International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

A SECURE EFFICIENT AND LIGHTWEIGHT ENCRYPTED DATA MATCHING TECHNIQUE FOR MULTI KEYWORD SEARCH IN THE CLOUD

V.Muthu Lakshmi [1] , Dr.N.Sumathi[2]

[1]M.Phil Research Scholar, Department of Computer science,
Sri Ramakrishna College of Arts and Science [Formerly S.N.R SONS College],
Coimbatore-641006, TamilNadu, (India)
[2]Head &Professor,
Department of Information Technology,
Sri Ramakrishna College of Arts and Science [Formerly S.N.R SONS College],
Coimbatore-641006, Tamil Nadu, (India)

ABSTRACT

Searchable encryption is a crucial cryptographic technique that enables searching capabilities over the encrypted data utilization in the multi owner scenario on the cloud. Many existing schemes support single keyword search with simultaneous encouraging the dynamic update operations on the outsourced data incorporating privacy requirement. In spite of multiple opportunities, it faces many challenges such as wastage of network bandwidth and computing resources. In order solve those issues the present novel secure and efficient encrypted data matching technique for multi keyword search as it eliminate the data violations with less computation process. This process uses the inverted index table and iterative deepening depth first search algorithm for data retrieval over encrypted data. Furthermore process is secured by implementing the symmetric searchable encryption scheme. Meanwhile the vector space model with Term Frequency and Inverse Document Frequency is used to construct the inverted index table. The Search results also blinded using phantom terms which avoids the search results from designated user. Additionally, the User trust Model is integrated into the proposed technique using similarity measures in terms of data provenance.. It is to identify the user violation on the accessed data and to share the information with the neighbour or other data owner of the group or community about user violation in terms of data integrity.

Keywords: Searchable Encryption, Multi keyword Search, Encrypted data Search,inverted Index, User Trust

I.INTRODUCTION

Due to numerous advantages of the cloud computing, data outsourcing has been increased large extent by data owner which leverages the high security assumptions [1]. In order to handle security requirement, many solution has been implemented in terms of encryption in literature though there exist some limitations. Moreover data querying through search keyword has directed towards new forum. Searchable encryption [2] is probably the best solution that brings up a secure solution. Despite its advantages, there exist various limitations by applying the Searchable encryption to search keywords which demand of huge amount bandwidth and server requirement for processing the query [3]. In this paper, we propose novel secure efficient and lightweight Encrypted data matching technique for multikeyword search.

The Proposed Secure model utilizes the inverted index table by employing keyword balanced binary tree and it constructs a tree based index structure for keywords in the document uploaded. The vector space model with Tf-Idf is used to construct the inverted index table for documents stored in the cloud environment. The iterative deepening depth first search algorithm [4] is been applied for data retrieval over encrypted data by achieving linear search time. Due to multiple update at data owner end, it complicates the retrieval process so it has been resolved by implementing the dynamic symmetric searchable encryption scheme. The Search results also blinded using phantom terms. User Trust is also computed finally using similarity measures [5].

The remainder of the paper is organized as follows: Section 2 discusses the related works in Secure encrypted data search over single keyword and its impacts against the performing security under multi user environment, Section 3 briefly discusses the proposed technique in terms of dynamic Symmetric Searchable encryption and generation of inverted index and Section 4 presents the experimental results on a huge number of documents. Section 5 discusses conclusion and future work.

II.LITERTURE SURVEY

There exist many techniques to secure encrypted data search and implemented efficiently. Each of these techniques follows some sort of security principles, among few performs nearly equivalent to the proposed framework is described as follows

Jiadi Yu et al. [6] described cloud secure computing has emerging as a promising pattern for data outsourcing and high-quality data services. However, concerns of sensitive information on cloud potentially cause privacy problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud. In this literature, data privacy issues using SSE has been addressed. Privacy issue from the aspect of similarity relevance and scheme robustness has been formulated for further uses. That server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. To eliminate the leakage, a two-round searchable encryption (TRSE) scheme that supports top-k Multi keyword retrieval has been utilized. In TRSE, a vector space model and homomorphism encryption is employed. The vector space model helps to provide sufficient search accuracy, and the homomorphism encryption enables users to involve in the ranking while the majority of computing

work is done on the server side by operations only on cipher text. As a result, information leakage can be eliminated and data security is ensured. Thorough security and performance analysis show that the proposed scheme guarantees high security and practical efficiency.

Zhangjie Fu et al. [7] described the consumer-centric cloud computing paradigm has been analysed as it is emerged as the development of smart electronic devices combined with the emerging cloud computing technologies. A variety of cloud services are delivered to the consumers with the promise that an effective and efficient cloud search service is achieved. For consumers, they want to find the most relevant products or data, which is highly desirable in the "pay-as-you use" cloud computing paradigm. As sensitive data (such as photo albums, emails, personal health records, financial records, etc.) are encrypted before outsourcing to cloud, traditional keyword search techniques are useless. Meanwhile, existing search approaches over encrypted cloud data support only exact or fuzzy keyword search, but not semantics-based multi-keyword ranked search. Therefore, how to enable an effective searchable system with support of ranked search remains a very challenging problem. This literature proposes an effective approach to solve the problem of multi-keyword ranked search over encrypted cloud data supporting synonym queries. The main contribution of this literature is summarized in two aspects: multi-keyword ranked search to achieve more accurate search results and synonym-based search to support synonym queries. Extensive experiments on real-world dataset were performed to validate the approach, showing that the proposed solution is very effective and efficient for multi keyword ranked searching in a cloud environment.

Hongwei Li et al. [8] described the mobile cloud computing is analysed as it is the fundamental application to outsource the mobile data to external cloud servers for scalable data storage. The outsourced data, however, need to be encrypted due to the privacy and confidentiality concerns of their owner. This results in the distinguished difficulties on the accurate search over the encrypted mobile cloud data. To tackle this issue, in this literature, the searchable encryption for multi-keyword ranked search over the storage data has been developed. Specifically, by considering the large number of outsourced documents (data) in the cloud, we utilize the relevance score and k-nearest neighbour techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. Within this framework, an efficient index to further improve the search efficiency, and adopt the blind storage system to conceal access pattern of the search user has been leveraged. Security analysis demonstrates that our scheme can achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlink ability, and concealing access pattern of the search user. Finally, using extensive simulations, system can achieve much improved efficiency in terms of search functionality and search time compared with the existing proposals.

Zhangjie Fu et al. [9] described the cloud computing, searchable encryption scheme over outsourced data has considered as important aspect. However, most existing works on encrypted search over outsourced cloud data follow the model of “one size fits all” and ignore personalized search intention. Moreover, most of them support only exact keyword search, which greatly affects data usability and user experience. So how to design a searchable encryption scheme that supports personalized search and improves user search experience remains a very challenging task. In this literature the problem of personalized multi-keyword ranked search over encrypted

data (PRSE) while preserving privacy in cloud computing has been solved. With the help of semantic ontology WordNet, a user interest model for individual user by analysing the user's search history, and adopt a scoring mechanism to express user interest smartly has been built. To address the limitations of the model of "one size fit all" and keyword exact search, two PRSE schemes for different search intentions has been proposed. Extensive experiments on real-world dataset validate our analysis and show that proposed solution is very efficient and effective.

Cong Wang et al. [10] Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results in the search over the encrypted data and further ensures the file retrieval accuracy. Specifically, the statistical measure approach is explored, from information retrieval to build a secure searchable index in terms of relevance score. , A one-to-many order-preserving mapping technique has enabled to properly protect those sensitive score information. It facilitate efficient server-side ranking without losing keyword privacy.

Zhihua Xia et al. [11] secure tree based search scheme over the encrypted cloud data, which support multi keyword search. The vector space model and tree based model combined in the index structure and query generation to provide keyword search. These scheme provisions together correct keyword search and flexible dynamic operation on document collections. In order to statistical attacks, spirit terms are additional to the key vector for outstanding search effects. Multi keyword search outline can understand linear search period and deal with the deletion and insertion of documents flexibly.

III.PROPOSED WORK

System model and Threat Model

Secured keyword Search is constructed with data owner module, data user module and Cloud administrator Module. Data owner enables the registration and file uploading functionality along the dataowner first builds a secure searchable tree index and then generates an encrypted documentcollection. Also data owner provided with securely distribution the key information, document updating responsibility and document decryption to the authorized data users. Data user is a requestor to the document of the data owner uploaded. Cloud administrator stores the encrypted document collection. The Attack launches cipher text-only attack as the cloud server only knows the encrypted document collection and the searchable indextree. In Term Frequency statistical attack, it deduces or even identifies certain keywords through analyzing histogram and value range of the corresponding frequency distributions.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

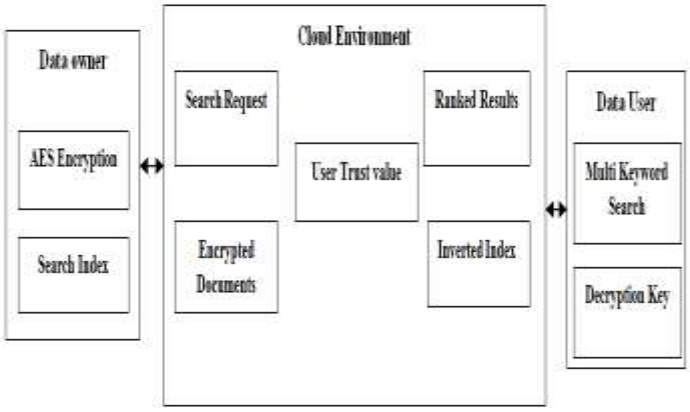

Figure 1: Architecture of proposed Multikeyword Encrypted data Search

3.1 Modelling AES Encryption Algorithm

It is a symmetric encryption algorithm. The encryption process uses a set of especially derived keys called round keys. These are applied, along with other operations, on an array of data that holds exactly the data to be encrypted. The block to be encrypted is just a sequence of 128 bits. AES works with byte and it converts the 128 bits into 16 bytes. Array is called state array.

3.2 Establishing Secure Lightweight Data Matching Algorithm

In this module, the data matching mechanism in terms of Index generation and data searching on the encrypted data and computing the trust of the user after obtaining the access to the data define. The Search model initializes based on vector space model and Tf-IDF for the documents in the cloud space. The process is as follows

3.3.1. Index Construction for Documents

In index construction, it is to construct a tree node for each document in the repository using index tree structure for easy searching and to carry out the iteration of the IDDFS. A tree node for each document in the collection has to be generated. These nodes are the leaf nodes of the index tree. Then, the internal tree nodes are generated based on these leaf nodes. Data structure of the node contain id, document, Identifier. This used to identify the document for keyword.

3.3.2. Iterative deepening Depth First Search

It is employed as retrieval process for keyword on the vector space model. It is termed as search strategy for the keyword and data retrieval process. It is developed with Result set which contains the relevance score and Identifier to query keyword. The data points in the list are ranked in descending order and conduct the deep

search on number of iteration. Result is updated frequently in specified time intervals. The detail architecture model is described in the figure1.

Algorithm: IDDFS Search

If (node u is not a leaf node)
Then
If (RScore(Du, Q)>Kth Score)
GDFS (u, hchild);
GDFS (u, lchild);
Else
return
End if
Else
If (RScore(Du, Q) <Kth Score)
Then
Delete the element with the smallest relevance score from RList
Insert a new element (RScore (Du, Q), U,FID) and sort all the elements of R List
End if
Return
End if

3.3.3. Query Confidentiality Preserving

The proposed process protects the Index Confidentiality and Query Confidentiality in the known cipher text mode is achieved by establishing the phantom terms. An experimental method to further improve the security is to break such exact equality into unable randomness to disturb the relevance score calculation.

3.3.4. Iterative User Trust Computation

This process is used to evaluate the disclosure of document of data owner by data user to another user is determined in terms of data provenance mechanism. It is carried out by employing similarity estimation or similarity matching algorithm such as classification algorithm to the cloud server to detect the misbehaving data user in the cloud server in order to revoke their access and provide information about their misbehaviour among the other user and data owners.

Algorithm: User Trust Computation

Data instance {d1, d2,d3}

Extracted data = D

Partition D

If ((P1|P2|P3)=(d1,d2d3))

Revoke ()

Else

Increase the User Rating

Revoke ()

Update the Trusted User inform with malicious data

Deny the user from the further access.

IV. RESULTS AND DISCUSSION

This section analyse the security of the Multikeyword search over the encrypted data. The different size of word format file or notepad file is taken for evaluation.

5.1. Precision

It is the number of real top-k documents in the retrieved document. Positive predictive value is the fraction of relevant instances among the retrieved instances. Precision is the number of correct feature divided by the number of all returned feature space.

$$\text{Precision} = \frac{\text{Truepositive}}{Truepositive + FalsePositive}$$

True positive is a number of real positive cases in the data and false negative is number of real negative cases in the data. The precision value will be increase due to phantom terms are added to the index vector to obscure the relevance score calculation, so that the cloud server cannot identify keywords by analysing the TF distributions of special keywords.

Proposed scheme retrieves the search results through exact calculation of document vector and query vector in order to represent its precision value. Many existing scheme such as similarity-based multi-keyword ranked search scheme, the basic scheme precision loss due to the clustering of sub-vectors during index construction. The test is repeated 16 times, and the average precision is 91 percentages. The search precision of scheme is described in the figure 2

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

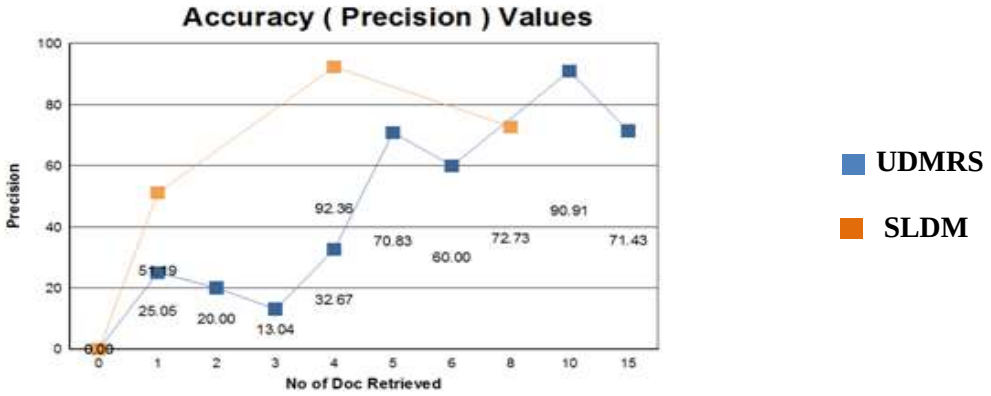

Figure 2: Performance measure in terms of Precision against the document retrieved

5.2. Privacy

The larger rank privacy denotes the higher security of the scheme. The privacy is also depends on the relevance score and standard deviation which is considered as balance parameter. The user trust is denoted as privacy value. The proposed mechanism produces the high security by calculating misbehaving user in the group. The Analysis of user trust behaviour against the data confidentiality is depicted in the figure 3.

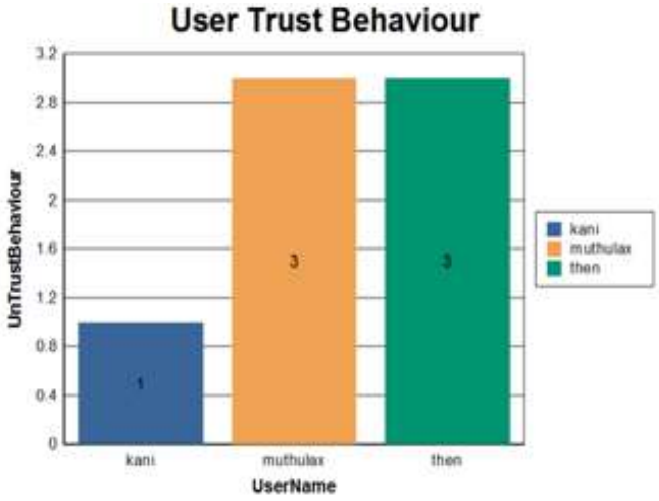

Figure3.Analysis of User trust behaviour against various data users

The figure 3 illustrates user trust is computed using the similarity measure. The proposed Scheme obtains better precision values and also achieves high search efficiency. In addition, the parallel execution of search process can increase the efficiency.

When the size of dictionary is fixed, the deletion of a document takes nearly logarithmic time with the size of document collection. The Precision measure for the number of document retrieved is described in table 1.

Table 1: Performance Comparison of Multi key search technique over encrypted data

Technique	Precision for 2 Documents retrieved	Precision for 4 Documents retrieved	Precision for 6 Documents retrieved
UDMRS– Existing System	0.20	0.326	0.60
SLDM – Proposed System	0.70	0.923	0.80

5.3. Efficiency Analysis

Time cost of index tree construction is almost linear with the size of document collection, and is proportional to the number of keywords in the dictionary. Due to the dimension extension, the index tree construction of proposed scheme is slightly more time-consuming. . Although the index tree construction consumes relatively much time at the data owner side, it is noteworthy that this is a one-time operation.

V.CONCLUSION AND FUTUREWORK

To design and implement a secure efficient encrypted data matching technique for several contributions has been made to proposed system in terms of constructing an index, vector for search keyword, employing iterative deepening depth first search for accurate retrieval of results. Also results is has been secured using phantom terms. The greatest challenges have been reached in this work towards maintaining high level of security against various kinds of threats, effectiveness and accuracy. The search process completely reduces the communication cost and time. The results yield that scheme is lightweight as processing steps of the secure model is less and revoke mechanism is faster when compared with other existing approaches. The system is tested with data's. Based on the results the scheme can term to be extremely lightweight.

In Future work, query expressiveness in terms of concept and semantic can be include in order to increase the accuracy of the retrieved results. The data owner needs to store the unencrypted index tree and the information that are necessary to recalculate the IDF values. Such an active data owner may not be very suitable for the cloud computing model. It could be a meaningful but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only, meanwhile reserving the ability to support multi-keyword ranked search.

REFERENCES

[1] Research Directorate Staff, "Securing the cloud with homomorphic encryption," Next Wave, vol. 20, no. 3, pp. 1–4, 2014.

[2] B. T. Prasanna and C. B. Akki, "A Comparative Study of Homomorphic and Searchable Encryption Schemes for Cloud Computing," Int. J. Adv. Stud. Comput. Sci. Eng. IJASCSE, vol. 4, no. 5, 2015.

[3] 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.

[4] K. Li, W. Zhang, C. Yang, and N. Yu, "Security Analysis on One-toMany Order Preserving Encryption-Based Cloud Data Search," IEEE Trans. Inf. Forensics Secur., vol. 10, no. 9, pp.1918–1926, 2015.

[5] S. Kamara and C. Papamanthou, "Parallel and dynamic searchable symmetric encryption,"Lect. Notes Comput.Sci. (including Subser.Lect. Notes Artif.Intell.Lect.NotesBioinformatics), vol. 7859 LNCS, pp. 258–274, 2013.

[6] Jiadi Yu ; Peng Lu ; Yanmin Zhu ; GuangtaoXue ; MingluLi " Toward SecureMultikeyword Top-k Retrieval over Encrypted Cloud Data IEEE Transactions on Dependable and Secure Computing (Volume: 10, Issue: 4, July-Aug. 2013.

[7]. Zhangjie Fu ; Xingming Sun ; Nigel Linge ; Lu Zhou Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonymqueryIEEEtransactions on Consumer Electronics, Volume: 60, Issue: 1, February 2014.

[8]. Hongwei Li ; Dongxiao Liu ; Yuanshun Dai ; Tom H. Luan ; Xuemin Sherman She Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage " IEEE Transactions on Emerging Topics in Computing Volume: 3, Issue: 1, March 2015.

[9]. Zhangjie Fu ; Fengxiao Huang ; KuiRen ; JianWeng ; Cong Wang "Privacy- Preserving Smart Semantic Search Based on Conceptual Graphs Over Encrypted OutsourcedData"IEEE Transactions on Information Forensics and Security, Volume: 12, Issue: 8, Aug. 2017.

[10]. Wang C., 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,Aug. 2012.

[11]. ZhihuaXia,Xinhui Wang, Xingming Sun, QianWang,"A Secure and Dynamic Multi- Keyword Ranked Search Scheme over Encrypted Cloud Data" IEEE Transactions on Parallel and Distributed Systems,Volume: 27, Issue: 2,pp. 340 – 352,2015.