International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-02, May 2018

A Secure Fuzzy Keyword Search in a Multi Owner Paradigm at Cloud Environment

R.S Padmaja, P.G Scholar, JNTUACE Anantapur, India, padmajaramadas27@gmail.com

Dr. S.Vasundra. Professor of CSE, JNTUACE Anantapur, India, vasundras.cse@jntua.ac.in

Abstract - **With the hasty exploitation of cloud computing the capacity to store the information by the customers in public cloud increases. For the security of records, customers should encipher the statistics before storing into the cloud which makes the data usage that includes statistics retrieval a critical assignment. Therefore, it is perfect to enable the quest service over enciphered cloud information for helping powerful and effective statistics revival over a big number of consumers of information and records from the cloud. Existing procedures on an enciphered cloud information search paid attention on a single keyword search and become ineffective while a massive amount of files are there, and therefore have not much support for the decisive multi keyword search. This paper advocates a weight search technique that supports the effective multi keyword ranked seek in the cloud computing system. Specifically, it first suggest a primary scheme, the use of polynomial characteristics to cover the enciphered keyword and search styles for decisive multi keyword ranked search. To enhance the performance of keyword search, this paper proposes the fuzzy keyword searching mechanism.**

Keywords — ***Cloud Computing, Fuzzy Keyword Search, Indexing, Multi keyword search, Ranked keywords, LSH function.***

INTRODUCTION

Cloud computing is a high fantasy perceiving of computing as an application, in which cloud clients can store their information in the cloud to relish the required applications from a communal puddle of resources. Its awesome flexibility and financial reserves inspiring people and corporations to outsource their complicated data management systems to the cloud. To safeguard statistics privacy and to dispute against unwanted access in the cloud, sensitive statistics, e.g., emails, individual fitness statistics, tax files, pictures, financial business, and so on., the information should be encrypted via information proprietors before storing data into the public cloud; this, however, antiquate the conventional data usage service is stationed totally on plaintext keyword seek. The minor solution of retrieving all the information and deciphering originally is simply unfeasible because of the massive measure of bandwidth value in cloud scale structures. Furthermore, other than casting off the nearby garage control, storing records into the cloud servers has no cause except they may be without difficulty searched and utilized. Thus, scrutinizing privacy securing and powerful seek carrier over encrypted cloud information has supreme importance. Consider the massive number of calls for information users and the large amount of outsourced information records within the cloud, this hassle is predominantly difficult as it's far extraordinarily troublesome to satisfy additionally the necessities of performance, machine usability and scalability. To acquire the efficient facts retrieval demand, the enormous quantities of records demand the cloud server to carry out end result significance ranking. Such ranked seek system permits statistics customers to discover the maximum relevant facts fastness, in place of the concern of categorizing through each match if the content material series. Ranked[9] seek also can gracefully eliminate useless community visitors via sending back the most relevant facts. For privacy security, such rating operation, however, need not to reveal any keyword related statistics. To improve the hunt result precision in addition to improve the consumer seeking experience, it's also important for such rating system to assist a couple of keywords seek, as a single keyword search regularly yields some distance to coarse outcomes. As it is not an unusual exercise indicated via today's internet engines like Google (e.g., Google seek), records users can also generally tend to offer a hard and fast of keywords in place of simplest one because, the indicator in their search hobby try to get back the maximum related information. And every keyword in the seek request is capable of assist sender down the search result in addition. However, how to practice it in the enciphered cloud data seek devices stays a totally difficult undertaking due to inherent protection and privacy boundaries, inclusive of various strict necessities like the records privacy, the keyword privacy, and plenty of others.

DOI : 10.18231/2454-9150.2018.0264

RELATED WORK

Cloud computing economically permits the paradigm of information service outsourcing. However, to safeguard knowledge privacy, sensitive cloud knowledge ought to be encrypted before outsourced to the industrial public cloud, that makes effective knowledge utilization service a really difficult task. Though ancient searchable coding techniques permit users to firmly search over encrypted knowledge through keywords, they support solely Boolean search[1] and aren't nevertheless sufficient to satisfy the effective knowledge utilization want that's inherently demanded by sizable amount of users and big amount of knowledge files in cloud. C. Wang [5] proposed a tendency to outline and solve the matter of secure hierarchic keyword search over encrypted cloud knowledge. hierarchic search greatly enhances system usability by sanctioning search result connexion ranking rather than causing undifferentiated results, and additional ensures the file retrieval accuracy. Specifically, we have a tendency to explore the applied math live approach, i.e., connexion score, from data retrieval to make a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly defend those sensitive score data. The ensuing style is in a position to facilitate economical server-side ranking while not losing keyword privacy. Thorough analysis shows that our planned resolution enjoys "as-strong-as-possible" security guarantee compared to previous searchable coding schemes, whereas properly realizing the goal of hierarchic keyword search. intensive experimental results demonstrate the potency of the planned resolution.

Z. Xu, W. Kang, R. Li, K. Yow, and C. Xu [2],[10] aim to offer a feasible answer for the problems of multi keyword ranked query over enciphered statistics in the cloud location. They first outlined the trouble, examined the existing answers and designed a unique set of rules called Multi Keyword ranked Query on Enciphered data (MKQE) to deal with the obstacles. MKQE adopted a partitioned matrices approach. When the amount of enciphered fact's increases and larger keywords want to be brought, the seeking framework can be evidently elevated with the minimum overhead. They additionally layout a novel trapdoor era set of rules that resolve the inoperative trouble within the lower back result set without dropping the safety of facts and security property. Moreover, the weights of the key phrases are taken into deliberation within the ranking set of rules while giving the result. The experiments verify that their method can attain higher overall concert with a great protection stage.

This paper formalize and resolve the difficulty of assisting effective privacy maintaining fuzzy search to attain powerful usage of stored enciphered statistics in cloud via J. Li et al. They layout an sophisticated technique (i.e., wildcard based method) to construct the fuzzy keyword units by applying a substantial commentary on the similarity measure of Edit distance; depending on the formulated fuzzy keyword sets, they, also suggest a decisive fuzzy keyword seek scheme. Through thorough safety evaluation, this displays that their anticipated solution is relaxed and privacy preserving even as successfully knowing the intention of fuzzy keyword seek.

M. Chuah and W. Hu [4][3] have proposed technique that let the users to perform fuzzy multi keyword searches on enciphered facts. Their approach allows clean inclusion of newly available statistics objects and need not to rebuild the entire index tree while new statistics is available. They additionally use co-occurrence chances to decide extra beneficial multi key phrases that may be related to the published enciphered information objects. This function gives quicker response instances for fetching reports with multi keyword queries. Their system also allows fuzzy multi keyword search. The interpretations of the proposed work results in higher creation time and lower garage value with the increase in the size of facts documents. In addition, the quest time using their technique is affordable and tends to be improved for multi keyword seeks in which again listing for individual keyword is large.

D. Boneh[6] proposed Identity Based Encryption (IBE) Scheme and has its own merits and demerits on quicker responses.

J. Hur[8] proposed attribute-based data sharing scheme in which keyword retrieval efficiency is less when compared to fuzzy keyword search.

Ranked keyword search on remotely stored records is executed through storing documents in the cloud and retrieve the documents by means of searching through the keywords. Recovered documents are tested in ranked order that is accomplished by the use of rating algorithm within the index page. Security for statistics saved in cloud is achieved through storing encrypted documents and privacy of data is maintained with the aid of presenting one-of-a-kind trapdoors to diverse customers. Ranked evaluation is accomplished by the way of score dynamics i.e. taking the consumer picks into attention and giving highest rank to person chosen file so that the person can get more results.

FRAMEWORK

A. *Proposed system overview*

Figure1 represents a model of multi owner paradigm at cloud environment for secure keyword search on enciphered data. It proposes an effective user verification protocol that hampers attackers from eavesdropping mystery keys and from the illicit users performing searches, however additionally permits data user authentication and revocation. This paper assemble a unique search protocol which is secure that

DOI : 10.18231/2454-9150.2018.0264

allows the cloud server to carry out ranked keyword search without understanding the tangible information of each key phrases and trapdoors and additionally let the data owners to encipher the keywords with their own keys and permits the clients.

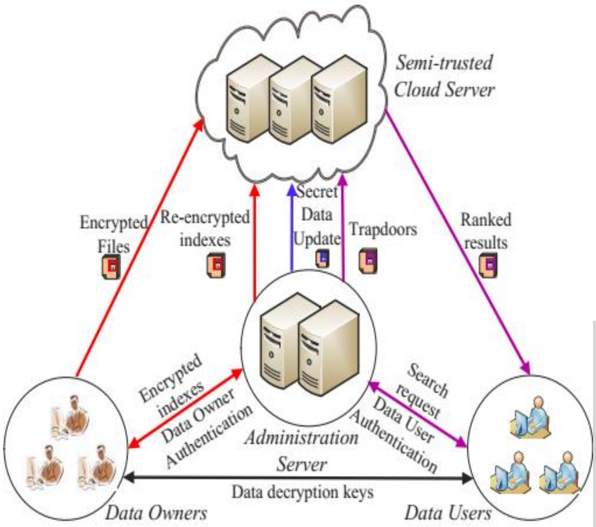

Figure: 1 Overview of the proposed Framework

The multi owner & multi user cloud computing version contains 4 modules as shown in Fig. 1; they are Admin Server, Cloud Server, Data Owner and Data User. Data owners consist of set of files. For decisive seek operations on those files which are in the encrypted form, information owners first constructs the index for the keyword set W derived from file, and then they send the index (I) to the Management server. Data proprietors encipher their documents and treasures the corresponding enciphered documents C to the cloud server. The received index is enciphered by the admin for the legal information proprietors and sends that re-enciphered index to the cloud server. Whenever the records consumer needs to search the keywords over the enciphered files that are stored in the cloud server, he first calculates the consequent trapdoors and sends to admin. Once the statistics person is verified by the admin, the management server in addition re-enciphers the trapdoors and publishes them to the cloud server. On deceiving the trapdoor T, the cloud server seeks the enciphered index I of every data proprietor and gives the related set of enciphered files. To enhance the document retrieval accuracy and store conversation cloud server return the pinnacle k applicable documents to the user. After retrieving the top k enciphered documents from cloud, the user will decipher those documents.

Fuzzy Searching Scheme

However, we need to improve the multi keyword search results over encrypted data by implementing new searching mechanism named as fuzzy searching scheme for multi owner paradigm. In this, whenever the user enters a keyword, it first converts into n-gram set and his similarity of keyword is calculated using Euclidean distance. It later uses Locality Sensitive Hashing (LSH) functions to produce the index and query. Because of LSH function, even when the keywords are wrongly spelled, it still could be hashed into the equivalent bits in the query vectors with high possibility and finds the matches. Finally, this method uses the inner product of the index vector and the query vector as the relevance score between queries and records. This proposed fuzzy search can solve the troubles of multi keyword search with high effectiveness and accurateness.

Suppose we have to store keyword: "**fuzzykey**"

- **N-grams:** fuz uzz zzy zyk yke key
- **Encrypted n-grams**: thr yu7 tf5 7yt lk8 eer
- Storing **thr** into **index_0** table
- Storing **yu7** into **index_1** table
- Storing **tf5** into **index_2** table
- Storing **7yt** into **index_3** table and so on….

Rather than storing all n-grams into single table we are storing the first n-gram in first table, second n-gram in second table and so on.

This will reduce the number of comparisons required to match the n-grams and hence faster the search results.

Consider the Scenario1, Admin want to upload the file on cloud and the keywords associated with the file, suppose admin uploaded a file abc.txt, the file will be encrypted and stored in the file system. Now the keywords that are associated with the files generate the n-grams and those n-grams will be encrypted using a secret key and the encrypted n-grams are stored on to the database. So we have now uploaded the file and the keywords which are both in encrypted form.

Now consider the 2nd scenario where users want to search for a file and the user entered the keyword language in a search box but misspelled the keyword and started searching.

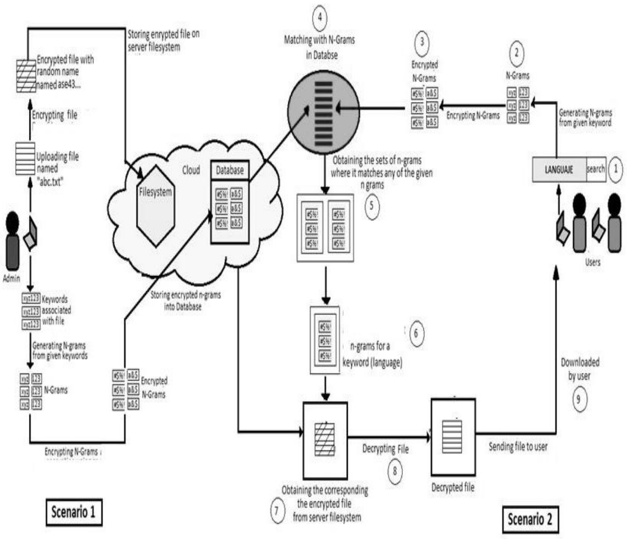
Storing enrypted file on server filesystem
Encrypted file with random name named ase43...
Encrypting file
Uploading file named "abc.txt"
Admin
Keywords associated with file
Generating N-grams from given keywords
N-Grams
Encrypted N-Grams
Encrypting N-Grams
Scenario 1
Cloud
Filesystem
Database
Storing encrypted n-grams into Database
4
Matching with N-Grams in Databse
3
Encrypted N-Grams
2
N-Grams
Encrypting N-Grams
Generating N-grams from given keyword
LANGUAJE
search
1
Users
Obtaining the sets of n-grams where it matches any of the given n grams
5
n-grams for a keyword (language)
6
Downloaded by user
9
Decrypting File
8
Decrypted file
Sending file to user
Obtaining the corresponding the encrypted file from server filesystem
7
Scenario 2

Figure 2: Working of the project

DOI : 10.18231/2454-9150.2018.0264

N-grams will be generated to that keyword also and encrypt those n-grams with the same secret key and obtains the set of n-grams that matches any of the encrypted n-grams of scenario1. According to the keyword in database the corresponding encrypted file is fetched form the cloud server. Then the user will decrypt the file.

B. Implementation Modules

1. Data Owner

Data proprietor scrambles the data for securing the facts in cloud using Commutative RSA before moving into the cloud. They likewise signify the doorway rights for the client who desires to get those archives. The entrance proper is a 2-state variable: consent conceded or authorization denied. Information owner makes a document tree in view of B tree and encodes the tree utilizing CRS.

2. Cloud server

Cloud server shops the scrambled data statistics and encoded document tree. It acknowledges the scrambled keywords (trapdoor) and offers back the coordinating statistics record in view of their pertinence. Information consumer can search for encoded records documents in cloud with scrambled catchphrases (trapdoor) and the cloud server will return the relevant results for the search request.

Experimental Analysis

In this experiment, we run the admin server and the data owner. The data user must register into the application. After registration of the data owner, he must authenticate to the admin server. After he upload the file into the cloud and that uploaded file will be stored in the server.

When the data user login into the system he also must authenticate to the cloud or admin server and he can search the keywords.

When compared to other techniques, fuzzy keyword search has its own advantages. If user searches normal or correct keywords on the server, then the server will display the results. Even though he will enter wrong keyword the fuzzy search will give the similar results to the data user and then he can download the file from the server. It greatly increases the system usability by returning the matching files and is especially useful in situations where misspelling is an issue. Along with the privacy preserving, fuzzy adds much benefits for data retrieval techniques.

In Figure3 it is observed that the performance is increasing with the use of Fuzzy based keyword search when compared to the exact keyword search.

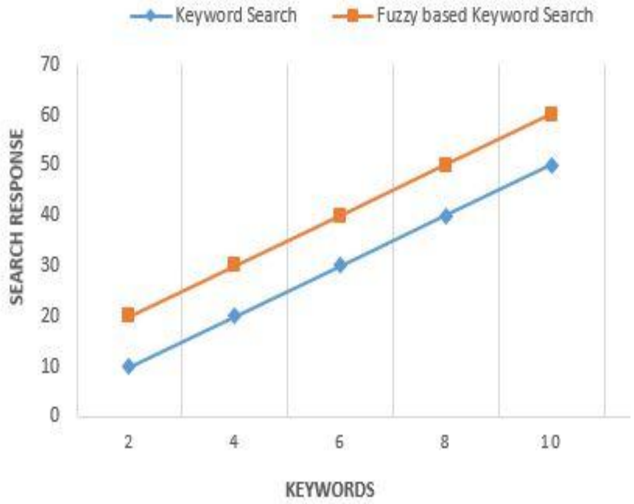

Figure 3: Performance graph

Conclusion

This paper implemented the fuzzy searching scheme to avoid the drawbacks of the multi-keyword search scheme. By implementing fuzzy keyword search, data users can get the efficient and exact results for the searched keywords. And it is efficient than exact keyword search for research and investigation. The design enables authenticated data users to attain secure, convenient, and efficient searches over a couple of statistics owners' records. To efficaciously authenticate information users and locate attackers who steal the name of the game key and carry out unlawful seeks, a singular dynamic mystery key era protocol and a new records user authentication protocol is suggested. To allow the cloud server to carry out cozy seek among a couple of owners' records encrypted with one of a kind secret keys, this paper systematically assemble a unique relaxed search protocol. For ranking query items and safe the security of importance appraisals among keywords and records, this paper embrace a solitary Additive Order and Privacy Preserving Function.

REFERENCES

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy preserving multi-keyword ranked search over encrypted cloud data,” Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 1, pp. 222–233, 2014.

[2] Z. Xu, W. Kang, R. Li, K. Yow, and C. Xu, “Efficient multikeyword ranked query on encrypted data in the cloud,” in Proc. IEEE Parallel and Distributed Systems (ICPADS’12), Singapore, Dec. 2012, pp. 244–251.

[3] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in Proc. IEEE INFOCOM’10, San Diego, CA, Mar. 2010, pp. 1–5.

[4] M. Chuah and W. Hu, “Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data,” in Proc. IEEE 31th International

DOI : 10.18231/2454-9150.2018.0264

Conference on Distributed Computing Systems (ICDCS'11), Minneapolis, MN, Jun. 2011, pp. 383–392.

[5] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 23, no. 8, pp. 1467–1479, 2012.

[6] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," SIAM Journal on Computing, vol. 32, no. 3, pp. 586–615, 2003.

[7] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. IEEE ASIACCS'13, Hangzhou, China, May 2013, pp. 71–81.

[8] J. Hur, "Improving security and efficiency in attribute-based data sharing," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, vol. 25, no. 10, pp. 2271–2282, 2013.

[9] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proc. ACM SIGMOD'04, Paris, France, Jun. 2004, pp. 563–574.

[10] T. Jung, X. Y. Li, Z. Wan, and M. Wan, "Privacy preserving cloud data access with multi-authorities," in Proc. IEEE INFOCOM'13, Turin, Italy, Apr. 2013, pp. 2625–263.

ACKNOWLEDGEMENTS

I thank Prof. S.Vasundra, Dr. Venkatesh for the guidance given for processing this work.

AUTHORS BIOGRAPHY

R.S.Padmaja received her B.Tech degree in Computer Science and Engineering from Padmavathi Mahilakalasala, Tirupathi, in 2014. Currently, she is pursuing her M.Tech in Computer Science and Engineering from JNTUA College of Engineering, Ananthapuramu, Andhra Pradesh, India. Her areas of interests include Cloud Computing and Network Security.

Dr. S.Vasundra is working as Professor & Head of the Department in Computer Science and Engineering, JNTUA College of Engineering Anantapur, Ananthapuramu, Andhra Pradesh, India. She received her Ph.D. from JNTUA University in the year 2011. Her areas of interests include Mobile Ad hoc Networks, Computer Networks, Data Mining, Cloud Computing and Data Science. She is a professional Body Member of ISTE, IE, IEEE and CSI.

DOI : 10.18231/2454-9150.2018.0264