International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-05, Issue-07, Oct 2019

# Providing Privacy For Keyword Search In Encrypted Cloud Using Term Frequency & Document Frequency

[1]Adarsh K Thampi, [2]Dr. S.P Swornambiga

[1]MPhil Scholar, [2]Associate Professor, [1,2]Department of Computer Application, CMS College of Science and Commerce, Coimbatore & India. [1]adarshkthampi@outlook.com, [2]swornagoms@gmail.com

**Abstract: Cloud computing is developing as an innovative computing model which offers a flexible and commercial strategy for data managing and resource distribution. Here the main problem to be addressed is security and privacy. It became a major worry for the cloud users, to solve this Searchable Encryption (SE) method is proposed to provide an efficient retrieval of encrypted data. But, the absence of trivial search is still a problem proving concern in existing SE systems. This paper proposed an effective privacy-preserving search technique over encrypted cloud data. Most of the existing work only support a single element search in queries which decreases the effectiveness. The key advantages of this proposed method are the ability to provide a keyword search in a single query with multiple instance. Term Frequency & Document Frequency (TFDF) strategy for giving a protected access to the cloud client. To guarantee this privacy TFDT provide a better security approach to the cloud data. One of the most well-known ways for this is through keyword-based retrieval. Keyword-based retrieval is search method which is applied in plaintext. Our proposed method is providing a satisfy search result using an Index based approach and this method was used based on term frequency and document frequency (TFDF) methodologies and provide a better searchable result and their outcome were provided.**

***Keywords - Cloud, Privacy, Encryption, Keyword Search and TFDF Approach.***

## I. INTRODUCTION

Cloud computing defines a variation of different types of computing ideas that comprise a large number of computers linked through a communication network. Broader concept of touched infrastructure and shared facilities laid the basis of cloud computing. The term cloud is not simply the synonym term for the Internet rather, the cloud is where we go to use technology when we need it, for as long as we need it and does not need to install anything on our desktop/pc [1]. It has been intended as the next group information technology construction for enterprises, due to its long list of extraordinary advantages in the IT history. Here the computing resources are shared by many users. The benefits of cloud are to everyone, from individual users to organizations. Moreover, there is no need to pay for the technology when we are not using it. The cloud thus can be both software and infrastructure [2]. One of the most popular ways for this is through keyword-based retrieval. Keyword-based retrieval is a regular search procedure generally applied in plaintext situations, in which clients recover applicable files dependent on keywords. An instrument called Keyword Search Based Encryption considers unscrambling the searched outcomes just as searching for wanted documents. Data encryption confines client's capacity to perform keyword search [3]. It also demands the protection of keyword privacy because few keywords contain important information about data files.

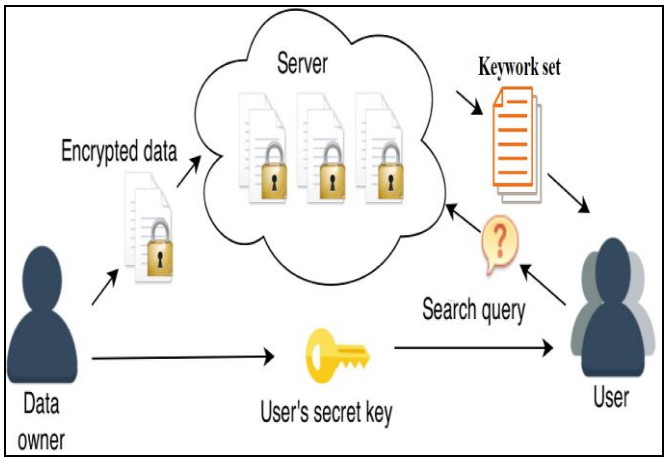


**Fig 1:** Architecture of Keyword Search in Cloud

 DOI : 10.35291/2454-9150.2019.0462 

Here encryption is implemented if both the owner and user of the data are different Privacy preserving keyword search enhances searchable encryption where the keyword search is performed on encrypted cloud data and in addition it will not reveal any data about either keyword or document. The user outsources the encrypted dataset and encrypted file to the remote server. With the encrypted query as input, server will retrieve the corresponding document. Without the security analysis for frequency information in their scheme, it is not clear whether such sensitive information disclosure could lead to keyword privacy [4]. While considering, the quantity of data clients and documents in cloud, it is monotonous for the search administration to give keyword inquiry and give result likeness positioning to meet the data retrieval need. Related deals with searchable encryption center around single keyword search, and infrequently separate the search results. Nonetheless, these ways are not very useful because of their high computational complexity for both the cloud sever and user [5, 6]. On the contrary, more functional precise intent solutions, such as searchable encryption (SE) schemes have made particular contributions in terms of efficiency, functionality and protection. Searchable encryption schemes permit the user to store the encrypted data to the cloud and execute keyword search over cipher-text. This paper proposed a TFDF method for providing a secure access to the cloud user.

The organization of this paper is as follows. In section 1 provides the introduction about our concept, Section 2 holds a literature review section, it shows various authors approaches, in Section 3 discussed about problem statement from existing work, in Section 4 provided about our proposed approach, in Section 5 proposed work experimental result and finally Section 6 contains the a conclusion about this paper.

## II. LITERATURE REVIEW

The most prominent work done by Chang Liu et al [7], Analyze the data can be put away in the cloud safely, individuals encode their data before redistributing to the cloud, which causes searching on a lot of encrypted data to turn into a requesting task. Customary searchable encryption plans give a scope of ways to deal with search on encrypted data, however they just help definite keyword search. Accurate keyword search isn't appropriate for cloud storage frameworks, since it doesn't permit clients making any spelling mistakes or configuration irregularities, which significantly decrease the framework ease of use. As far as we could possibly know, the moderately most practical plan distributed so far which supports fuzzy keyword search is the "Trump card based Fuzzy Set Construction". In this paper creator present the "Word reference based Fuzzy Set Construction", in which every keyword is comparing with considerably less fuzzy keywords. This improvement enormously decreases the record size, subsequently diminishing the storage and correspondence overheads.

Research in the keyword and document security proposed by Hoi Ting Poon [8] discover their enthusiasm for the zone of privacy-secured searching as businesses keep on receiving cloud advances. Much of the recent efforts have been towards incorporating more advanced searching techniques. Although many have proposed solutions for conjunctive keyword search, it is only recently that researchers began exploring phrase search over encrypted data. In this paper, author present a scheme that incorporates both functionalities. Their solution makes use of symmetric encryption, which provides computational and storage efficiency over schemes based on public key encryption. By considering the statistical properties of natural languages, they were able to design indexes that significantly reduce storage cost when compared to existing solutions. Their solution allows for simple ranking of results and requires a low storage cost while providing document and keyword security. By using both the index and the encrypted documents to performs searches, our scheme is also currently the only phrase search scheme capable of searching for non-indexed keywords.

In the paper "A Practical and Secure Multi-keyword Search Method over Encrypted Cloud Data", Cengiz Orencik ; Murat Kantarcioglu et al [9] dissect the cloud computing innovations become increasingly more well known each year, the same number of associations will in general re-appropriate their data using strong and quick administrations of clouds while bringing down the expense of equipment ownership. In spite of the fact that its advantages are invited, privacy is as yet an outstanding worry that should be tended to. Creator propose an effective privacy-saving search technique over encrypted cloud data that uses minhash capacities. The vast majority of the work in writing can just help a solitary element search in inquiries which diminishes the adequacy. One of the primary points of interest of Their proposed technique is the capacity of multi-keyword search in a solitary question. The proposed technique is demonstrated to fulfill versatile semantic security definition. They likewise join a compelling positioning ability that depends on term frequency-backwards document frequency (tf-idf) estimations of keyword document sets. Their examination exhibits that the proposed plan is demonstrated to be privacy-safeguarding, proficient and successful.

Efficient Multi-keyword Search over Encrypted Data in Untrusted Cloud Environment, P. Pandiaraja ; P. Vijayakumar [10] present cloud computing paradigm, with each business moving to the cloud computing, it has become mandatory that the data be moved to the public cloud environment and managing the same without any security breach. The main problem was privacy not

 DOI : 10.35291/2454-9150.2019.0462 

maintain to secure access the data in local and public cloud storage, and another problem not securely to searching the encrypted data in biggest cloud storage .Many of the researchers to be motivated this work under the single data owner setting model and multi-owner data setting model. As more and more data are stored in the databases provided by a single cloud service provider, the data is owned by multi persons. In this approach, we have proposed a new scheme to deal with preserving the privacy in encrypted cloud data using multi-keyword searching. To enable cloud server to perform access securely to searching the data with no cognizance about the real data of both keywords and trapdoors.

The Paper, "Fuzzy Multi-Keyword Query on Encrypted Data in the Cloud", Xiu-Jin Shi ; Sheng-Ping Hu [11] Find the security of cloud data has become a difficult problem that needs to be solved urgently. Data encryption is an effective means to ensure the security of data in the cloud. However, the retrieval of encrypted data is very different from the retrieval of plaintext. The traditional plaintext retrieval method is no longer suitable for the retrieval of the cipher text. In the previous fuzzy search strategy, although fuzzy keyword sets that building on keyword dictionary can realize the fuzzy search. But when the keyword set is constantly increasing, the fuzzy set will be exponentially growing. So, the retrieval efficiency will be very difficult to make people satisfied. On the basis of the existing encryption data fuzzy search scheme, abandoned the original fuzzy set of keywords, the search request of keywords and keyword Dictionary of keywords matching correction method to realize the fuzzy search, improving the efficiency of retrieval. At the same time, the Chinese and English data mixed more and more, and the existing fuzzy retrieval scheme supports only a single language retrieval, using English and Chinese comparison table convert Chinese keyword to English keyword to support Chinese and English keyword search, making the precision ratio improved, greatly improving the user's search experience.

Nagesh Jadhav [12] with the appealing choices of cloud computing, cloud becomes a significant infrastructure of enterprise IT. An outsized amount of information is being outsourced to the cloud. Before outsourcing the data is being encrypted. Cryptography makes straightforward but important functionalities like search operations over cloud information difficult. The quality and economical plaintext keyword search technique has no result on encrypted information. The prevailing searchable cryptography schemes support exclusively precise keyword search, not support linguistics-based search. Therefore, they propose a research theme whereby the linguistics relationship and word of the question keyword are thought-about with the help of information structures.

## III. PROBLEM STATEMENT

The data storage in cloud is one of important feature. In Cloud computing, data owners are persuaded to re-appropriate their gigantic data the executives' frameworks from neighborhood destinations to business open cloud for incredible adaptability and monetary advantages. Security is one of the frequently referred to issues with cloud registering. Cloud clients face security dangers both from outside and inside the cloud. The cloud client is in charge of use level security. The cloud provider is responsible for physical security. When users outsource their private data onto the cloud, the cloud service providers can control and monitor the data without having permission from the data owner. They were many approaches were implemented by many researchers but it provides a security issues, in existing work they concentrate on privacy-preserving keyword search. But the provide a security issues, third party may easily access the data without knowing the cloud user and word seeks over encoded also provided. issue of a safe distributed storage administration provides a great disadvantage over this. Elements of an encryption scheme and the execution protocol for encrypted query processing take long time to process it.

## IV. PROPOSED WORK

Cloud computing store data into the cloud on interest top notch applications and administrations. Data owners can be remembered from the weight of data storage and support. In the event that cloud server isn't in the equivalent confided in area, the data might be in danger. Data encryption makes successful data use that there could be huge measure of redistributed data files. In any case, for securing privacy of the data, delicate data must be encrypted before putting away, which obsoletes customary data usage dependent on plaintext keyword search. In this way, empowering an encrypted cloud data search administration is valuable. This paper proposed a TFDF strategy for giving a protected access to the cloud client. To guarantee this privacy, clients subsequently expected to scramble the data before re-appropriating it onto cloud. Yet, scrambling the data to be transferred will enormously influence the data use. In addition, data owners may impart their re-appropriated data to various clients. The data clients then again need to recover the data files they are keen on. One of the most well known ways for this is through keyword-based retrieval. Keyword-based retrieval is an average search method broadly applied in plaintext situations, in which clients recover applicable files dependent on keywords. Clients can recover files through keyword-based search as opposed to recovering all the encrypted files back. Such strategy can be applied in plain content search situations

 DOI : 10.35291/2454-9150.2019.0462 

*A. Data Preprocessing*

In cloud provider for storing the data, here data have to process for the further usages and the document of the user must be processed. The document preprocessing performs all the text operations on the dataset and extract meaningful keywords. For providing a keyword search mechanism it must be considered as to extract the keyword to process for the keyword search results. Here keyword processing KP is KP={F_L ,F_s ,F_st} is a data preprocessing function consist of various functions, lexical analysis (F_L), stop word removal (F_s) and stemming (F_st). The term frequency and inverse document frequencies are calculated to compute the weight of each term.

On the basis of the weight, the top terms are selected for creation the dictionary keyword set.

$$W_{ik} = TF * DF = TF * \frac{1}{DF} = F_{ik} * \log \frac{N}{N_k}$$

Here F_ik=Frequency of term i in document.

$N_k$=no. of documents which contains term i.

keyword set is generated by extracting important keywords from all documents.

Here the polynomial function is implemented used for hiding the encrypted keywords and search patterns. To provide privacy search, a secure inner product method is used.

*B. Keyword Search using TFDF*

The data is encrypted and stored in a remote server so clients can't play out a plaintext keyword search to recover the data. Searchable encryption is generally utilized in cloud storage to perform keyword search. The searchable file is encrypted and stored with a separate key. A unique identifier allotted for each record and each user can only access one record. The encrypted keyword index is uploaded along with the encrypted file. All the words in the file are encrypted as the keyword index. If the user or client looking for the particular document, they can retrieve the document with the respective encrypted keyword. This paper proposed a TFDF method for providing a secure access to the cloud user. TFDF stands for term frequency document frequency, and the TFDF weight is a weight often used in information retrieval and text extraction using a keyword search mechanism. The mechanism performs in two terms: the first computes the normalized Term Frequency (TF). The occasions a word shows up in a document, partitioned by the absolute number of words in that document; the subsequent term is the Document Frequency (DF), processed as the quantity of the documents in the amount isolated by the quantity of documents where the particular term appears. Term frequency estimates how as often as possible a term happens in a document. Since each document is diverse long, it is conceivable that a term would show up significantly more occasions in long documents than shorter ones.

In the case of the term frequency TF(t,d), the simplest choice is to use the raw count of a term in a document, i.e. the number of times that term t occurs in document d. If we denote the raw count by $F_{t,d}$, then the simplest TF scheme is TF (t,d) = $F_{t,d}$.

$$TF(t,d) = \frac{1}{2} + \frac{1}{2} * \frac{F_{t,d}}{\max\{F_{t,d} : t \in d\}}$$

Document frequency measures how important a term is. While computing TF, all terms are considered equally important. However, it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. fraction of the documents that contain the word, obtained by dividing the total number of documents by the number of documents containing the term

$$IDF(T.D) = \log \frac{N}{\{d \in D : t \in d\}}$$

We now combine the definitions of term frequency and inverse document frequency, to produce a composite weight for each term in each document. The tfdf scheme assigns to term t a weight in document b given by

$TFDT_{t,d} = TF_{t,d}$ X $DF_t$

For each document Di ∈ D, the set of features $FE_i$ = { $FE_{i1}$, . . . , $FE_{iz}$} that characterize the document is extracted. In our case, those features are composed of two values FEij = (Wij , RSij ). The first one is a keyword $W_{ij}$ of the sensitive document. The second one is the relevancy score (RS), which is based on TFDF value of the keyword $W_{ij}$ for document $D_i$

Once the data is encrypted it is uploaded into the cloud server along with the relevant words for particular document which will be used as index for searching. The importance for the particular word in the document is identified using TFIDF. This enhancement of indexing only relevant words will reduce the index size and it retrieves only the more relevant documents. In the case of previous method, all the words are indexed. If common word is used as documents and so comparing to the present method it results more irrelevant documents.

If you are using *Word*, use either the Microsoft Equation Editor or the *MathType* add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation *or* MathType Equation). "Float over text" should *not* be selected.

## V. EXPERIMENTAL RESULT

This section provides a analyze the proposed method to validate the efficiency and effectiveness of our proposed scheme. For this the system is implemented by Java language. Here we consider documents retrieval mechanism and also search encryption methods. The fact that each of the documents were retrieved by the search engine implies

 DOI : 10.35291/2454-9150.2019.0462 

that they are all at least marginally relevant to the query. In proposed system, the TF-IDF value of each word is used as a key for each word encryption. The key differs for each word encryption. Unless the attacker has the knowledge of method (i.e., TF-IDF) used for encryption, it is very difficult to hack the contents.

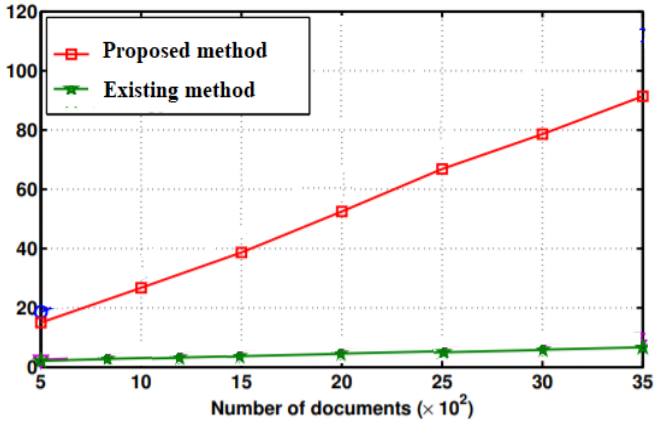


**Chart 1:** Keyword search comparison

This model retrieves the most relevant documents from a large document collection in the cloud based on the keyword search method. From chart 1 it shows the comparison of the existing and proposed method which provide a no. of keywords searches in a document.

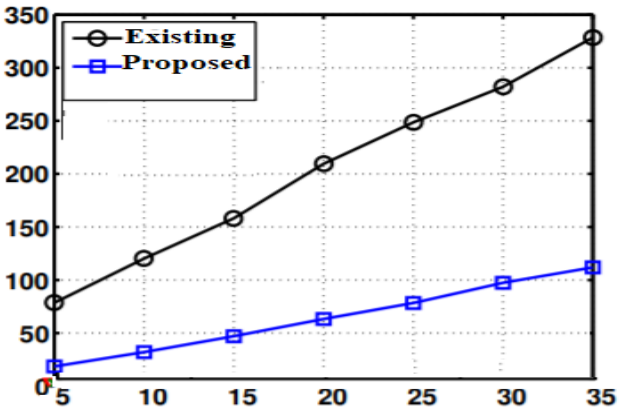


**Chart 2:** Security Comparison

Our method produces a high no. of searchable data and provide more security which were shown chart 2.

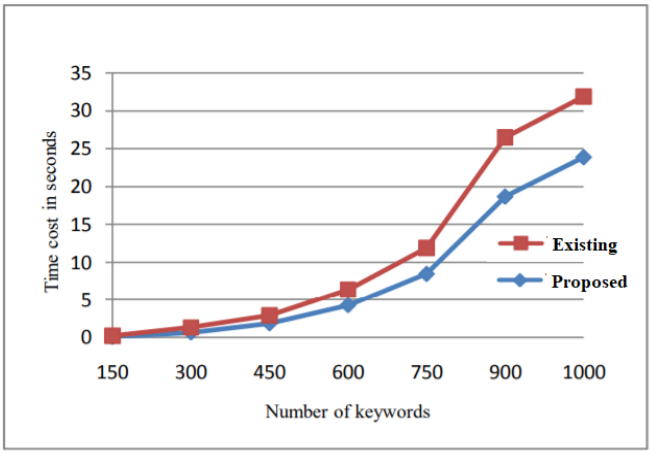


**Chart 3:** Time Cost Performance comparison

Therefore, this experimental result is performed in well manner by the comparison of time. In existing it take more time to encrypt the data and also more energy but our proposed method produces less time utilization. And also to ensures that the search algorithm is conducted in a secure way to protect important search privacy in the whole search process.

## VI. CONCLUSION

Security is one of the great issues in the cloud architecture. Most of the researchers were concentrating on this domain to solve the security and privacy issue, but they were failed in some lack cases. This paper analyzes various security threads and propose a TFDF method for the utilization of Keyword search mechanism in the cloud. Encrypted Keyword search provide a better security for the cloud user. This approach ensures that only the most relevant items are retrieved by the user and it also provide more secure access to the cloud user. We analyze our proposed work and provide our experimental results which shows better result compared with existing one.

## VII. REFERENCES

[1]. Li M, Yu S, Ren K, Lou W, "Securing personal health records in cloud computing: patient centric and fine grained data access control in multi owner settings". Springer, Berlin Heidelberg, pp 89–106, 2010.

[2]. S . Kamara and K. Lauter, "Cryptographic cloud storage, in RLCPS", Springer, January 2010.

[3]. Wang Jie, Yu Xiao, Zhao Ming, Wang Yon, "A Novel Dynamic Ranked Fuzzy Key-word Search Over Cloud Encrypted Data", IEEE, 2014.

[4]. Salam, MdIftekhar, et al, "Implementation of searchable symmetric encryption for privacy-preserving keyword search on cloud storage." Human-centric Computing and Information Sciences 5, no. 1, 2015.

[5]. S.Menaka and N.Radha "Text Classification using Keyword Extraction Technique" International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Volume 3, Issue 12, December 2013.

[6]. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data", Proc. IEEE 30th Intl Conf. Distributed Computing Systems (ICDCS), 2010.

[7]. Chang Liu et al, "Fuzzy Keyword Search on Encrypted Cloud Storage Data With Small Index", 978-1-61284-204-2/11, IEEE CCIS2011.

[8]. Hoi Ting Poon, "An Efficient Conjunctive Keyword and Phase Search Scheme for Encrypted Cloud Storage Systems", **IEEE Xplore:** 20 August 2015.

 DOI : 10.35291/2454-9150.2019.0462 

[9]. Cengiz Orencik; Murat Kantarcioglu, "A Practical and Secure Multi-keyword Search Method over Encrypted Cloud Data", IEEE Xplore: 02 December 2013.

[10]. P. Pandiaraja ; P. Vijayakumar, "Efficient Multi-keyword Search over Encrypted Data in Untrusted Cloud Environment", IEEE Xplore: 05 October 2017.

[11]. Nagesh Jadhav, "Semantic Search Supporting Similarity Ranking Over Encrypted Private Cloud Data", ISSN 2349-4409, IJEERT.

[12]. Prasanthi Sreekumari, "Privacy-Preserving Keyword Search Schemes over Encrypted Cloud Data: An Extensive Analysis" IEEE, 2018.

[13]. Q. Xu, H. Shen, Y. Sang and H. Tian, "Privacy-Preserving Ranked Fuzzy Keyword Search over Encrypted Cloud Data," 2013 International Conference on Parallel and Distributed Computing, Applications and Technologies, Taipei, 2013.

[14]. X. Jiang, J. Yu, F. Kong, X. Cheng and R. Hao, "A Novel Privacy Preserving Keyword Search Scheme over Encrypted Cloud Data," 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, 2015, pp. 836-839. 2015.

[15]. Tseng, Ching-Yang, et al, "Efficient privacy-preserving multi-keyword ranked search utilizing document replication and partition." CCNC, IEEE, 2015.

 DOI : 10.35291/2454-9150.2019.0462 