Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776790

A NOVEL APPROACH FOR FUZZY KEYWORD SEARCHABLE ENCRYPTION USING N-GRAM IN CRYPTOGRAPHIC CLOUD ENVIRONMENT

[1]Panchal Mital Nikunj, [2]Dr. Dushyantsinh B. Rathod
[1]PhD Scholar, Faculty of Engineering and Technology, Sankalchand Patel University,Visnagar, Gujarat, India.
[2]Professor & HOD, Computer Engineering Department, Ahmedabad Institute of Technology, Ahmedabad
[1]panchal.mital@gmail.com ,[2]dushyantsinh.rathod@gmail.com

Abstract: Cloud computing has been now a very familiar word among IT fraternities since almost two decades. The benefits of cloud services are limit less. Cloud computing provides several benefits for users, like the ability to access computing resources as needed, and the flexibility to pay for use. The convenience of having access to resources, data, and information on demand is a big plus for consumers. They only have to pay for what they use. Users and businesses should be able to verify the charges they're being billed for cloud computing and storage solutions. These services offer various capabilities to store and process data. But, on the other hand, consumers face serious difficulties in finding the most suitable cloud services. Storing data on data storage servers, such as mail servers and file servers, in encrypted form reduces security and privacy risks. Encryption is the process of encoding messages or information so that only authorized parties can read it. Multiple encryption techniques are developed to encrypt the data. Encryption of search results has become an important problem. There are major three reasons for it. (1) search engines usages to search data is the only way to find suitable material on internet (2) Data is stored over cloud. (3) there is a serious trust issues about data misuse. Many computer scientists are now interested in the problem of encrypted search, as it is an important part of major processes over internet. Some methods are more practical, secure, or flexible than others. These schemes do not support accurate search result when there is a spelling mistake. To address these challenges proposed schemes is developed works on fuzzy keyword search. The proposed scheme developed with n-gram technique for comparison of keywords on server with AES +ABE encryption algorithm for privacy preserving of the data.
Keywords: , cryptography, encryption schemes, fuzzy keywords, trapdoor

I. INTRODUCTION

Cloud storage lets users store data in the cloud and access it from any device. Cloud servers are becoming increasingly popular for hosting local data, as they are convenient, affordable, and easy to access. Data is encrypted over the cloud to protect your privacy. Encryption is a way to encode messages so that only authorized people can read them. Multiple encryption techniques have been developed to encrypt the data. This leads to a key problem in encrypted data search. There are two key-based encryption algorithms, symmetric encryption algorithm (also called key algorithm) and asymmetric encryption algorithm (or called public key

algorithm).The difference is that symmetric encryption algorithms use the same key for encryption and decryption (or the decryption key is easily derived from the encryption key), while asymmetric encryption algorithms use different keys for encryption and decryption, and the decryption key cannot be derived from the encryption key.

Here are the major risks linked with cloud storage:

Data loss: sharing sensitive information to a wrong person can risk the data loss and data misuse. User without knowing the authorized person is acting as third-party shares confidential information can be a victim of cyber threats.

Data leakage: Hackers know that cloud storage services are valuable targets, because they offer a quick and easy way to store valuable data. Of course, all reputable cloud storage services take security very seriously. These companies have systems in place to protect your data from unauthorized access. However, users themselves create opportunities for theft, such as by using simple passwords for their cloud storage accounts rather than the complex hard-to-crack ones that are enforced within your organization..

Vulnerable to attack: Cloud storage providers have full access to your data, which puts your data at risk from security and technical issues with the providers themselves.

In today's world most of the communication is done using electronic media. Data security plays a vital role in such communication. Hence there is a need to protect data from malicious attacks. This can be achieved by Cryptography. The earlier encryption algorithm is Data Encryption Standard (DES) which has several loopholes such as small key size and sensible to brute force attack etc. and it cannot be provide high level, efficient and exportable security. These loopholes overcome by a new algorithm called as Advanced Encryption Standard (AES).

However, it is a very difficult to search the most suitable services or products for ordinary consumers, as there are so many services and products in cloud. Meanwhile, data outsourcing enables the data owner and the cloud service provider not in a same trusted domain, making the data owner not manage data in real time. It is a common practice to encrypt sensitive information before outsourcing.

II. RELATED WORK

Techniques that help in maintaining privacy and data security are derived from mathematical concepts and predefined algorithms. Some of the most used techniques are:

Phases in searchable encryption is inevitable. Multiple phases that deals with trapdoor generation, key generation, encryption, index generation and decryption of documents are categorized in different phases. The major advantage is to use them for separate stages on platforms[1]. Privacy leakage is not an issue that provide meaningful information to trespasser as its generated with probabilistic based trapdoor.

Cryptography - Cryptography is the process of generating secret codes, enabling the confidentiality of communication through an insecure channel[1]. It protects against unauthorized parties by preventing unauthorized alteration of use. Generally speaking, it uses a cryptographic system to transform a plaintext into a cipher text. using most of the time a key. It has different Encryption and Decryption algorithms to do so.

Cipher Text - This is the scrambled message produced as output from Encryption algorithm. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different cipher texts.

Encryption - Encryption is the process of converting data, in plain text format into a meaningless cipher text by means of a suitable algorithm. The algorithm takes secret key and plain text as input and produces cipher text.

Decryption - Decryption is converting the meaningless cipher text into the original information using decryption algorithms. The decryption algorithm is inverse of encryption algorithm. This takes key and cipher text as input and produces original plain text[3].

Symmetric Key Cryptography - It uses the same secret (private) key to encrypt and decrypt its data. It requires that the secret key be known by the party encrypting the data and the party decrypting the data.

Asymmetric Key Cryptography - Asymmetric uses both a public and private key. This allows for distribution of your public key to anyone with which they can encrypt the data they want to send securely and then it can only be decoded by the person having the private key[3].

To search over encryption data multiple search techniques are invented which are described below.
In the paper of [3] they proposed an efficient verifiable keyword-based semantic search scheme. Comparing to most of the existing searchable encryption schemes, the proposed scheme is more practical and flexible, better suiting users different search intensions. Moreover, the proposed scheme protects data privacy and supports verifiable searchability, in the presence of the semi honest server in the cloud computing environment.

In the paper of [3] proposed an effective approach to solve the problem of synonym based multi-keyword ranked search over encrypted cloud data. The main contributions are summarized in two aspects; synonym based search and similarity ranked search. The search results can be synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. The vector space model is adopted combined with cosine measure, which is popular in information retrieval field, to evaluate the similarity between search request and document.

In the paper [4] solved the problem of multi-keyword ranked search over encrypted cloud data and establish a variety of privacy requirements. Among various multi keyword semantics, we choose the efficient similarity measure of coordinate matching i.e. as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords and use inner product similarity to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi keyword semantic without privacy breaches, they proposed a basic idea of MRSE using secure inner product computation. Then they have given two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

In the paper [4] proposed the first verifiable SSE scheme, which offers data privacy. Verifiable searchability and efficiency, in the presence of an unusually strong adversarial server in cloud computing environment. The rigorous security analysis together with through experimental evaluations on a resource-constrained device using real data sets confirms that the proposed VSSE realizes our design goals and is a promising solution to mediate the conflicts between data usability and data privacy in such scenario.

I reviewed related work and illustrate the difference of different keyword-based search techniques. Li at al [6] firstly proposes a fuzzy keyword search scheme over encrypted cloud data. Wang et al [5] proposed a secure ranked search scheme. This scheme supports only single keyword search while fuzzy keyword search scheme tacks the problem of minor typos and format inconsistencies. Chai et al [4] propose a verifiable search scheme which can prove correctness and completeness of result efficiently. Based on VSSE and fuzzy keyword search, Wang et al [6] proposes a scheme supporting both verification and fuzzy search, but the scheme ignores result ranking.

One is that most of these schemes supports only exact keyword search. That means returned result is completely dependent on whether query terms users enter match pre-set keywords. The other one is that most existing searchable schemes assume that the cloud server is honest-but-curious. However, Chai et al [4] notice that the cloud may be selfish to save its computation or download bandwidth. That is, the cloud server might conduct only a fraction of search of search operation or return a part of result honestly.

Besides, Fu et al [9] recently proposed a multi-keyword search scheme in encrypted cloud environment which can achieve synonym query. The main contribution of the scheme is that it solves the problem of synonym search.

III. Encrypted Domain Model

A. Architecture of search over cloud data

Architecture for search services involves two different entities: the data owner and the cloud server. The data owner has a collection of data documents that are outsourced to the cloud server in encrypted form. To enable the ability to search encrypted documents for effective

data utilization, data owners first create an encrypted searchable index before outsourcing, and then index. And outsource both the encrypted document collection to the cloud server.

Figure 1 .Architecture of search over cloud data[1]

B. Processing Steps of Encryption

1. Single keyword search- A non-linear substitution step where each byte is replaced with another according to a lookup table (S-box).
2. Multi keyword search - A transposition step where each row of the state is shifted cyclically a certain number of times.
3. Fuzzy keyword search - A mixing operation which operates on the columns of the state, combining the four bytes in each column.

As keyword in files and looked through words in secret entrances have been covered up to safeguard the protection, it is challenging to find out the right relating connections. As per the calculation Decrypt [3], on the off chance that there are adequately not right comparing connections, the records can't be accurately decoded to get the characters of archives.

C. Processing Steps of Search technique

1) SETUP – In This Algorithm The Data Owner Initiates The Scheme To Generate The Random Key And A Secret Key.
2) GENINDEX - To further develop the pursuit effectiveness, an image based tree to store components in a limited image set is constructed.
3) PREPROCESS: The information proprietor examines the plaintext record assortment D and concentrates the particular keywords of D, meant as W.

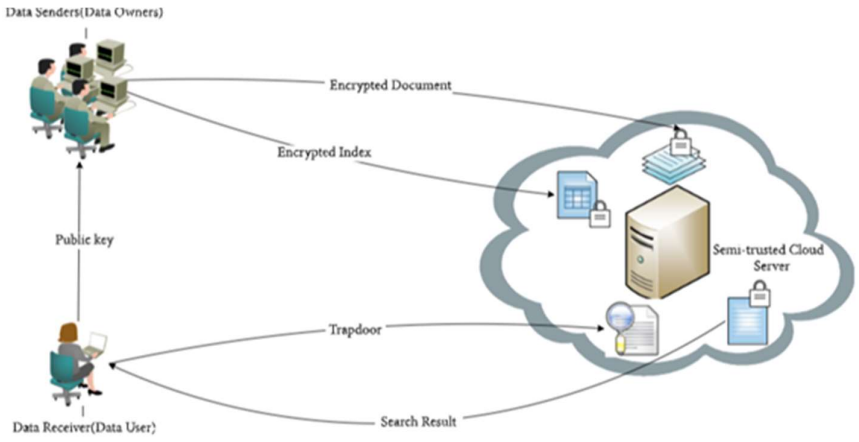

Figure 2. Process of Searchable Encryption

4) The information proprietor processes the score of all particular keywords on premise of presence in number of reports from assortment.
5) GENQUERY - When the client inputs the inquiry terms Q, first forms term likeness tree TST(Q,v,m) and executes catchphrase semantic expansion, getting the drawn out question.
6) SEARCH - Upon getting the keywords, the cloud server plays out the inquiry activity over the list G. The hunt is basically to track down a way in G as per the pursuit demand, from the root hub to the leaf hub. The presence of a way shows that the questioned words occur no less than one of the designated information records.
7) CHECK AND RANK - When the client gets the positioned result from the cloud server, he can confirm the accuracy and fulfillment of item.

D. K-NN Model to secure keywords

KNN permits productive calculation of the k-closest neighbors over a scrambled informational index and will be applied to encode both our records and queries. Keywords from the list of words signify the security boundary, which ought to be sufficiently long to protect brute force attacks. Given a vector v, its I-the component is indicated by a word of vector. Encryption is performed aspects before encryption. Accordingly, the cycle of adding fake aspects in can be ignored. Efficiency is achieved to fuzzy search the encrypted data is achieved. Semantic security measurements are bunched with query word, trapdoor as combined input criteria. The scheme can define different words from the input and stimulate the model to perform the probability of produce multiple words matched.

IV . PROPOSED METHODOLOGY

The system contains three entities: a cloud server, a data owner and a receiver. Moreover, the data owner wants to send confidential files to the cloud which are allowed the assigned receiver to access the data. The exact procedures are as follows: First, the data owner extracts a group of keywords from documents and builds an secure index including keyword ciphertexts and documents. Second, the data owner encrypts the files by symmetric encryption and uploads the encrypted file and keyword ciphertext index to the server. Third, the receiver generates a trapdoor for a query keyword and sends it to the server. Finally, after receiving the trapdoor, cloud server runs the test algorithm and outputs the search results.

V PROPOSED ALGORITHM

Step 1 The Data Owner runs the KeyGen(s) with k-anonymity algorithm to generate Apub and Apriv.
Step 2 It gives Apub to the User.
Step 3 The User can adaptively generate the trapdoor TW for any keyword $W \in \{0, 1\} *$ of his choice.
Step 4 The user A sends the server two words W0, W1 on which it wishes to be the keywords. (The attacker did not previously ask for the trapdoors TW0 or TW1)
Step 5 The user can continue to generate for trapdoors TW for any keyword W of his choice as long as $W \neq W0$, W1.
Step 6 Eventually, the user receives the files containing w0 or w1.

VI. IMPLEMENTATION

The proposed algorithm is implemented in windows machine in core i5 processor , 8 GB RAM with JAVA programming language with the help of external libraries for encryption and indexing process. The effectiveness of the proposed algorithm is further compared with different approaches. The implementation screenshots are added below:

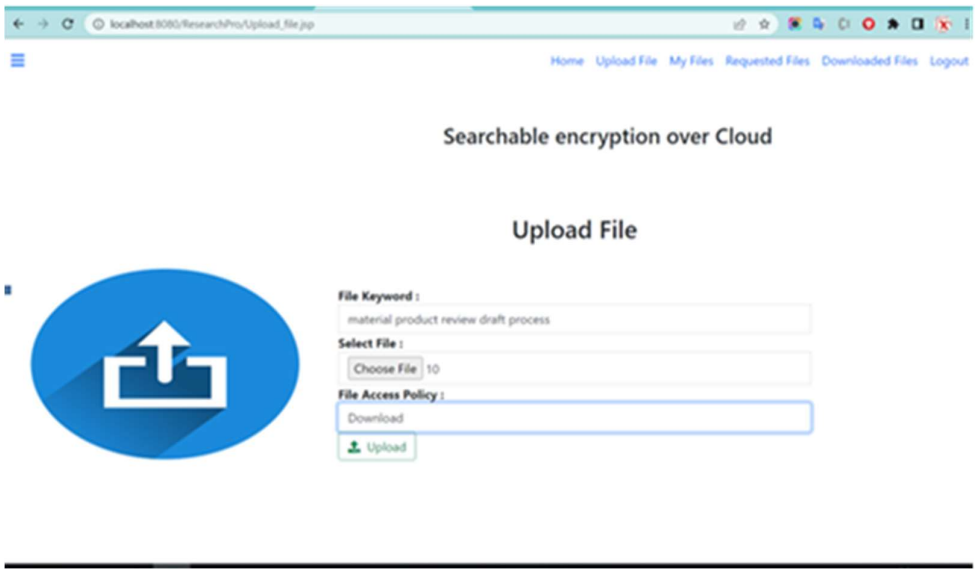

Figure 3 : File upload by data owner on cloud server

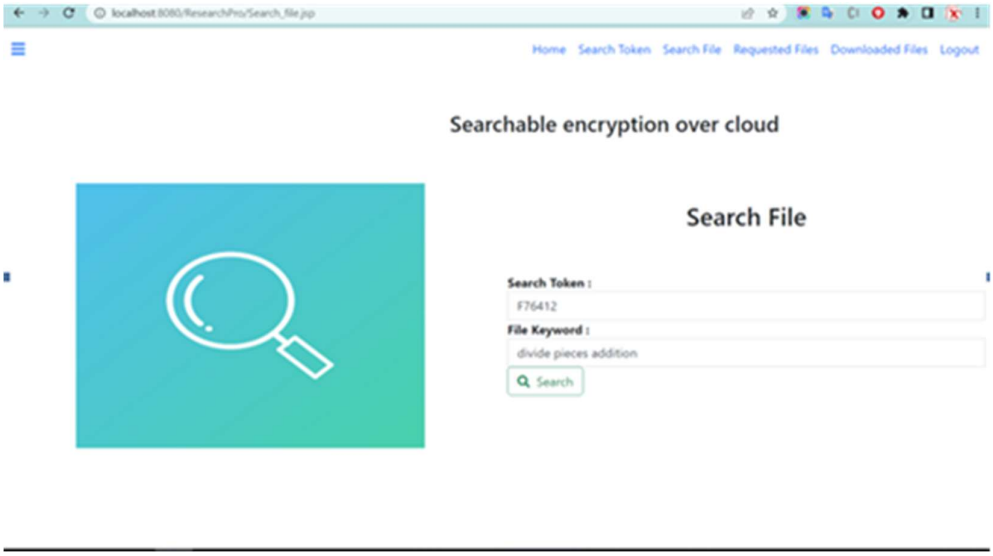

Figure 4: Search File with keywords and trapdoor by data user

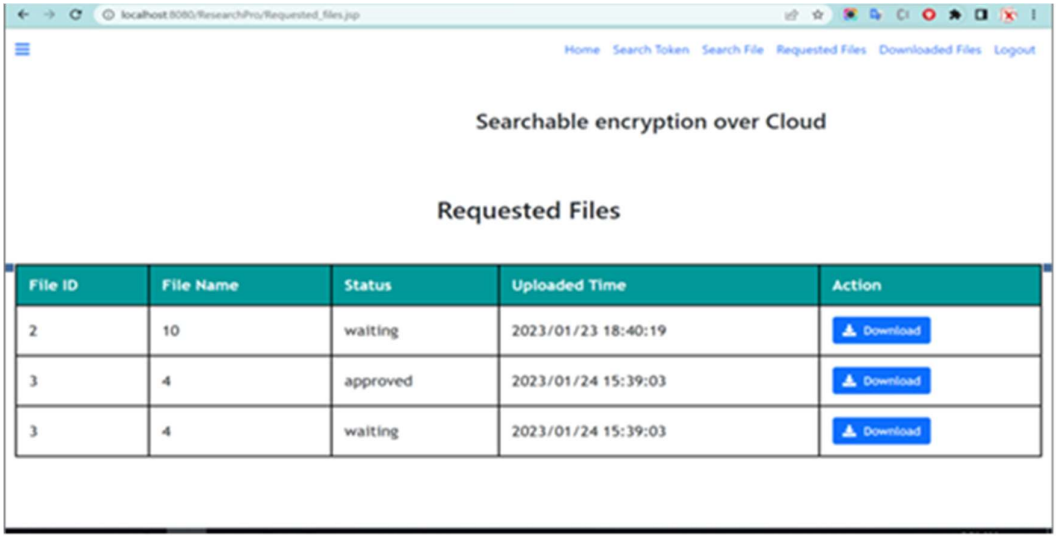

Figure 5 : Requested file from cloud server

VII. RESULTS AND COMPARISION

We implemented our proposed work with existing approaches with Enron Email Dataset. As shown in table-1 we got the better time complexity as compared to other approaches for different size of documents.

size of documents (MB)	**1000**	**1500**	**2000**	**3000**
bloom filter[2]	18	40	58	80
tree-based index[7]	40	75	105	148
btcs search[8]	320	600	800	1000
Proposed Algorithm	16.52	34.7	49.2	65.901

Table 1: Time complexity comparison (in seconds)

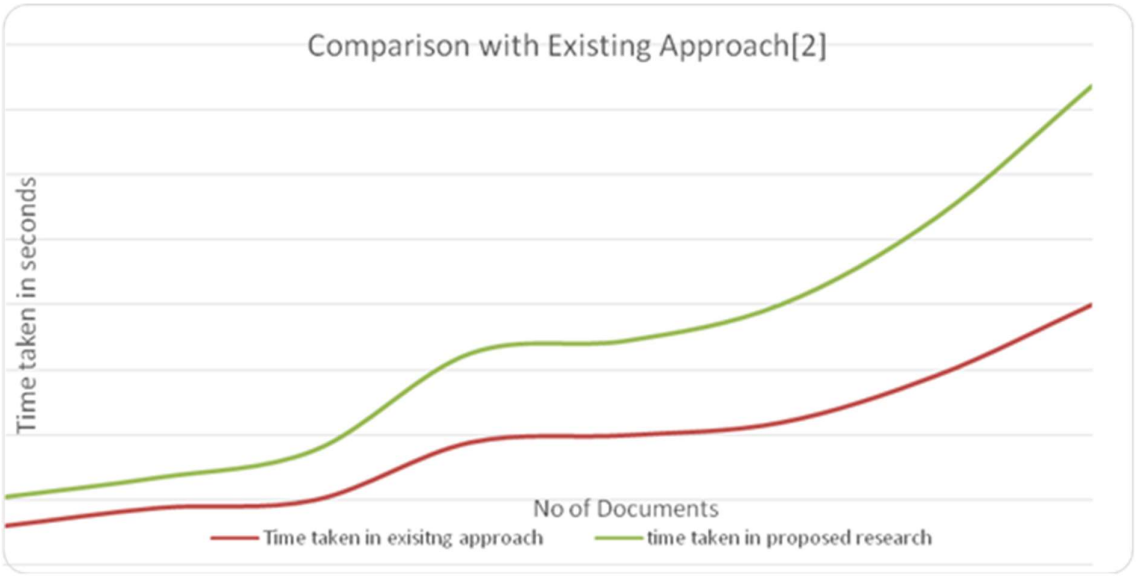

Figure 6: Comparison with existing approach

Also as shown in figure 6 we have compared our proposed work with [2] and found that our system gives less time as compared to [2] for whole search process.

VIII. CONCLUSION

Multiple techniques are invented and analyzed to perform searchable encryption based on type of keywords. Fuzzy keywords are encrypted in addition to the words formation even ambiguously related to the word entered but found in the document. Fuzzy keyword searchable encryption techniques generation is proposed by us is compared with existing approach ad we got the satisfactory results for fuzzy keyword search. We also get less time to perform the search process.

IX. REFERENCES

[1] Yuan Ping 1, Wei Song 2, Zhili Zhang 1 , Weiping Wang 3 and Baocang Wang 2,4,, " A Multi-Keyword Searchable Encryption Scheme Based on Probability Trapdoor over Encryption Cloud Data," MDPI, 2020.

[2] Li, Mengmeng, et al. "Multi-keyword Fuzzy Search over Encrypted Cloud Storage Data." Procedia Computer Science 187 (2021): 365-370.

[3] Meng Meng Li, Guijuan Wang, Suhui Liu, and Jiguo Yu, " Multi-keyword Fuzzy Search over Encrypted Cloud Storage Data ," Elsevier, school of computer science,P.R.china, 2021.

[4] Jing Chen, Kun He, Lan Deng, Quan Yuan, Ruiying Du, Yang Xiang, and Jie Wu, "EliMFS: Achieving Efficient, Leakage-resilient, and Multi-keyword Fuzzy Search on Encrypted Cloud Data," Proceedings of IEEE 2017.

[5] Qin. Liu and Yu. Peng, Shuyu Pei, Jie Wu, Tao Peng, Guojun Wang, "Prime Inner Product Encoding for Effective Wildcard-based Multi-Keyword Fuzzy Search," Proceedings of IEEE 2020.

[6] Jing Chen, Kun He, Lan Deng, Quan Yuan, Ruiying Du, Yang Xiang, and Jie Wu, EliMFS: Achieving Efficient, Leakage-resilient, and Multi-keyword Fuzzy Search on Encrypted Cloud Data – IEEE-2017

[7] Li, Jin, and Xiaofeng Chen. "Efficient multi-user keyword search over encrypted data in cloud computing." Computing and Informatics 32.4 (2013): 723-738.

[8] Xinrui Ge1, Jia Yu 1,2,3, Chengyu Hu 4, Hanlin Zhang1, And Rong Hao1, "Enabling Efficient Verifiable Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proceedings of IEEE-ACCESS,2018.

[9] J Hong Zhu1, Zhuolin Mei1(B), Bing Wu1, Hongbo Li1, and Zongmin Cui2, "Fuzzy Keyword Search and Access Control over Ciphertexts in Cloud Computing," Proceedings of SPRINGER 2017, School of computer science and technology, wuhan, china.

[10] Shaojing Fu1,2 · Qi Zhang1 · Nan Jia1 · Ming Xu1, "A Privacy-preserving Fuzzy Search Scheme Supporting Logic Query over Encrypted Cloud Data," SPRINGER-2020.

[11] Jin Li, Xiaofeng Chen, "Efficient multi-user keyword search over encrypted data in cloud computing," Computing and Informatics, Vol. 32, 2013, 723-738,.