International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 23 Issue 9 – DECEMBER 2016.

SECURE SEARCH SCHEMA OVER ENCRYPTED SECURED CLOUD STORAGE IN A HIERARCHICAL CLUSTERING COMPUTATION

B.KRANTHI KUMAR[#1] and VIKRUTHI RAMU[*2]

[#] *PG Scholar, Kakinada Institute Of Engineering & Technology Department of Computer Science, JNTUK,A.P, India*

[*] *Assistant Prof, Dept of CSE, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA*

***Abstract*— There is many secure objections in a multi-user strategy. First, all the users habitually keep the same secure key for trapdoor origination in a symmetric SE scheme. In this case, the revocation of the user is big demanding. If it is needed to revoke a user in this scheme, we demand to rebuild the index and disseminate the new secure keys to all the authorized users. Second, symmetric SE schemes frequently consider that all the data users are trustworthy with believable in nature. Which is not at all practical and a dishonest data user will make use advantage of it which also may lead to many security issues. In the proposed work, we improve the SE scheme to grasp these challenge problems. In our enhancement model, a hierarchical clustering method is designed to guide more search semantics with an additional feature of making the system to cope the demand for fast ciphertext search in large scale environments such as Big Data scenarios. The proposed hierarchical approach clusters the cloud documents with the minimum relevance threshold, and then divides the resulting clusters into sub-clusters until the necessity on the maximum size of cluster is reached. In the search module part of the proposed system, this advent model can influence a linear computational complexity against an exponential size increase of document collection. Experimental results prove the effectiveness of the proposed scheme.**

***Index Terms*— Multi-user model, Searchable encryption, Hierarchical index, linear model and relevance score.**

I. INTRODUCTION

Data processing as a service has become a vital role in the cloud storage systems. The user's upload their data to the cloud in the remote settings [1]. With the help of cloud server, the data can be processed; manipulated and outsourcing process can be done by the cloud user. The outsourced information may contain some sensitive data. These sensitive data should be protected from any threats or misuse of the data. While encoding the information, the data may loss, this might pose some challenges. It pertains to usability effect of outsourced data over encrypted data.

The better instance is the Google search where it uses Secure Socket Layer (SSL) to protect the data like credentials details etc from search and retrieval system. When the user enters the queries, the websites links to the concern web pages, and appears in the result. The owner develops several keywords depends on the outsourced data [2]. These keywords are further encrypted at the cloud server. The user explores the search over the data, the keywords are then matched with outsourced data and thus the query result is retrieved. To achieve the similar search efficiency and precision over encrypted data as that of plaintext keyword search, an extensive body of research has been developed in literature.

In this paper, we propose an enhanced search data over encrypted cloud storage data. By building a hierarchical cluster index, the files are arranged into clustered vectors. The threshold is set, in order to process the maximal size of the clusters. The objective of the study is to efficiently search and retrieve the files in lessened time. The rest of the paper is organized as follows: Section II describes the previous work done by other researchers. Section III describes the proposed work. Section IV describes the results and discussion of the proposed design. Atlast concluded in section V.

II. RELATED WORK

There has been lot of research study conducted in the multi-keyword search over encrypted data. A symmetric oriented encryption scheme is proposed by *Cash et al*, that achieves high efficiency with better security schemes. *Cao et al* proposed the multi-keyword ranking scheme with the use of k-nearest neighbors. *Naveed et al* studied about the dynamic searchable encryption that conceals the pattern of search user. *Sun et al* proposed a multi-watchword content that contain catchphrases to search the encrypted data over the cloud server. *Yu et al* proposed top k multi -keyword search that make use of homomorphic encryption to ensure the security.

Quin Liuy et al depicted that when the data is searched by the CSP, is facing some security issue. The solution provided is that only authorized user can access the data. The authors study targeted to maintain the preserve the privacy of the data. The basic technique used is the Public Key

Encryption system. The task of the service provider is to decode the ciphertext. The decryption data may reduce the cost complexity. Similarly, decryption process is also studied by the *Boneh et al*. *Ming li et al* proposed a private keyword search in order to protect the data while the searching operations done in multidimensional keywords. The study is further extended by the proposed APKS+ that maintain the privacy of the query.

In contrast, cong studied about the search operations in terms of cost analysis and network traffic. An indexing scheme is maintained to avoid redundancy of the data. Boneh et al. [7] proposed a PKC based search scheme when he inspired from the identity based encryption. This scheme is initially well suited for single query only. By using this scheme anyone with public can write to cloud but the user having a private key can only allow performing searching operations on cloud. By taking base of this technique, a number of methodologies have been implemented to filter the searching techniques. One of the great author Li [8] proposed another predictive encryption technique which is based on the hierarchical encryption. This technique build authorized keyword search technique over cloud. Like other techniques, this technique also gives search access to the authorized users and non-authorized users will not get access to search. In spite of effectiveness of these schemes it has biggest drawback that it is computationally expensive.

III. PROPOSED WORK

This section depicts the encrypted search schema over encrypted cloud storage in hierarchical clustering model. Let us assume a cloud storage that comprised of entities like data owner, data user, trusted third party and cloud server. In order to access the cloud service, the user should get enroll with the cloud server. The data owner contains a colossal of documents D to be forward to the cloud server in an encrypted form. Each document contains a unique ID before outsourcing to the cloud server. The index is maintained for the integrity of the data. The trusted third party will use index to find out accurate document. The proposed algorithm is presented in Fig. 3.1.

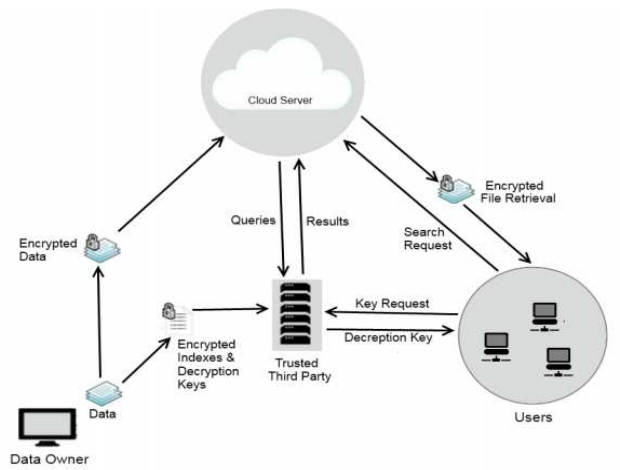

Fig.3.1 System Architecture

In order to enhance the efficiency of the search, a hierarchical computational scheme is developed. In this relevance score between query and the documents is estimated, to improve the searching schemes. By ignoring the irrelevant fields of the data, the search speed is enhanced. By the use of clustering methods, the documents are clustered. Here, the relevance score is the index used for evaluating the randomly arranged documents. The hierarchical method is used for obtaining better clustering results. And also, the cluster size determines the query efficiency and searching results. Thus, the cluster size creates greater impacts towards the accurate information systems. In this proposed work, we used hierarchical index to outsource the document D, to cloud server. The unique feature is that, every document is indexed in vector format. In the search phase, cloud server calculates the relevance score between the query and documents by computing the inner product of the query vector and document vectors and returns the target documents to user according to the top k relevance score. The proposed algorithm is executed in three algorithms, namely, Hierarchical index algorithm, Dynamic k-means and Quality Hierarchical clustering (QHC).

i) Hierarchical index algorithm:

- Input: Secret key {SK, k) and dataset D.
- Obtaining dictionary Dn from D.
- For each file in D
- Build a file vector fv_i.
- Invoke QHC to build the hierarchical index.
- Extend its dimension by analyzing its data cluster DC, Collection of cluster value in the dimensional vector.
- Then the index I, is encrypted using matrix multiplication and then outsourced to the cloud server.

ii) Dynamic K-means algorithm:

- Input: Set of k clusters with centers C.
- Defining the threshold TH.
- While k is static
- New cluster centers C' is created by k-means algorithm.
- For every new cluster center C'
- Find the minimum relevance score min(si)
- If min(si)< TH
- Insert a new cluster center k= k+1
- Operation executes until k is steady.

iii) Quality Hierarchical Clustering (QHC):

- Input: File and the set of Threshold TH.
- Cluster set is formed by dynamic k-means.
- While new cluster set is found, Ci.
- If the cluster size Ci > TH
- Fragment the clusters into sub-clusters Ci+1.
- Until all clusters meets the cluster's size prerequisites.

By employing our proposed scheme, we ensure the following design goals like correctness, freshness and completeness and privacy requirements like data, index and keyword.

IV. EXPERIMENTAL RESULTS

In this section, the performance analysis of the proposed scheme is described. Let us consider a dynamic data collection from the cloud server. Using naïve bayes concept, the documents are downloaded and indexed. Since these sorts

of data collection incurs higher cost and higher storage space. By our proposed scheme, the documents are indexed, encrypted and forwarded to the cloud server. The performance indices such as search precision and rank privacy are studied.

a) Search precision: It validates the satisfaction of the users. Depends upon the relevance score between documents and the query, the search precision is defined. It is given by the eqn. 1.

$$P_q = \frac{\sum_{i=1}^{k'} S(qw, d_i)}{\sum_{i=1}^{k} S(qw, d_i)}$$

Here, k_0 denotes the number of files retrieved by the evaluated method, k denotes the number of files retrieved by plain text search, qw represents query vector, d_i represents document vector, and S is a function to compute the relevance score between qw and d_i.

b) Rank Privacy: Rank privacy depicts the information leakage of the search results. It is given by the eqn. 2.

$$P_k = \sum_{i=1}^{k} \frac{P_i}{k}$$

Here, k denotes the number of top-k retrieved documents, c_{i0} is the ranking of document d_i in the retrieved top-k documents, c_i is the actual ranking of document di in the data set, and P_i is set to k if greater than k.

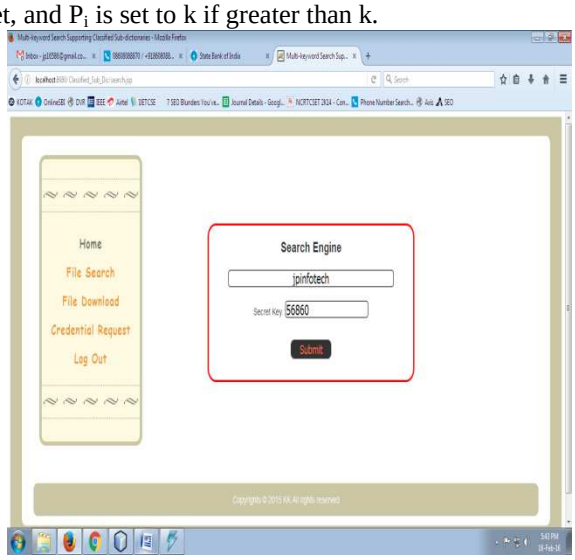

Fig.2. Entering the secret key.

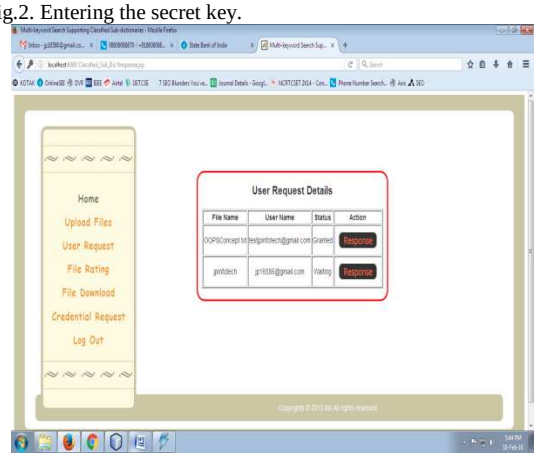

Fig.3. User's file requests

Fig.4 File is downloaded by user, once request is verified and approved.

V. CONCLUSION

In this paper, we studied about the encrypted data search in the scenario of cloud storage. By exploring the relationship between the files over the encrypted files and presents the design methods of semantic search. We introduced an enhanced hierarchical computation over the encrypted cloud storage. It is executed in three algorithms that data analysis over the encrypted cloud data. We have pin-pointed and diagnosed the main issues that are to be satisfied for secured data utilization are keyword privacy, Data privacy, Index privacy, Query Privacy, Fine-grained Search, Scalability, Efficiency, Result ranking, Index confidentiality, Query confidentiality, Query unlinkability, semantic security and Trapdoor unlinkability. Performance indices such as search precision and rank privacy were studied.

REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, ``An SMDP-based service model for interdomain resource allocation in mobile cloud net-works," IEEE Trans. Veh. Technol., vol. 61, no. 5, pp. 22222232, Jun. 2012.

[2] M. M. E. A. Mahmoud and X. Shen, ``A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 10, pp. 18051818, Oct. 2012.

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Y. Luo, ``Exploiting geo-distributed clouds for a e-health monitoring system with minimum service delay and privacy preservation," IEEE J. Biomed. Health Inform., vol. 18, no. 2, pp. 430439, Mar. 2014.

[4] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, ``A survey of mobile cloud computing: Architecture, applications, and approaches," Wireless Commun. Mobile Comput., vol. 13, no. 18, pp. 15871611, Dec. 2013.

[5] H. Li, Y. Dai, L. Tian, and H. Yang, ``Identity-based authentication for cloud computing," in Cloud Computing. Berlin, Germany: Springer-Verlag, 2009, pp. 157166.

[6] W. Sun, et al., ``Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2013, pp. 7182.

[7] B.Wang, S. Yu,W. Lou, andY. T. Hou, ``Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 21122120.

[8] E. Stefanov, C. Papamanthou, and E. Shi, ``Practical dynamic searchable encryption with small leakage," in Proc. NDSS, Feb. 2014.

[9] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, ``Secure dynamic search-able symmetric encryption with constant document update cost," in Proc. GLOBECOM, Anaheim, CA, USA, 2014.

[10] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Ro³u, and M. Steiner, ``Highly-scalable searchable symmetric encryption with support for Boolean queries," in Proc. CRYPTO, 2013, pp. 353373.

AUTHOR PROFILE

B KRANTHI KUMAR is a student of Kakinada Institute Of Engineering & Technology affiliated to JNTUK, Kakinada, pursuing M.Tech (Computer Science) His Area of interest includes Cloud Computing and its objectives in all current trends and techniques in Computer Science.

VIKRUTHI RAMU M.TECH is working as Assistant Professor, Department of Computer Science & Engineering, Dept of CSE, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA and Completed B.Tech in Gudlavalleru Engineering college, gudlavalleru M.Tech in SRKR, BHIMAVARAM. Exp:- 06.yrs Interested Area: Data Mining.