# A Secure and Dynamic Multi Keyword Search on Encrypted Data in Cloud

[1]Ms. Priyanka M. Abhale, [2] Prof. M.B. Vaidya
[1]ME ,[2]Associate Professor
[1]Department of Computer Engineering,
[1]Amrutvahini college of Engineering, Sangamner, India

*Abstract :* Cloud storage is very famous because it provides extra benefits over the traditional storage solutions. The encryption techniques offer safety in cloud it plays a primary function in outsourced to the cloud statistics. The encrypted facts in cloud retrieving over the cloud garage are complex. There are many looking strategies for retrieving the encrypted statistics from the cloud. This paper specializes in a Multi-key-word ranked Search mechanism over encrypted records, which gives secured statistics retrieval with excessive performance. It concludes that, Multi-key-word ranked search is the most useful Method for looking for encrypted statistics inside the Cloud. It gives more performance than single key-word. Searchable Encryption (SE) permits a server of cloud to do keyword searching on encrypted facts on the non - appearance of the information clients without studying the fundamental plain texts. Where, the previous accessible encryption plans guide, only one or conjunctive keyword search, even as a different scheme capable of carry out expressive keyword search from bilinear pairings over composite request group. This paper, suggest an expressive public-key accessible encryption scheme in the top- request agencies, which permits catchphrase seeks rules (i.e., predicates, get entry in the systems) to communicate in conjunctive, disjunctive or any monotonic Boolean formulation and accomplishes significant overall execution development over existing plans.

***Index Terms*** **- Cloud computing, Trapdoor, expressiveness, access structure, encrypted keywords.**

## I. INTRODUCTION

Recently as a new industrial version, computers that do work for you, but that are stored somewhere else and maintained by other companies has attracted lots interest from both the world of college and industry. A major gain of cloud is that it useful things supplies completely and totally unlimited storage abilities and elastic aid provisioning. In order to reduce the capital and operational costs for hardware and software program, plenty of IT businesses and people are paying someone else to do something their statistics to cloud servers instead of building and keeping their own statistics facilities. Cloud computing has been thought about as a new version of large business IT basic equipment needed for a business or society to operate, that can organize big aid of calculating, storage and packages, and enable customers to enjoy existing everywhere, convenient and on demand network access to a shared pool of configurable calculating useful things supplies with excellent wasting very little while working or producing something and very little money-based overhead.

In public system storage cloud, as information can be put away in distributed data centres, all the data centres there won't not be a one central authority that controls. Other than the administrators of the storage cloud providers themselves would have the get to the data if its keep in plain course of action. To protect the security of the data, data proprietors use cryptographic methodology to encode the data by one means or another that solely customers who are allowed to get the data as depicted by the access policies having capacity to do as such. We tend to raise to display approach as a strategy frame encoded data get to.

The cloud provider providers (CSP) that preserve the statistics for users might also get speaking the truth about something bad to customers sensitive records on the not being there present of known approval. A famous approach to protect the facts confidentiality is to turn into secret code the statistics in advance than paying someone else to do something. However, this can purpose a big cost in phrases of statistics usability. Downloading all the statistics from the cloud and change secret codes into readable messages within a large area is obviously not having common sense. In this paper, we recommend a public-key based totally communicating a lot of thought or emotion SE layout in most important-order groups, that's specially good for key-word searching for over unreadable in situations of a couple of data owners and many statistics users which include the cloud-based college records system that hosts paid someone else to do something available data from many colleges or from different colleges.

In the above cloud-based college system, to and the connection among department and stud name or class, person who works to find information may also problem to search question with an access structure (i.e. Predicate) (department = computer AND (stud name = Ram OR class = ME)). In order to help data use and sharing, it is incredibly clearly connected with or related to have a Searchable Encryption (SE) layout which permits the cloud provider company to look over unreadable data for the legal clients (which include scientific researchers or college authority) without studying statistics about the hidden plaintext In the realistic programs, look for predicates (i.e., guidelines) should be communicating a lot of thought or emotion such that they can be expressed as not having a connection, conjunction, or any Boolean system of very important phrases. [1].

## II. PROBLEM DEFINATION

Develop a framework to determine how to securely search any document from cloud in the form of encrypted data with the help of TRAPDOOR And also how to Store data in Secure form on cloud . To protect the user privacy from third party or unauthorized access, encryption algorithm is uses as well as RSBS algorithm is use to reduce the searching time.

## III. REVIEW OF LITERATURE

Wang, N. Cao, K. Ren, and W. Lou. [2] This paper define the problem of secure ranked key-word search over encrypted cloud facts and give effective protocol this will satisfy the secure ranked search operations using some piece of information over

keyword The data owner outsources encrypted files and their index to the cloud server then this encrypted file converted in byte stream. The data user send search request to server after that server identifies the particular user and sends files using ranking to users.

Zhangjie Fu, Xinle Wu, Chaowen Guan and Xingming Sun, [3] This paper propose an efficient multi keyword fuzzy ranked search scheme this is capable of address the above point out issues. First, we develop a brand new method of keyword transformation primarily based on the uni-gram, if you want to concurrently improve the accuracy and creates the capability to handle other spelling mistakes. The design goal are it help spelling errors like netward or netwrok , it take guarantee of privacy and maintain safety, the support of report and key-word updating, it generate the result in step with score. Stemming algorithm, Bloom Filter, Locality-Sensitive Hashing (LSH) this 3 essential strategies are used in this layout. The stemming, bloom and encryption is achieve by the usage of trapdoor technology center. The seek time and index production is important in trapdoor.

B. Wang, S. Yu, W. Lou, and Y. T. Hou, [4] In this paper we characterize and take care of the testing issue of protection saving multi-catchphrase positioned seek over encoded cloud information (MRSE), and build up an arrangement of strict protection prerequisites for such a safe cloud information use framework to end up plainly a reality. Among different multi-watchword semantics, we pick the effective standard of "facilitate coordinating", i.e., whatever number matches as could be expected under the circumstances, to catch the similitude between look question and information reports, and further utilize "internal item closeness" to quantitatively formalize such rule for comparability estimation. We initially propose an essential MRSE plot utilizing secure inward item calculation, and afterward altogether enhance it to meet distinctive protection necessities in two levels of danger models.

C. Wang, K. Ren, S. Yu, and K. M. R. Urs, [5] In this paper, we distinguish the framework prerequisites and difficulties towards accomplishing security guaranteed accessible outsourced cloud information administrations. This paper display a general system for this, utilizing accessible encryption methods, which permits encoded information to be sought by clients without spilling data about the information itself and clients inquiries. The factual measure approach, i.e., significance score, from data recovery to construct a safe accessible file, and build up a one-to-many request saving mapping system to legitimately ensure those delicate score data. The subsequent plan can encourage productive server side positioning without losing watchword protection.

Zhihua Xia, Xinhui Wang, Xingming Sun, and Qian Wang, [6] In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. The generation of query and also construction of index merge the vector model and TF-IDF model. In this Greedy Depth-first Search algorithm and KNN algorithm are used for tree based construction of keyword searching in multi keyword rank search and to encrypt the index and query. This paper describe the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme based on two secure search schemes BDMRS and EDMRS schemes. The UDMRS scheme is constructed using the KNN algorithm and the privacy preserving is achieved by BDMRS scheme.

M. Kuzu, M. S. Islam, and M. Kantarcioglu ,[7] This paper offers an efficient scheme for similarity seek over encrypted facts. To achieve this, this utilize a state-of-threat algorithm for fast near neighbour seek in high dimensional spaces referred to as locality sensitive hashing. To make certain the confidentiality of the touchy information, This paper offer a rigorous safety definition and show the security of the scheme beneath the provided definition. In addition, this paper provide a actual international software of this scheme and verify the theoretical results with empirical observations on a actual dataset. Locality Sensitive Hashing (LSH) is an approximation algorithm for near neighbor search in high dimensional spaces . The basic idea of LSH is to use a set of hash functions to map objects into several buckets such that similar objects share a bucket with high probability .

## IV. SYSTEM ARCHITECTURE AND OVERVIEW

The system architecture of key-word search is proven in Figure, which is composed of 5 entities: a trusted trapdoor technology center who publishes the machine parameter and holds a master non-public key and is answerable for trapdoor era for the system, records owners who outsource encrypted information to a public cloud, data customers who're privileged to go looking and get right of entry to encrypted information, and a designated cloud server who executes the key-word seek operations for records users. To permit the cloud server to look over ciphertexts, the data proprietors append each encrypted file with encrypted key phrases. A records consumer issues a trapdoor request via sending a key-word get admission to structure to the trapdoor generation middle which generates and returns a trapdoor corresponding to the access structure.

There are four main contributions of this paper that follow:-

1) This system is based on the secure multi keyword ranked search over encrypted data in cloud and provides a protocol which fulfills the secure ranked search functionality with little relevance score information leakage against keyword privacy.
2) In contrast to previous solutions on multiple keywords search our scheme can achieve higher search efficiency by executing our "Ranked Serial Binary Search" algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.
3) Our search process is very efficient it performs multiple keyword matching in one round and display the sorted list of documents and also it can greatly reduce the search time and the storage cost of the searchable index.
4) The user interest model is used it maintain the history of searchable keywords with respective documents in the cloud and extensive experimental results demonstrate the effectiveness and efficiency of the proposed solution.

*A. System Architecture*

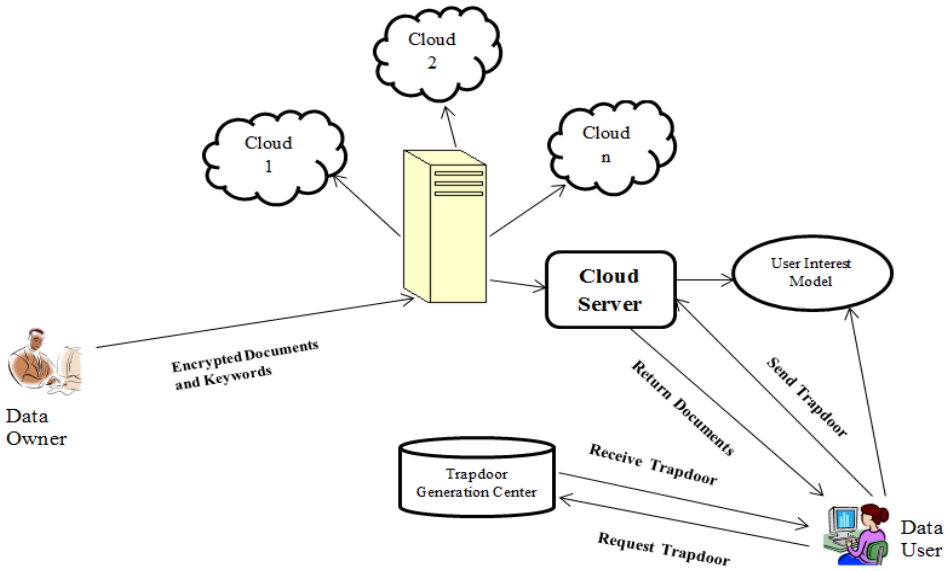


Fig. 1. Multi-keyword Search Architecture

*B. System Modules:*

1) Data Owner: In this module, Data owner will browse encrypt and upload the files. views all the uploaded files and transaction build on the files uploaded. Data owner comprises of data objects, authorization rules, Data encryption, key encryption, transaction administration. In order to prevent CSP from accessing data object, it is uploaded in encrypted on cloud and manage keys on key distribution centres.

2) Cloud Server: Cloud server will see all the uploaded files with encrypted attribute, authorize the clients furthermore, data owner and view the attackers and the transaction expand on the roles and the related files and furthermore the search transactions. Cloud is in charge of client validation of end client and data owners. This can be helpful for Inter-cloud situations, where information can go through various CSPs.

3) Data User: In this module, the client will register established on roles and search for the documents based on Content keyboard and request to the file and download with the secrete key for the Corresponding file from the cloud and downloads the file.

4) Trapdoor: A relied on trapdoor technology centre who publishes the device parameter and holds a master personal key and is answerable for trapdoor era for the machine.

5) User Interest Model: A consumer model is the gathering and categorization of statistics associated with a particular user. This model holds the history of searching keyword and also first of all cloud servers go to this model for checking the previously searched information.

*C. Algorithm:*

The Expressive keyword search scheme consists of five algorithm Setup, sKeyGen, Trapdoor, Encrypt , Test.

1) Setup(λ ,U) (PK,MSK). The setup algorithm takes the security parameter and the attribute universe description U as the input. It outputs the public parameters PK and a master secret key MSK.

2) Key Gen(MSK, S) SK. The key generation algorithm takes the master secret key MSK and a set of attributes S as input. It outputs a secret key SK.

3) Trapdoor(pars, pks, msk )TM. Taking the public parameter pars, the server public key pks and an access structure as the input, generates a trapdoor TM. This algorithm is run by the trapdoor centre.

4) Encrypt(PK,M,A) CT. The encryption algorithm takes the public parameters PK, a message M, and an access structure A as input. The algorithm will encrypt M and produce a ciphertext CT .

5) Test(pars, sks, CT, TM) → 1/0. Taking the public parameter pars, the server private key sks, a ciphertext CT associated with a keywords set W and a trapdoor TM for an access structure as the input, and generate outputs either 1 when the ciphertext satisfies the access structure of the trapdoor TM or 0 otherwise. This algorithm is run by the designated server.

The algorithm is executed by the owner to encrypt the keyword and produce a searchable encrypted index as follows:

1. Encrypt the document index using cipher under owners private key and users public key

2. Build a tree I for document collection as index,
3. Where document identifiers are stored in leaf nodes.
4. Let N represents a node in I, and we denote its form as ⟨fid, lc, rc, switch⟩. If N is a leaf node, fid is the document identifier, lc is the left child and rc right child.
5. If the children of node N have the same switch form, we add node N to the document group.

**Ranked Serial Binary Search (RSBS) algorithm:**

Input : Noised trapdoor: t1

The number of document to return: k

Encrypted record indexes: E

Output : Document request: D

1. Create the scores as an N zeros
2. For i = 1 to N do
3. For n = 1 to E do
4. Find the keywords seems in any of the s slice of the report
5. End for
6. End for
7. Sorted, indices=sort, (scores)
8. Acquire the top-k files
9. Go back D

## V. EXPERIMENTAL DETAILS

There are many components in this system and each component plays an important role in the ranked multi keyword search fig 2. is a Data Owner module from which the collection of all the documents uploaded by data user in which it shows the data owner name, file uploaded time and also a size of the document. The encrypted collection along with the secure index stored to the remote server. In this module, the data owners should be able to upload the files. The records are encoded before the documents are transferred to the cloud. The data owners are given a choice to enter the keywords for the record that are transferred to the server. These keywords are utilized for the ordering reason which enables the search to return values rapidly. These files when once available on the cloud, the data users should be able to search using keywords.

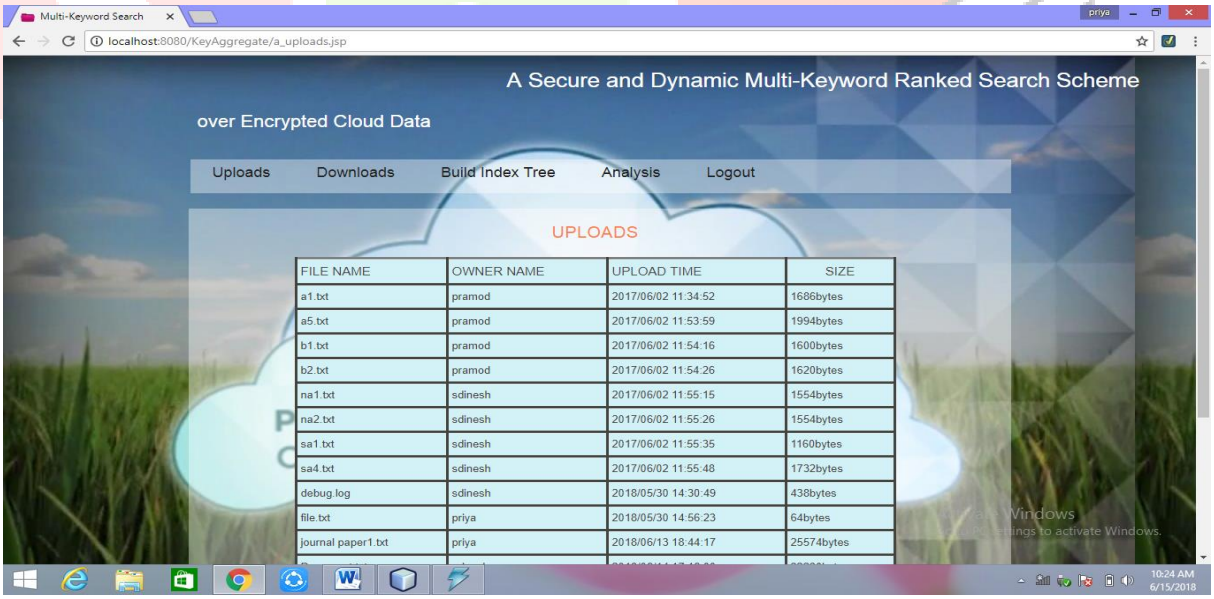


Fig.2. Data Owner Module

Fig 3 demonstrates the trapdoor generation. To recover the documents containing keywords the user needs to request public key to create trapdoors; If disconnected these owners information can't be recovered in time. If not client will get general public key and make one trapdoor for a keyword set utilizing TrapdoorGen algorithm. Initially the data user consolidates the query to influence them to seem as though one query then client will process the trapdoor of the search demand of connected keyword under his private key.

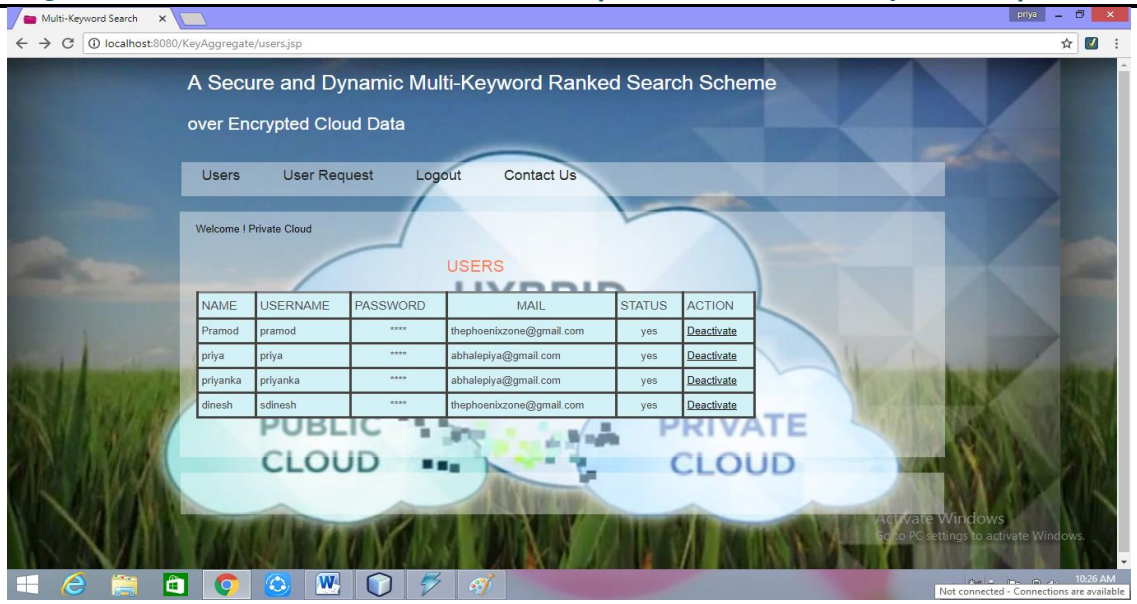


Fig.3. TPA Module

Fig.4 shows the search operation executed at the cloud server side comprises of registering the internal item count for all the files in the dataset. The file set or archive list while the quantity of keyword in the query enter by client. This is natural in light of the fact that the inquiry procedure needs to go over all the files in the dataset before the cloud server can get the final result. The inward item calculation is just identified with the length of the list, so the calculation time changes.


Fig.4. Multi-keyword search

## VI. RESULT ANALYSIS

We implement our scheme in java based on the java programming language . For testing we have hosted each entity on different machines. The cloud TPA has core i-3 processor with 4 gb RAM. Client system has i3 processor with 2 gb ram. On every system java runtime environment JRE-1.7 is installed. For development we have used jdk 1.7 and NetBeans IDE are used. For Database storage we have used mysql 5.3 database and also we have done the JDBC database connectivity.

The analysis graph shows that to encrypt and partition of our file data not vary to much with file size. To complete this analysis we perform execution on different files of different size and allows us to distinguish the actions to be taken to return the number of documents with respective time.

This time taken is for the operations such as key generation, file encryption, file partition and file upload on cloud. So analysis result shows that there is no much time variance even if file size increased. This graph shows the how much time it take to find the keyword in particular document and show the list of documents in which this keyword is appears. By using access structure user get the list of top document in which the data is stored after that user download the file .

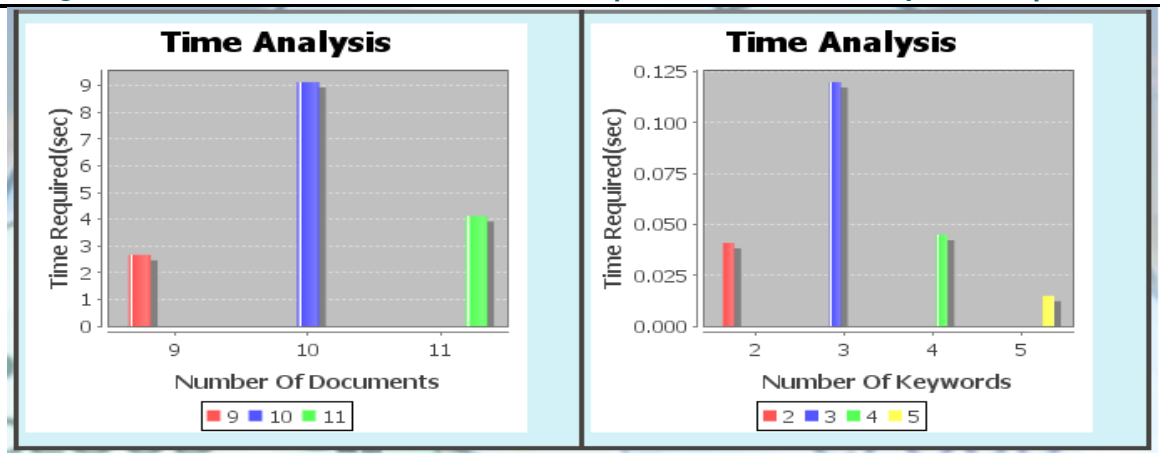


Fig.5.The time analysis of keywords and documents

In the previous works, when the client needs to recover records containing each of several keywords, he should give the server trapdoors for every one of the keywords exclusively and depend on a crossing point activity. This arrangement isn't attractive, it requires O(n) m search time, where n is the no. of documents and m is the no. of keywords in conjunctive string. As it were, the sever needs O(n) search time for every keyword in conjunctive string. While in our work, the client processes one trapdoor for all conjunctive keyword and sends it to the remote server and with one round over all conjunctive keywords, a server calls Search Index algorithm once on each trapdoor.

## VII. CONCLUSION

This paper focus on enhancing the efficiency and the security of multi-keyword top-k comparability search over encrypted information. At that point, keeping in mind the end goal to enhance the search efficiency, we design the collection of multi-keyword top-k search scheme, which separates the dictionary into numerous groups and just needs to store the top k documents of each word aggregate when building list or index. In future this proposed key-word searching technique proves efficient and return top most documents or files corresponding to submitted search terms. This proposed system reduces the searching time the usage of Ranked Serial Binary Search (RSBS) set of rules.

In Future we can extend our work by focusing users searching habits. Also can add sentiment analysis on users review about particular file, while suggesting cloud file to users. We will investigate supporting other multi keyword semantics (e.g., weighted question) over encrypted information, honesty check of rank request in search result and protection ensures in the stronger threat model.

## ACKNOWLEDGMENT

I would like to thank my project guide Prof M. B. Vaidya and project coordinator prof. S. K. Sonkar, Department of Computer Engineering, AVCOE, Sangamner for the valuable advice, support and the interest shown in this project by timely suggestions in this work. I would also like to thank my teachers for their encouragement, guidance, understanding and support.

## REFERENCES

[1] Hui Cui, Zhiguo Wan, Robert H. Deng, Guilin Wang, and Yingjiu Li, "*Efficient and Expressive Keyword Search Over Encrypted Data in Cloud*", Transactions on Dependable and Secure Computing Journal Of , Vol. , No., 2016.

[2] 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. 14671479, Aug. 2012.

[3] Zhangjie Fu, Member, IEEE, Xinle Wu, Chaowen Guan, Xingming Sun, Senior Member, IEEE, and Kui Ren, Fellow, IEEE, "*Toward Efcient MultiKeyword Fuzzy Search Over Encrypted Outsourced Data With Accuracy Improvement*", IEEE Transactions On Information Forensics And Security, Vol. 11, No. 12, December 2016.

[4] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "*Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud*," in Proc.IEEE INFOCOM, 2014, pp. 2112–2120.

[5] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "*Achieving usable and privacy-assured similarity search over outsourced cloud data,"* in Proc. IEEE INFOCOM, 2012, pp. 451–459.

[6] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang, Member, IEEE, "*A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data"*, IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, February 2016.

[7] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "*Efficient similarity search over encrypted data*", in Proc. IEEE 28th Int. Conf. Data Eng., 2012, pp. 11561167.

[8] Priya S , Ambika R, "*A Multi-keyword Ranked Search Scheme that is Dynamic and Secure over Cloud Data",* International Journal of Innovative Research in Science, Engineering and Technology Vol. 5, Special Issue 10, May 2016.

[9] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, "*Secure ranked multi-keyword search for multiple data owners in cloud computing",* in Dependable Syst. Networks (DSN), IEEE 44th Annu. IEEE/IFIP Int. Conf., 2014, pp. 276286.

[10] Xiaofeng Ding, Member, IEEE, Peng Liu and Hai Jin, Senior Member, IEEE, "*Privacy-Preserving Multi-keyword Top-k Similarity Search Over Encrypted Data",* IEEE Transactions on Dependable and Secure Computing,2016.

[11] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "*Public key encryption with keyword search"*, in Advances in Cryptology Euro crypt 2004. Springer, 2004, pp. 506522.

[12] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "*Searchable symmetric encryption: Improved denitions and effcient constructions",* in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 7988.

[13] E. Shen, E. Shi, and B. Waters, "*Predicate privacy in encryption systems"*, in Theory of Cryptography. Springer, 2009, pp. 457473.

[14] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "*Privacy-preserving multi keyword ranked search over encrypted cloud data",* IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222233, 2014.

[15] J. Baek, R. Safavi-Naini, and W. Susilo, *"Public key encryption with keyword search revisited",* in Computational Science and Its Applications. Springer, 2008, pp. 12491259.

[16] Jie Shi, Junzuo Lai , Yingjiu Li , Robert H. Deng , Jian Weng, *"Authorized Keyword Search on Encrypted Data"* Singapore Management University, 2012.

[17] Sasikala.K, Karthik.S, Santhana Mani.J "*Improved Approach For Encrypted Content Search In Mobile Cloud Service "* IRJET Volume: 03 Issue: 05 | May-2016.