IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Multi Keyword Search Within An Encrypted Text Using Tf-Idf Based Trapdoor Function

L. Soumya[1*], B.S. Vamsi Krishna [2]

[1]Dept. of CSE, MVGR College of engineering, Vizianagaram , Andhrapradhesh, India
[2]Dept. of CSE, MVGR College of engineering, Vizianagaram , Andhrapradhesh, India

[*]*Corresponding Author: sowmya.lavudi9@gmailcom, Tel.: +91-94409-14598*

Available online at: www.ijcseonline.org

Received: 20/Mar/2018, Revised: 28/Mar/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

***Abstract*—** Many people can save sensitive data on remote servers, provide data access from the admin to data users. As the stored information can hold important information, before uploading the data to the cloud, the information must be encrypted. If any cloud user wants to retrieve any file then they no need to check every file in the cloud. Data user can utilize keyword-based document retrieval. This paper suggests a technique to retrieve the encrypted information from the data store through multiple keywords. This technique instantaneously maintains active update operations like deleting as well as inserting records. Particularly, TFIDF is preferable for building index as well as generating query. Here we establish a tree-based file index to offer multi keyword ranked search. For encrypting and decrypting the files, AES algorithm is used and simultaneously offer the exact relevance score frequency among encrypted indexed files. In this technique, decryption can be performed before downloading the file from the database. General tests are directed to show the productivity of the proposed scheme.

***Keywords*—**Multi-Keyword Search, Security, TF-IDF, Indexing, Search Query.

I. INTRODUCTION

The cloud computing gives calculating usefulness and hires out the computing as well as storing capabilities to users in such a structure, user can distantly store information on the data store. Cloud computing defines a various types of computing theories that contain a huge number of computers associated through a communication network. Broader idea of converged substructure as well as shared facilities layered the establishment of cloud computing. The word cloud remains not simply the alternative word for Internet. Cloud is the place we go to utilize innovation when we require it until we need it and no need to install cloud on our system. Additionally we no need to pay for cloud when we are not utilizing it. Cloud can be both software and infrastructure. Security is one of the most often-cited objections to cloud computing. By this attractive option, each people as well as organizations triggered to save their information to the cloud. Data service providers can provide the security for the data owner's information. To secure the information, data must be encrypted before saving into the cloud. But, it costs high in terms of data usage. The previous procedures on key phrase based data extraction; those are basically performed on unencrypted data and cannot be openly used on the encrypted information. Decryption will performed on after downloading the information is totally impractical. So as to address the above issue, we have planned some broadly useful solutions by AES technique. More viable solutions, like searchable encryption (SE) schemes have particularly granted efficiency, functionality and security. Searchable encryption techniques enable users for saving the encrypted information into the cloud by performing keyword search for text file which is encrypted. Various searching schemes are proposed to achieve different search results, such as single keyword search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. In these techniques, multi-keyword ranked search is much better to obtain effective results. Few dynamic techniques were performed recently for supporting inserting/deleting actions on collection of different documents. This project initiates a secured multi keyword ranked search plot over the cipher text data, thereby helping multi-key phrase ranked search and dynamic operation on the collection of text documents and decrypt the files before downloading the file. Particularly, TF × IDF model (Relevance frequency model) is utilized for building an index and generating query for multi-keyword ranked search.

This paper is organised in the following manner. Section I starts with the importance of keyword searching techniques for retrieving documents, Section II contains the related work on various keyword searching techniques. Proceeding ahead, Section III discussed the problem statements and existing approaches. Section IV provides details to the baseline approaches and continues developing methodology of revised versions of multi-keyword searching technique.

Section V describes the Experiments performed and discusses results achieved by the proposed model. Finally, Section VI concludes research work with future directions.

II. RELATED WORK

Searchable Encryption design can be used for storing the encrypted information into data store and perform keyword searching over cipher text information by a client. Storing and securing the data can be proposed by authors in [1],[2],[3],[4]. In the view of various cryptographic methods, searchable encryption schemes could be developed using public key cryptography [5],[6] or symmetric key based cryptography [7], [8], [9], [10].

Wang *et al.* in [9] [8] proposed the importance of ranked keyword searching over encrypted information. The work initiated 2 methods for single keyword search over encrypted text data. Their method was an extends [10] and they add the security to it. The client produces a deterministic trapdoor which means the same trapdoor is created at the time of searching a specific keyword. The advanced technique offers secure dynamic inverted index and saves the RF encrypted. But, in [11] it was proposed a effective differential attack on the aforementioned scheme.

Kamara *et al.* in [12] explained a dynamic searchable symmetric encryption scheme extends [10]. The method encourages inserting, deleting or updating a file. Despite the fact that it could be named as a development in the field of SE, but it depends on a deterministic trapdoor which means same trapdoor is produced a similar term without fail.

Wang *et al.* in [13] introduced a range search scheme on encrypted text data.This design is in fact momentous as it is not reliant upon a specific geometric shape and backings Axis-parallel Rectangles, Circles, Non-axis parallel Rectangles and triangles. Be that as it may, their plan doesn't give ranking causing in further network traffic.

Tang in [14] initiated a symmetric searchable multiparty encryption plot (MPSE) which is an augmentation of [15], Presents a Follow calculation that dispenses a token to admin for disseminating among the client of index table. This token approves the user to execute the search on index table. The plan encourages the dynamic clients yet doesn't permit dynamic databases. Meanwhile, trapdoor is deterministic, therefore, the plan is vulnerable for deterministic assaults. Their plan utilizes forward index that is an index for each document due to which the ranking is impossible.

III. PROBLEM STATEMENT AND EXISTING APPROACHES

DATA RETRIEVAL TECHNIQUE:

Data Retrieval involves retrieving the required information from database. Documents and queries are the data to be retrieved. Based on the query the user can easily retrieve the data as per their concern. Further the Database system, software can browse the required file in the database. The extracted data may be downloaded by using various searchable techniques.

SEARCHABLE ENCRYPTION:

Searchable encryption techniques are very important to make the cloud server efficient for the users the encryption. But the old techniques are not provide efficient results to make the system actual efficient for practical execution. So to make the cloud server effective for practical execution we have proposed another searchable technique for efficient information retrieval alongside keeping up the privacy of groups or an individual user. In this technique there are two kinds of entities one is data owner and another is data clients. The data owners have the right to upload documents, and all the access rights related to the documents. Data user will search the files through the database by using queries and can download files. But the main problem in all this is the security which is primarily maintained by authenticating the data owner as well as data clients to maintain validity of the data.

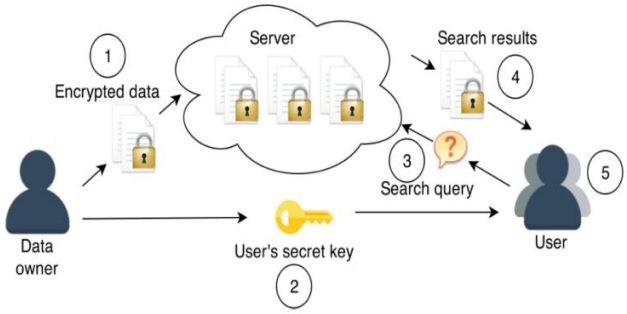

Figure. 1. Framework for representing the search over encrypted data.

RANKED KEYWORD SEARCHING:

Ranked search technique allows retrieval of the matching files in a ranked order based on certain relevance model (e.g., keyword relevance frequency), thus stepping forward for practical execution of privacy information in the Cloud Computing. This technique proposes the ranked searchable symmetric encryption scheme to accomplish our design goal on both usability and security of the system.

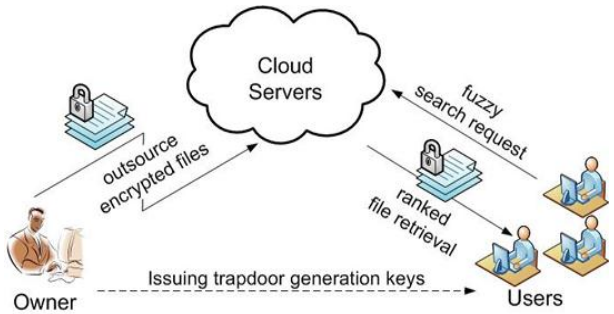

Figure. 2. Framework for representing the Ranked Keyword search over encrypted data.

MULTI-KEYWORD RANKED SEARCHING:

Multi-keyword ranked search scheme perform more effectively than other search techniques. In this technique data user retrieve the data files by placing different multiple keywords in the search query. Newly, few dynamic plans were implemented for uploading as well as deleting activities over documents gathered. These effectively perform as it is profoundly conceivable such that data owners have to refresh their data over cloud server.

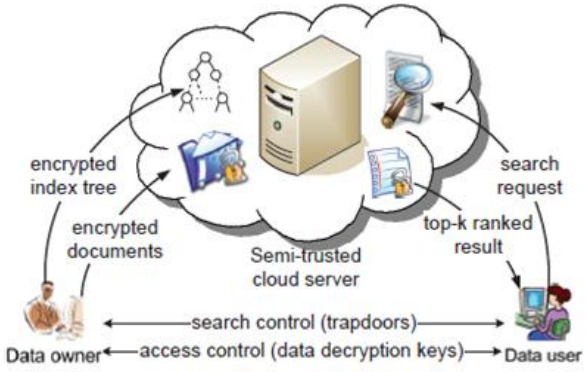

Figure. 3. Framework for representing the Multi-Keyword Ranked search over encrypted data.

EXISTING SYSTEM

The existing methods of keyword based search retrieves only the plain text instead of retrieving the encrypted text. Also these include the procedure of decrypting the file only once it is downloaded. These existing multi keyword search methods retrieve data files depending on the keywords, that won't give worthy outcome ranking functionality.

CONTRIBUTION:

1. We prefer multi keyword ranked search scheme to search a particular file in the search index by using multiple keywords.
2. We inspect further enhancements for the multi ranked search method to retain additional search results and dynamic data process.
3. We identify the issue in multi keyword ranked searching over encrypted text and prefer some privacy methods for a secure data operation system.
4. Thorough examination on security and Efficiency affirmation of the proposed plans is determined, and testing on real-world data set additionally demonstrates the proposed methods unquestionably acquire low overhead on calculation and communication. In this paper we propose a strategy where it includes saving decrypted document specifically into the client's system.

PROPOSED SYSTEM:

"Multi Keyword Search With in an Encrypted Text using TF-IDF based Trapdoor Function" In this paper we build up a special tree-based index structure along with greedy Depth-first Search algorithm to gives a proficient multi-key phrase ranked search. The developed plan can accomplish sub-linear search time and manage the deleting/inserting of documents effortlessly.

- Plenty of works proposed different threat models to achieve various search functionalities.
- Encryption and keywords generation are performed by TF_IDF model.
- Users easily searching the required file by placing multiple keywords in the query.
- Proposes the process of automatically decrypting the file before downloading the file itself.

IV. METHODOLOGY

Data owner used to store their data in public servers because of large size of data, Large size of data can cause the huge cost. But privacy is the key issue for storing of data then plain text data is in text format which can be easily edited or hacked so most of the data is stored in covered format so retrieving the data which is in secured format is main challenge and need to get the relevant data meanwhile the large number of documents are stored in cloud it is necessary to search the data by using the keyword technique so multi keyword technique is one of the most important factor and challenging and also need to retrieve the relevant document with the relevance keyword.

SYSTEM ARCHITECUTRE:

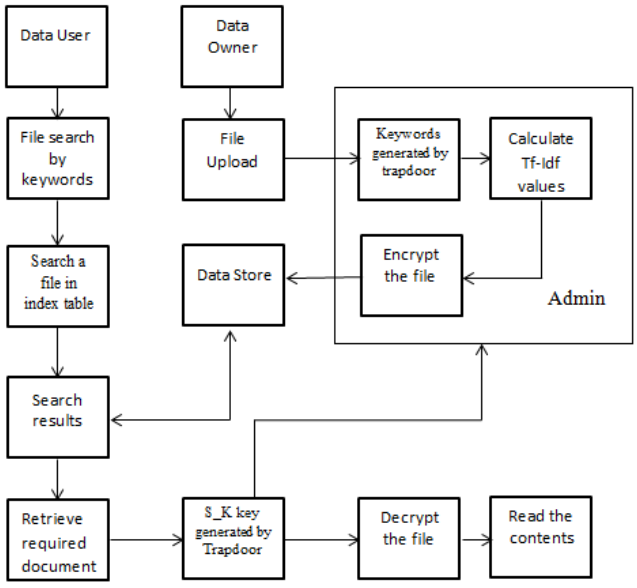

Figure. 4. Architecture of the Retrieving encrypted cloud data.

ADMIN: Admin have the rights to manage both data owner and data user. Admin is a main authority of this application; he maintains all user's data means data users and data owners information. He creates the data owner profiles and he accept or reject the user registration details.

DATA OWNER/USER: That can be an organization or an individual originally possessing Sensitive data to be store in the cloud. In proposed system data owner is industry owner. Here the user may Data Owner and Data User, User after his authenticating he can search cloud the based search, and the keywords Thas input, this generates a corresponding trapdoor. Data owner builds a searchable index I which is encrypted by the symmetric key SK and then outsourced to the cloud server. After the construction of index table, the collection of documents can be encrypted and outsourced independently. The keywords placed suggested by Trapdoor.

MULTI-KEYWORD SEARCH : Data outsourcing is the major threat in the distributed computing where the information is stored in the secured format of encryption and after that recovering the information from the cloud storage with the relevant data is a main task since it requires the accurate information recovery system so we propose another approach of closest neighbourhood query retrieval process with the keyword searching method on secured cloud information so we utilize the process of the key administration where the client and owner have the full trust and the information is more secured transposed with the authorization of security with the key to encrypt and decrypt the information so we execute a scheme for effective information retrieval using AES algorithm.

Trapdoor: Trapdoor mainly designed for generating keywords and symmetric key. When data owner upload a file data store then multiple keywords and symmetric key can be generated by trapdoor and when user send the request to find a file by using multiple keywords then the symmetric key can be forwarded by the trapdoor. Admin need not bother about the data in the server and searching requirements for both Owner and User.

Relevance Frequency: Relevance Frequency used for identifying frequently occurring words in files to generate the keywords. This function assigns a relevance score for every matching result to a specified search query.

TF×IDF model can be used for index construction and query generation to provide multi-keyword ranked search.

TF(w) = (No. of times words w appears in a record) / (Total no. of words in the record).

IDF(w) = log_e(Total no. of records / No.of files with word w in it).

Relevance Frequency = TF*IDF

ENCRYPTION AND DECRYPTION ALGORITHM:

For greater flexibility, data owners outsource their data from local systems to remote servers. Security issues are to be considered if sensitive data is outsourced into the open source. For securing data, they are encrypted before uploading. Search on plain text is an easy task. But it is complicated in cipher text files. Several searchable symmetric encryption methods are there for cipher text search. Be that as it may, they do not think about the relevance of record and allow only single keyword in a search query. For a large information management environment, various rank based searches are important. Existing plans that supports ranked search have low efficiency and security. So to have security in data recovery, an encryption method called AES Encryption is being utilized.

AES Algorithm can be utilized for both encryption and decryption process.

Greedy Depth First Search:

Greedy Depth First Search can be used for searching the relevant document from the huge number of document collection in data store. This algorithm can search the file based on shortest path first concept.

MODULE DESCRIPTION:

Multi keyword search within an encrypted text will be performed by these following steps. Those are

Module 1: Initially Data owner and data user must be authenticated by Admin. Admin have the rights to accept or decline the authentication of data owner and data user.

Module 2: When data owner wants to save his/her data into data store, they must be request to trapdoor for generating the keywords for the particular record. If data owner upload any document to the store then they send requests to trapdoor.

Module 3: When admin receives a request from data owner to the trapdoor, admin can execute that request then calculate Tf/Idf values and perform encryption analysis on that particular record. Then generates symmetric_key and frequently repeated keywords.

Module 4: If any authenticated user can request a particular file by placing multiple keywords into the search query then the application displays the relevant documents in a ranked order. Users select the required document and send the key-request to trapdoor. Trapdoor can check the key-request and send to user in a secured manner.

User copies that s_k and downloads the file.

V. EXPERIMENTAL RESULTS

Security Analysis:

Index confidentiality is maintained in the proposed method for the admin it is not possible to gather the information about data vector and multi-keyword query

vector. Moreover TF-IDF values, patterns of query and search cannot be deducted by the admin. Relevant scoring of files based on multi-keyword query is performed in the cloud application but index confidentiality is maintained as index is symmetrically encrypted. The user query I is symmetrically encrypted. In this proposed scheme for same search query, the trapdoor generates different keywords. Maintains keyword privacy in traditional SSE plan it is possible for the trusted outsider to think about the querying keyword by investigating the access pattern, search pattern and term frequency. Also, they bolsters just single keyword query. Be that as it may, the proposed system supports multi-keyword query and hide the term distribution details. Consequently preserves keyword privacy.

Performance Analysis:

Index Building When a data owner uploads a file into server; trapdoor generates the keywords and encrypted using AES encryption method. The queried keywords are encrypted and plotted to the data vector. Length of that wordlist (secure index for each document) and the number of documents in the files determines the data vectors dimensionality. On the other hand dimensionality of the data vector directly determines cost for mapping. That is as number of documents increases the file size increases, as document size increases the length of secure index increases. Figure below shows time cost for building secure index for various documents with different number of keywords.

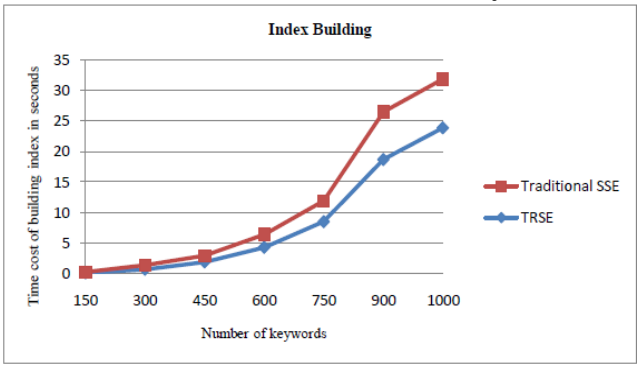

Figure 5 : Index Building

From the above graph it can be inferred that the proposed scheme requires less time than traditional SSE scheme to build a secure index.

Query Vector Generation The time to build a query vector depends on the length of secure index for every file. Thus the total of keywords in the user query has less influence on vector generation.

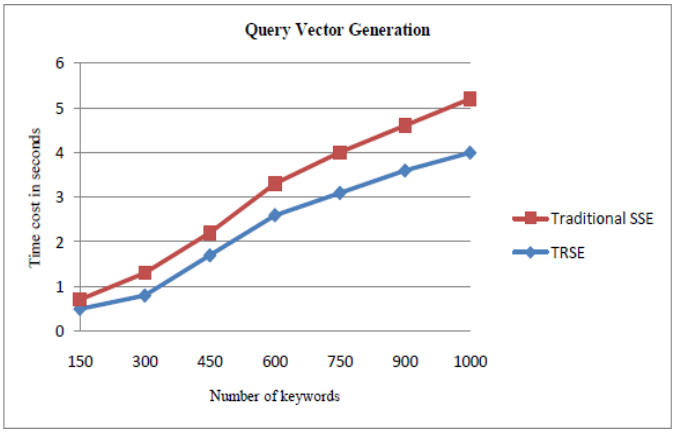

Figure 6: Query Vector Generation

VI. CONCLUSION AND FUTURE WORK

In this paper, a secure, efficient and dynamic search scheme was proposed for supporting accurate multi-keyword ranked search as well as dynamic deleting/inserting the documents. We create index using relevance frequency. AES algorithm is used for encrypting and decrypting the files. Experimental results determine the efficiency of our proposed system. This work can be extended towards developing content based searching over the encrypted images and video files. These solutions are under process in collaboration with varies encoding/decoding techniques.

REFERENCES

[1] K. Ren, C.Wang, Q.Wang *et al.*, "Security challenges for the public cloud," *IEEE Internet Computing*, vol. 16, no. 1, pp. 69–73, 2012.

[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Financial Cryptography and Data Security*. Springer, 2010, pp. 136–149.

[3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.

[4] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," *Journal of the ACM (JACM)*, vol. 43, no. 3, pp. 431–473, 1996.

[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Advances in Cryptology- Eurocrypt 2004*. Springer, 2004, pp. 506–522.

[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in *Advances in Cryptology-CRYPTO 2007*. Springer, 2007, pp. 50–67.

[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*. IEEE, 2000, pp. 44–55.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 253–262, 2010.

[9] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 8, pp. 1467 1479, 2012.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption," *Proc. 13th ACM Conf. Comput. Commun. Secur. - CCS '06*, p. 79, 2006.
[11] K. Li, W. Zhang, C. Yang, and N. Yu, "Security Analysis on One-to-Many Order Preserving Encryption-Based Cloud Data Search," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 9, pp. 1918–1926, 2015.
[12] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," *2012 ACM Conf. Comput. Commun. Secur.*,pp. 965–976, 2012.
[13] B. Wang, S. Member, M. Li, and H. Wang, "Geometric Range Search on Encrypted Spatial Data," *IEEE Trans. Inf. Forensics Secur.*, vol.11, no. 4, pp. 704–719, 2016.
[14] Q. Tang, "Nothing is for free: Security in searching shared and encrypted data," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 11, pp.1943–1952, 2014.
[15] R. Popa and N. Zeldovich, "Multi-key searchable encryption," pp. 1–18, 2013.

Authors Profile

Ms. Soumya is pursuing Master of Technology in MVGR College of Engineering, Vizianagaram, Adhra pradesh.

Mr. B.S.Vamsikrishna is a Senior Assistant Proffessor in MVGR College of Engineering, Vizianagaram, Andhra pradesh.

