cosine similarity measure. For achieving the higher search efficiency, he explained a tree-based algorithm. The writer explains the importance of maintaining the privacy of the multi-watchwords text search scheme Multi-watchword search and result ranking search was supported by a search index based on frequency and vector space with the cos angle similarity to quip with a high ranking accuracy. The traditional linear search can be replaced by a ore efficient practical search algorithm.To make the search privacy more improved we give an idea of two more secured indexes to cope up with all the requirements called as cipher text and background model respectively. Therefore lastly we give a demonstration of the cipher text and background models by mentioning and performing some calculations on some data.

Our dataset used for our calculations is the same dataset that was used by Bing wang[5] from the last 10 years IEEE INFOCOM publications. The total watch words extracted by us is 5733 and the average number of watchwords in the research paper is 145. The maximum and minimum watchwords presented in the paper is 110 and 173 respectively. The writer of the paper tells that we can build a mix matching by the correct design of algorithm and there is no need of expanding the index file. The proposed scheme forbids the use of a dictionary and supports the multi-watchword search and that too without expanding the index. The complete experiment on our scheme proves that the scheme given by us is more safe and accurate. The scheme proposed in our paper is very much efficient of practical matching and that too without expanding the index file. And as it is not expanding the index file the maintenance cost is also cut down. There fore it proves that our experimented scheme is accurate, efficient and safe upto the mark . According to our expectations this is the very first work that achieves all the requirements which would be very helpful in the near future while dealing with the cloud server.

Wei Zhang has also experimented and performed the calculations on the same dataset used by us in our scheme but this time he used internet request for comments also .The new dataset contains 6865 text files which has a size of approx 347MB. A system known as hemantic word frequency counter is used for extracting watchwords from the RFC file. The same method is also depicted in our paper. The algorithms of Additive Order and Privacy Preserving Function family techniques is also used by Wei Zhang A safe search rule set for the search without knowing the details of the actual dataset of watchwords and gateway was implemented by Zhang. Our schemes deal with the multi-watchwords search with the use of a multi owner model. A novel secure search protocol is also constructed to let the cloud servers perform a safe search. The efficiency of our given schemes are checked by calculations performed on real database but this scheme is only valid for the single owners . our approach proves to be efficient for the large database.

The another great men David cash [6]used the datasets that was artificially prepared by the use of the US census data provided by the ClueWeb. The semantic functions and features of this database contain atomic text columns .This database was made for the multiple clients which supported the conjuctions and the database that was provided by the clueweb was processed for the single watchword in the multi customer settings. This effort was able to bring several factors which was ignored by the earlier theories into action.These efforts includes low level space and accurate output., Therefore several modifications was implemented to our theory for the optimal use of the prototypes characteristics which were specially designed to overcome all the factors. The performance of our system is calculated by using two very large datasets: a maintained database with approx 100 million records and lots of watchwords per record. Wikipedia was used as a subset while creating the collection of the webpages. Moreover, we focused 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 observed that the performance to be limited by the observation in the database access patterns. The work that will be done in the future eill be focused more on more better and optimal ways to design such techniques.The last paper by David Cash offered an optimal server index size which was completely parallel in search and was approx with no leakage . Many modifications were already made to overcome such factors and design an optimal design of such masking technique

The theory given by Qin Liu[7] in this paper that the search that is efficient with watchword security , data security and semantic security by public key encryption. The communication is reduced for end users by using partial decipherment by the use of CSP. The private key encrypted by the users is submitted to the cloud servers safely and regains the documents.

Limitation: - The cost of the communication and computation is high,the cost of encryption and decryption of data is also very high.

The scheme[8] of searchable encryption scheme is user firmly in look for the encrypted data before applying decryption on it.the traditional scheme only supports the Boolean operations on it., without capturing any applicability of the files in the search result. If directly applied in larger database which outsource the cloud environment they go through the next coming of shorter duration.

Limitations: - we donot get relevant data in single – watchword search as it is searched without ranking process.

This search was proven by Jaidi[9] and was proven to be exploited in 2 spherical search encodings In first one the user provides the multiple watchword 'REQ' 'Ẅ as encrypted question for achieving the knowledge, watchword security and make gateway (REQ, PK) as Tw and then redirects it to the cloud server. Then in the cloud server the score is calculated from the encrypted index for the usage of files. Cloud server also redirects the result to the client or the user. The second spherical calculates the file ranking with k prime scores using the N key decipher. After the second spherical gets completed it is expected that the file ranking is completed on the customer side and the rating is completed on the server side.

Limitation: - To cut the cipher text size methods like contraction and confining is used but till then the key size is larger than the requirement. The cost of the communication is very high and if the gateway's size is too large it will not create effective results.

Another theory for the showing of cipher text model and background model using low communication cost was provided by Ning[10]. For this the coordinate matching is performed for the multi watchword search which uses real number similarity to quantitatively assess similarity for ranking files.

Limitation: - Two different gateways are created which exploits the security leakage problem of gateway unlink ability which proves to be weakening the watchword security. MRSE has small standard deviation σ which in turn weakens the keyword privacy. There is a problem of integrity of ordering of the rank..

The planned strategy has helped to solve the subject of safe rank watchword search over the safe cloud knowledge.This kind of search greatly able the system reusability by returning the matching files in an ascending order which is related to the safe requirement so build one more step towards good use of secure knowledge in Cloud Computing. These papers mentioned here has fulfilled and solved the difficult aspect of security which saves the economical multi watchword.

CONCLUSION

It ended that searchable secret writing techniques are ready to offer safe search and safe and secure knowledge for clients . It builds a reversed index which can be searched and that stocks lists of mapping from watchwords (watchwords) to the files which are related to their files that contain these watchwords. Once clients uses the input a gateway is generated for these watchword (watchwords) and submitted to the cloud server. The comparison between the gateway and main index is done by the cloud servers, and at the end it returns the information of each user and each file which contains these watchwords (watchwords) upon receiving the gateway.

FUTURE SCOPE

Several researches have been done on the multiple watchwords searching with retaining its privacy which are very helpful for study or make another research on it. This paper explains the system of keyword search to provide the modified 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