AN EFFICIENT PRIVACY-PRESERVING RANKED KEYWORD SEARCH METHOD

E. Nirmala¹, S. Muthurajkumar², D. Subitha³

E. Nirmala Corresponding Author

1. Research Scholar, Department of Computer Technology, Madras Institute of Technology (MIT) Campus, Anna University, Chennai, India. niralalochan@gmail.com
2. Assistant Professor, Department of Computer Technology, Madras Institute of Technology (MIT) Campus, Anna University, Chennai, India.
3. Assistant Professor, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Vel Nagar, Chennai. subithaece@veltech.edu.in

Abstract

The keyword searching algorithm plays a vital role in cloud computing Architecture. Here, the searching technique is applied as a tool to encrypted data set which is outsourced. This technique works well, by either directed on multi keyword findings with exact match or one keyword search with fuzzy logic with a result of true or false. Though the existing techniques come across sensible difficulty in execution upon the encrypted data set, in the execution of multiple keyword based searching, keyword correction is difficult to achieve. To overcome with this problem, we construct a highly efficient, multiple corrections in keyword searching algorithm using ranking scheme. Multiple corrections like spell checks and errors. In the existing system of Author Wang and others scheme, they were unable to address the aforementioned problems, But in our method of multiple keyword searching technique, the Fuzzy Gramm algorithm helps to detect and handle the spelling errors by making a function call. Additionally the keywords with similar ancestry can be found by making use of stemming procedure. Alongside with that rank keyword is used to achieve accurate checking with matched text file content. Thus the proposed approach enhances security and reliability to the maximum extent compared to the existing system.

Keywords: Fuzzy Gramm algorithm, Bloom Filter Algorithm, query base searching, multiple keyword searching algorithms

1. INTRODUCTION

In the cloud environment more sensitive information of the corporate and business has put forth in the public medium, this could have certain difficulties in maintenance and control over the data. To resolve the problem data is outsourced. This improves the protection on data content. This motivation leads to the development of keyword searching algorithm in cloud. The searching of the keyword is carried through the precise keyword. In the traditional algorithms, but the complexity in executing the algorithm was comparatively less, due to it's the cost and time. Effective keyword searching algorithms is used based on the previous search of the user by considering the semantic and distance of the enumerated the keyword [1] the proposed search algorithm was purely based on the correspondence checking with the usage of advanced techniques like fuzzy identification. By means of sets of keyword related to potential identification of user related technique like storage complexity by having redundant data sets been reduced to the maximum extent. However, it was very difficult for the data owners to search the keyword over the encrypted data. The advanced communication leads to high computational overhead [2]. To reduce the cost of infrastructure the software business firms used to outsource the data. This mitigates the risk of ownership too. But secrecy is a question mark. In this scenario the new technique called minhash related functions been introduced. This minimizes the risk by using many keywords...
searching with the help of single query. This proposed technique is proven based on the term inverse document frequency. By applying constraints using query author proved the efficiency by preserving the secrecy of the stored data content[4]. Even though the privacy is preserved using minhash functions, for sensitive content like electronic health records and financial data like public insurance and all has heavy overhead. to overcome with this problem data structure algorithm like B+ tree is used to calculate the competence of querying. It used to calculate the rate at which query could be executed over the cloud data within a stipulated span of time. To enhance the secrecy towards the data KNN based algorithm is used over the encrypted data. this is reduces the index value of the cloud storage medium and also increase the level of extraction[5]. Even though privacy is preserved using B+ tree algorithms, the cost incurred towards management of this process and ease of accessibility is the crucial challenging factor in future to be visited. To achieve this Secrecy preserving multi keyword searching algorithms has been introduced.

This secrecy preserving scheme is going to be verified before starting the search based on the keyword by using rank. This is implemented along with Multi Dimensional processing Statements. To ensure the security once again, indexing schemes has been used. This method is evaluated, hence proven for its enhanced security by preserving its sensitive data content [6]. Using the images the story could be learnt by the children. For this efficient technique related to stories could be narrated using images. This laid down the path to generate the relevant images to the words dynamically by using machines. To achieve this kind of implementation natural language handling techniques has been used. To narrate the story using words, two steps are followed in this scenario. Firstly, identifying the keywords related to the characters in the story, along with its situation. Secondly image capturing progression will be carried out with the help of deep learning model for every image retrieved with multimedia Data Base. Moreover Google search engine is also used. Through this methodology, index, purify the validation of the images are visualized for safe and flexible environment [7]. Drag the most visited website using the technology called deep learning. This helps to identify the hidden files too. The objective of such an environment is to provide a special feature to identify the needed information not only through the used links, or history of drag and drop instead using the deep learning methodology dragging could be done in deeper manner using the various links and country codes with the means of URL’s. Efficiency in execution is achieved by suitable ranked searching based on the requested keyword [8].

2. LITERATURE SURVEY

Cheng Guo et.al.,[9] proposed a system to provide secured searching algorithm to enhance the security over the encrypted data by using Bloom Filter. While the data is sent for outsourcing in the cloud need to undergo the process of encrypted security, which is not available in the traditional system, So overcome with the existing problem, author proposes a new model by utilizing the storage space to enhance the operations during run time. To achieve this feature multi keyword based ranked search algorithm is used with reduced cost with maximum efficiency. Cost reduction is the advantageous feature over the existing approach. Bloom Filter technique is used to generate the index search tree, which helps to reduce the cost. In future refinement for Bloom’s Filter methodology could be developed to minimize the cost and complexity in execution.

MomoKyozuka et.al.,[10] Proposed a system to search for a book with the depiction of story inside. The users may search for a specific book based on the title with the description, may have a little bit idea about the story used as an example, in such a situation author introduced query relaxation scheme, for relevant search with vague memory on keywords. If not finding a relevant book from the data base may be difficult. This retrieval based on searching algorithm is executed successfully by forecasting the manifestation of the keyword search using random probabilistic technique. Here the subset of the searching keyword is taken into consideration to calculate the rank. After finding the rank based on the relevant description, even though user has a vague memory could be identified with no of searches made during stipulated span of time and decision is arrived with a probabilistic approach. In future, the queries relaxed based on vague memory could be used along with ranking based relaxed
queries to reduce the projected rank of the user data. This could be used in different situations where users wish to generate many queries from the basic query.

Sanjumol et.al.,[11] Proposed a General Location Aware Ranking queries model to equip the user with a spatial object, keyword to predict the purpose of ranking. Here the calculative part of rank is computed based on the spatial position of the object, textual message and the no of entities related to the requested data in the data base. The author introduces a technique like nearest location based querying. While, querying about the hotel nearest neighboring hotels will also be provided along with the added information like provisional store, theatre, health centre, and gym. Author constructed this work with three structures like R-tree, reversed files and notion tree. Whereas R-tree is used to process the location where the data is related to, Reversed file structure helps to track the keywords from the internet resources, and the notion tree will hold the information connected to mathematical data. The combination of aforementioned structure satisfies the user needed information by using ranked searching algorithms constructed as a framework. The advantage is reduced time for ranked search based on the factor location identification. In future the reversed file system is combined with filtering technique to reduce the searching time and complexity.

3. System Architecture:

![Fig.1. System architecture of query base searching](image)

Data Owner: The Data Files of Owners like wider organizations, business firms or Information technology oriented industries used to transfer to cloud storage, to minimize the cost of a storage device as well as mitigate the risk associated with security. Nowadays, while sending the data to the cloud environment, the owner of the data encrypts the content to safeguard from threat. For this, in our system the provision is made to the Data to ensure the privacy of the data, encryption is permitted along with his original data of information repository IR.

Cloud Server: Whenever, the Cloud Server receives the request from the Data Owner, immediately the request is processed, with the help of the query processor by exploring the organized methodologies. Technically with the help of searching techniques called tree traversal. Indexing is the factor upon which each of the text is checked on the basis of similar key words. Sub tree level checking is carried out to get the right text file requested by the user. As soon as, the user requested data is found, immediately reported to the Data Owner after thorough checking. This ensures the confidentiality with the confirmation factor like key value with secret sharing methodology.
User Query Processor: Whenever the user needs to access the data from the cloud, he has to get the access permission from the cloud server for his authorized verification. This has been implemented through a control mechanism kind of secured key word. After such user verification, the index checking is allowed to the user data. This factor of indexing prevents the data from the third party access. The query processor accepts the data from the cloud server, and processes it with verification procedure, and allows checking for the index with assurance of security constraint with permissible guarantee like a keyword exchange mechanism.

Cloud Data Base: Cloud data base used to store the encrypted outsourced user information to check for authorization. By generating the query user can access the data. Before retrieval of data, the data undergoes for stringent testing like owner authorization. Top secret key is shared by both the parties to prevent from malicious access. The algorithm for top secret key tsk searching algorithm is used to retrieve the data accurately. This rank based multiple keyword searching algorithm is executed with less time and cost complexity by means of tree structural searching.

4. Algorithm Construction

The binary tree is used to search the keyword based on top secret key tsk. The data structure of a node N is defined as \{FID, Du, Blf, Ic, rc\} in which the left child and right child represents lc, rc. Whenever the File is identified based on the user authorization called as tsk, by searching the leaf node ln with a Bloom Filter structure Blf for every user requirements. The Bloom filtering data structure helps to find whether the needed element is available in the set or not by utilizing a reduced time and space complexity. This is coded with hash function \(Hf(i=1, 2, \ldots\ldots hr)\) where the index value of \(Hf\) can be \{0, 1\} \({}\rightarrow\{1, bf\}\. Where every bit position can be mapped with \(hr(1), hr(2), hr(3), \ldots, hr(e)\. With the above rules along with that the \(hr(e)\) value is set to one. While, we wanted to know whether the element \(e\) belongs to the set \(S\), for which we can just check for the bit position with the corresponding value set with one. According to Bloom Filter algorithm, if the bit position is set to zero, then it signifies the chosen element completely does not belong to the set \(S\). This is called as positive false indication. In this scenario, false positive indication of an element \(e\) could be avoided by setting the valuable parameters in a right position of hash table.

4.1 Steps for Construction of Index Tree are as follows;

Step 1: makeIndexTreeTi (i=1, 2, 3, \ldots m) in the document set, create a equivalent leaf node for it by following procedures

Step 1.1: Create a un Document Identification for the set Docti, and define the identifier as (FILEID) of the set S as last child Nodecn.

Step 1.2: Produce the Indexed Vector IDXV for the word to be found in keyword Dictionary KWD for which \(w_f1, w_f2, \ldots, w_fm\) Where the KW- is what the key word lengthiness, this is the length of the Dictionary Key word. Where every aspect of Dictd, [j] is the methodized with term frequency vectorTFV with value of \(W_j\) in the document docti.

Step 1.3: Create the Bloom Filter Algorithm BFA for docti by using the relevant hash function hashsk, where hashsk{0, 1\} \({}\rightarrow\{1, bf\}, because of this, document d with its relevant keyword and their positions of BFdocti, with its relevant hashsk(1), hashsk(2), hashsk(3), hashsk(4), hashsk(5), \ldots, hashsk(constk), will be set to 1.

Step 2: Create the Index Tree for searching the key word In whose end child nodes are mcnodes in the step which is stated above. The FileID for the document Dictij and BFdocti, of the node of internal in natures are initially set to null or zero.
Step 3: For each arrival to the internal node INTN, BFA, could be calculated based on Bloom Filter’s Algorithmic rule called as left, right and root node recurrently. More specifically BFA$_a$[i], BFA$_a$[j] or BFA$_i$ where “or” is used as a Boolean Operator as OR functionality.

Step 4: The output of the index tree is one based on the execution.

4.1.1. Algorithm for searching in the Index Tree(BFA$_k$; TreeIndexNodecn)

Step 1: if the child node cn is the internal node then
Step 2: Initializing the count with zero
Step 3: for every I of KWD1 to the length of BFA$_u$[i] IndexVectorIV1) then
    count++;
Step 4: end if
end for
Step 6: if (count > pkey/a)
    IndexTreeSearching(BFA$_k$, u, l);
    IndexSearchTree(BFAIndexTreeSearching(BFA$_k$, u, r);
end if
Step 7: return the present node to u;
end if.

4.1.2. Highly Secured Algorithm with Update procedure:

Step 1: Generate the top secret key tsk to the lengthiness of set S becomes (ni+Un)bits and Mas1, Mas2 are changed to (ni+Un)+(ni+Un) 2disjoint matrices.

Step 2: When Creating the Index Vector IV1 in an aforementioned step2 of Tree Construction Algorithm with a function called makeIndexTreeTi (i=1,2,3,……m) is also extended to (ni+Un) with the extended dimensions, with the encryption of searched document based on rank doctdtj (1,2,3,………Un).

Step 3: A Set is assigned with a random value one. Before establishing trapdoor to enhance he Security by extending the query vector to (n+Un) positions with variable position called as VP with the extended dimension called as EDIm.

Step 4: Compute similarity among m variable with n different dimensions.

5. Performance Analysis

With an above stated algorithm we can extent maximum security to the data with index tree creation function based on the keyword searching concept. Enhanced security could be provided by generating top secret key tsk by verifying the requirements in detail. By using the Bloom Filter algorithm we can calculate randomly selected value to the variable set to one. With this secret way to enhance security guarantees the secured keyword searching based on the user request is the result we achieved with the comparative study of Basic Model without any data leakage.
| No. of Keywords | Time Consumption Existing Approach | Time Consumption New Approach |
|----------------|----------------------------------|-------------------------------|
| 0              | 7                                | 7                             |
| 2              | 4                                | 4                             |
| 3              | 6                                | 6                             |
| 4              | 8                                | 8                             |
| 5              | 11                               | 11                            |
| 6              | 18                               | 17                            |
| 7              | 22                               | 21                            |
| 8              | 26                               | 25                            |
| 9              | 34                               | 31                            |
| 10             | 32                               | 35                            |

Table-1 Cost and Time Generation of a Secrecy Approach

Fig.2 The Time cost of generation of Top-Secret Key Gap between Existing System with New Modern Approach with KW size of 10

| No. of Keywords | Time Consumption Existing TSK Approach | Time Consumption New TSK Approach |
|----------------|--------------------------------------|----------------------------------|
| 5              | 13                                   | 13.9                             |
| 10             | 13.1                                 | 14.1                             |
| 15             | 12.8                                 | 14.2                             |
Table 2 Cost and Time Generation of a Secrecy Approach

| No. of Documents to be searched | Keyword Searching Time in Existing Approach | Keyword Searching Time in Proposed Approach |
|--------------------------------|--------------------------------------------|--------------------------------------------|
| 10                            | 12                                         | 14                                         |
| 15                            | 20                                         | 22                                         |
| 20                            | 24                                         | 26                                         |
| 25                            | 25                                         | 27                                         |
| 30                            | 26                                         | 28                                         |
| 35                            | 27                                         | 29                                         |
| 40                            | 29                                         | 31                                         |
| 45                            | 30                                         | 33                                         |

Fig. 3 Query keywords of varied size available within the dictionary of keywords
In the aforementioned figure 2 and 3 Table 1 & 2 proves better performance using Java on Linux Server with Intel 2.9GHz Processor. This data set consists of 1,29,000 different abstracts with the description of NSF awards 1990-2003. We randomly chose the different files among the dataset to check for the availability using rank-based searching, compared with the existing conventional system and hence our system has proven with the better result with performance accuracy of 99% and reduced time consumption with 5% less than the conventional model with high compatibility.

6. Conclusion

Here our proposed approach proven with the rank based search scheme over the conventional cryptographic system with multiple keyword searching algorithms. Hence, dynamically the results were generated with high efficiency. This multi keyword searching algorithm works well with Searching of a Index Tree based on the rank. For this Bloom Filtering Technique is implemented to justify the work. Finally the above generated algorithms helped to achieve the design goal of our system with 99% accuracy and 5% reduced execution time as compared to the conventional algorithms.
References

1. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., and Lou, W., "Fuzzy keyword search over encrypted data in cloud computing" In Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM’10). pp. 441–445. IEEE, San Diego, CA, USA (2010)

2. Wang, C., Ren, K., Yu, S., Mahendra, K.” Achieving usable and privacy-assured similarity search over out-sourced cloud data” In: Proceedings of the 31th IEEE International Conference on Computer Communications (INFOCOM’12). pp. 451-459. IEEE, Orlando, USA (2012)

3. E. Ramprasath, P. Manojkumar and P. Veena, "Induction motor analysis using labview", Proceeding on International Conference on Electrical Engineering and Technology, vol. 2, no. 5, pp. 498, 2015.

4. Cheng Guo, Pengxu Tian, Chin-Chen Chang, "Privacy preserving weighted similarity search scheme for encrypted data" in the journal of proceedings of the Institution of Engineering Technology, IET, Vol. 13 Iss. 1, pp. 61-69,2018.DOI: 10.1049/iet-ifs.2018.518

5. Jian Xu, Xinyu Huang, Geng Yang, Yuanyuan Wu “,An Efficient Multi-keyword top-k Search Scheme over Encrypted Cloud Data”, In the proceedings of the 15th International Symposium on Pervasive Systems, Algorithms and Networks IEEE 2018.DOI 10.1109/I-SPAN.2018.00059

6. Lanxiang Chen, Zhenchao Chen, Kim-Kwang Raymond Choo, Chin-Chen Chang, Hung-Min Sun”, Memory Leakage-Resilient Dynamic and Verifiable Multi-keyword Ranked Search on Encrypted Smart Body Sensor Network Data”, in the journal of Advanced Software and Data Engineering for Secure Societies IEEE Access Vol.7,No1.pg.18772 – 18782 IEEE 2019.

7. Jezia Zakraout1, Samir Elloumi,JihadMohamadAlja Am, Ssdok Ben Yahia,” Improving Arabic Text to Image Mapping Using a Robust Machine Learning Technique”, Advanced Software and Data Engineering for Secure Societies IEEE. Vol. 7,No.1,Pg.18772 – 18782,2019. DOI:https://doi.org/10.1109/ACCESS.2019.2896713

8. Vishal.V.Mahale,Agaskhind,,Amruta V. Pandit,Agaskhind,” Advanced Web Crawler For Deep Web Interface Using Binary Vector & Page Rank”,Published in 2ndInternational Conference on I-SMAC (IoT in Social Mobile Analytics and Cloud) Conference Proceedings 2018,Pg.500-503.DOI: https://doi.org/10.1109/I-SMAC.2018.8653765

9. ChengGuo , Ruhanzhuang, Chin-Chen Chang , and QiongqiongYuani,” Dynamic Multi-Keyword Ranked Search Based on Bloom Filter Over Encrypted Cloud Data”, Published in IEEE Access Vol.7.pg.18772-18782,IEEE 2019,

10. MomoKyozukaKeishi Tajima,” Ranking Methods for Query Relaxation in Book Search”, /WIC/ACM International Conference on Web Intelligence (WI) 2018 IEEE Conference Proceedings. DOI:https://doi.org/10.1109/WI.2018.00-51

11. Sanjumol A, Reena Mary George,” An efficient Nearest Neighbor search method For Spatial Keyword Query Processing”, Published in ICIRCS-Conference proceedings ofInternational Conference on Inventive Research in Computing Applications(ICIRCA),IEEE 2018.

12. Z. Xia, X. Wang, X. Sun, and Q. Wang, “A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data,” IEEE Trans.ParallelDistrib. Syst., Vol. 27, No. 2, pp. 340 _352, Jan. 2016 DOI: https://doi.org/10.1109/TPDS.2015.2401003

13. Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, “Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing," IEICE Trans. Commun., vol. E98-B, no. 1,pp. 190_200, 2015.DOI:http://dx.doi.org/10.1587/transcom.E98.B.190
14. Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving effective cloud search services: Multi-keyword ranked search over encrypted cloud data supporting synonym query," IEEE Trans. Consum. Electron., vol. 60, no. 1, pp. 164-172, Feb. 2014.