ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor & UGC Approved Journal)*

*Website: www.ijirset.com*

**Vol. 6, Issue 8, August 2017**

# Keyword-Based Search Mechanism for Data Stored in Cloud

Shwetha [1], Shreya Ranjan [2], Dr.Vishwanth Y [3]

P.G Student, Department of Information Science, New Horizon College of Engineering Marathalli, Bangalore, India [1]

P.G Student, Department of Information Science, New Horizon College of Engineering Marathalli,, Bangalore, India [2]

Senior Assistant Professor, Department of Information Science, New Horizon College of Engineering Marathalli, Bangalore, India.[3]

**ABSTRACT:** Peoples now a day's rate of storing their personal and professional data on the cloud increasing rapidly. Data is getting transferred to remote area in larger chunks without knowing whether the server on which data is transferred is a trusted server or not. There is possibility of losing its contents in large size of data. In order to maintain privacy of documents which is being transferred on cloud environment, encryption of the document should be made before outsourcing it to the cloud. But after placing encrypted documents on cloud retrieving of the data is a tedious work. In order to retrieve the data on cloud computing there is so many techniques available. Among them keyword enabled search of the encrypted data is one of the most efficient and popular technique. In most of the case single keyword based technique are used. Because of its limitation not full fills the all it need. To increase efficiency and rapidness in searching multi-key word search can be applied. This paper represents a survey on a secure search scheme handled by single-keyword or multi-keyword ranked search over encrypted cloud data

**KEYWORDS:** Single keyword, Multi-keyword search, Ranked search, Encrypted cloud data, Security

## I. INTRODUCTION

Cloud computing depends on internet to delivery computer processing resources and data to computers and other devices on demand. The cloud costs of building and maintaining a private storage infrastructure. The company or organization's private and sensitive data like personnel file, emails, company records etc which is to be shared between selected company employees is stored and centralized into cloud server. That selected company employees choose customer, individuals or enterprises can transfer their local complex data system into the cloud to overcome from certain content from that data.

Since bulk of information are being added into cloud server there will be insecure feelings that anyone may hack these data. Hence Because of encrypted data on cloud we need to apply encrypted form of search to retrieve it. One of the most accepted ways is keyword based search method which allows customer to retrieve content of their interest.

**Searching Techniques:**

There several searching techniques involved to retrieve the encrypted data from cloud. Some of them as follow Searchable Encryption: It uses keywords as a tool to search complete encrypted data securely

- Fuzzy keyword searchable encryption

Fuzzy keyword search generally increases the system usability by returning the matching files when users searching inputs which are exactly matching the pre-defined keyword or closest possible matching files based on keyword similarity semantics when exact match fails

The most relatively feasible scheme published till now which supports fuzzy keyword search is the wildcard based fuzzy set. This technique eliminates the need for enumerating all the fuzzy keyword and resulted size of the fuzzy keyword set is reduced.

- Multi-keyword searchable encryptions

These systems build a secure index structure and send it along with encrypted document to the remote server. The specific users submit their request as secrete trapdoors which are integrated properly with stored index information.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Then the server uses the received trapdoor to search over the stored index and retrieves the matching encrypted document. These systems are designed to handle either a single keyword search or a Boolean scheme is not practical in the real world.

- Single Keyword searchable encryption

It is the most significant single-keyword search over encrypted cloud data. This scheme usually builds an encrypted searchable index such that, its content is hidden to the server, unless it is given to the specific trapdoor generated via secrete key(s).It always supports single keyword search. The main drawback of this scheme, it's not confronting able enough to express complex information needs.

- Plaintext fuzzy keyword search

These fuzzy searches have now reached in the information retrieval communication where it has received attention in the context of plaintext. The problem arises in the traditional information access paradigm by allowing user to search using without try and see approach for the finding relevant information based on approximate string matching. This retrieval construction suffers from the dictionary and statistics attack and fails to achieve the search privacy.

- These Boolean system allows user to specify their information using combination of Booleans operator AND, OR and NOT. It has several disadvantages there is no features of document ranking and it is difficult for a user to make a good search result.

## II. RELATED WORKS

**LITERATURE SURVEY:**

This literature survey mainly concentrated towards single keyword and multi keyword based encryption and also included other searching techniques

**1. Single keyword search**

- C.Wang et al,[1] proposed searching method to improve the efficiency of ranked keyword search algorithm. He also mentioned that one to many order preserving mapping function which allows the effective RSSE to be designed. The main disadvantage of single keyword search using with or without ranking is it will not retrieve data of user interest and it also have privacy issue.
- Deepali D.Rane et al,[2] proposed how to preserve the privacy of data on cloud and also sensitivity of keyword through ranking, indexing method. Ranking of the result introduced to improve the search result correctness and also improve the user searching experience. This research mainly discussed and implemented Lucene indexing algorithm. Demerits of existing system may be
  1. Single-keyword searching without ranking
  2. Boolean-keyword searching without ranking
  3. Single-keyword searching with ranking
  4. Rarely sorting of the result i.e. no index creation or ranking
  5. Single user search
- Y. –C.Chang et al[3] proposed to user assist for searching any kind of data using just keyword and downloading the data whenever is required
- D. Song D,Wagmer et al[4] proposed a searchable encryption in which each content treated as document under two-layered encryption construction

2. **Multi keyword search**

- Zhihua Xia et, al [5] proposed that a secure tree-based scheme over the encrypted cloud data, which supports multi keyword ranked search and dynamic operations on the document collection. This scheme supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. They construct a special keyword balanced binary tree as the index, and proposed a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. Also the parallel search process can be carried out to further decrease the time cost. By using the secure KNN algorithm, the security of the scheme is protected against two threat models. The Advantages of the proposed system are searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain and a secure tree-based search scheme over the

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. The disadvantages are the cloud service providers (CSPs) that keep the data for users may access user's sensitive information without authorization. In order to protect the data confidentiality is to encrypt the data before outsourcing. But this will become expensive in terms of data usability.

- Bing Wang et.al [6] proposed that problem of building a searchable encryption scheme based on the inverted index to overcome the aforementioned limitations. They construct our scheme through a series of novel designs based on the private set intersection protocol in [17]. They achieve secure and private matching between the query trapdoor and the secure index. They design a novel trapdoor generation algorithm so that the query related inverted lists are combined together secretly without letting the cloud server know which inverted lists are retrieved
- Yanzhi Ren et.al, [7] proposed a light-weight search approach that supports efficient multi-keyword ranked search in cloud computing system. The basic scheme employs the polynomial function to hide the encrypted keyword and search patterns for efficient multi-keyword ranked search. Then improve the basic scheme and propose a privacy-preserving scheme which utilizes the secure inner product method for protecting the privacy of the searched multi-keywords. The advantage of the proposed system is it analyzes the privacy guarantee of the proposed scheme and conduct extensive experiments based on the real-world dataset. The disadvantage is there is a possibility of leakage of information in cloud.
- Mikhail Strizhov et.al, [8] proposed a searchable encryption technique that enables secure searches over encrypted data stored on remote servers. They defined and solved the problem in multi-keyword ranked search over encrypted cloud data. In particular, they present an efficient similarity searchable encryption scheme that supports multi-keyword semantics. The solution is based on two building blocks: Term Frequency Inverse Document Frequency (TF-IDF) measurement and ring-LWE-based variant of homomorphism cryptosystem. The plus point of this system is it returns the matching data items in a ranked ordered manner. The Disadvantage is it supports only single keyword searches in traditional system.

Other searching techniques:

- E.-J. Goh et al, [9] proposed a technique that uses Bloom filters in order to construct the indexes for the data files. Bloom filter containing trapdoors (for each file) of all distinct words is built up and stored on the server. For searching a particular word, the user must generate the search request by computing the trapdoor of the word and sends it to the server. The server upon receiving the request performs tests to check if any Bloom filter holds the trapdoor of the query word and if so, it returns the corresponding file identifiers.

- Jun Zhou et.al, [10] proposed a more efficient verifiable outsourced computation of encrypted data EVOC from any one-way trapdoor function is proposed by combining a newly devised privacy-preserving data aggregation supporting both addition and multiplication operations with Yao's Garbled Circuit. The advantage is it proves the security of the proposed efficient privacy-preserving data aggregation scheme.

- Jin Li et al, [12] proposed a latest searching technique called fuzzy keyword search. They focused on enabling effective and privacy preserving fuzzy keyword search in Cloud Computing. To the best of knowledge, they formalize for the first time the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy.

## III. PROPOSED METHODLOGY

Mainly system has two functions. One of them is resource function and second one is customer function. Resource function deals with transactional values. The application function issues the query value and collects the information from the information source which is in encrypted standard.

The system is consists of four most important modules. They are data owner, data user, cloud Service provider and key generator as shown in figure 1.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

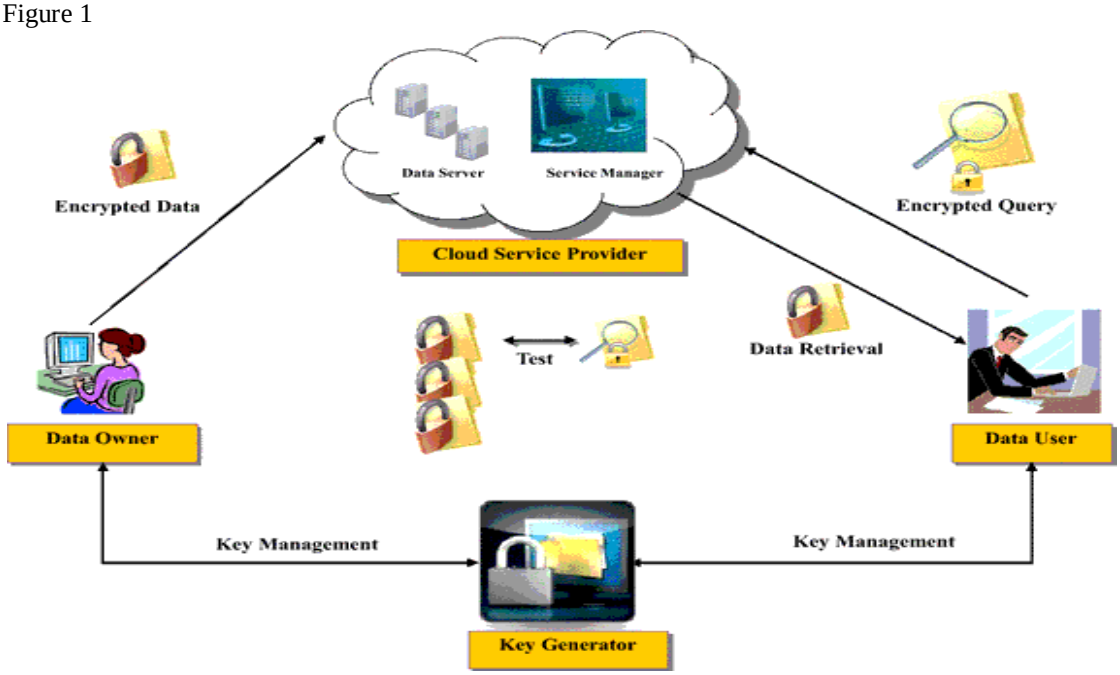
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Figure 1




Data owner: The main duty of the data owner to generate and encrypt the data and uploads them to the cloud server. It will be of form organization or in individual data user: There are many data users in the system who are the subscriber to the cloud storage. Data users acts as a entity which sends encrypted queries to the cloud service provider to search for specific encrypted data.

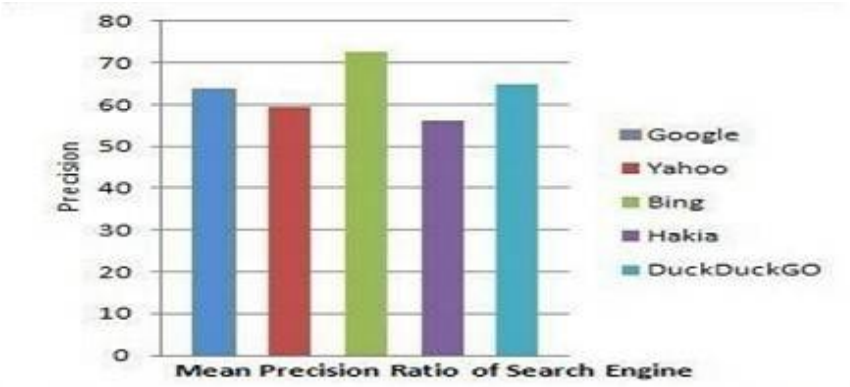
Cloud Service provider: It provides data storage and retrieval service to the data user. It mainly consists of cloud data server and cloud service manager.

Key generator: It responsible for user specific key generation and distribution during setup of the system

## IV. EXPERIMENTAL RESULT

Below figure2 shows overall graphical representation of mean precision ratio of search engine for first 20 document. Figure 2 and 3 represents graphical representation for queries.

Figure.2




Precision ratio of search engines for first 20 documents

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

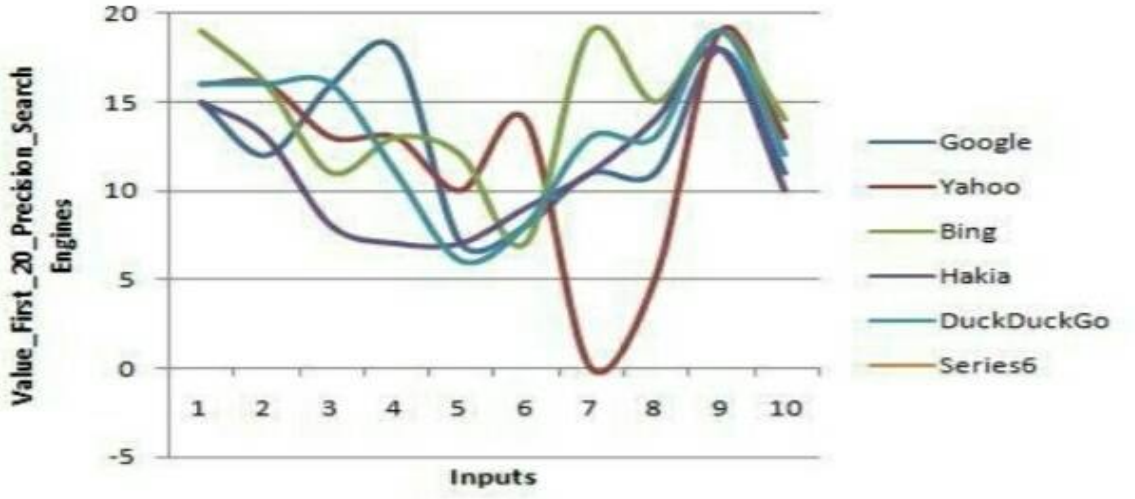
*(A High Impact Factor & UGC Approved Journal)*

*Website: www.ijirset.com*

**Vol. 6, Issue 8, August 2017**

It is concluded that figure 3 that semantic search engine like Bing and DuckDuckGo retrieve more efficient documents than keyword based search engine like Google and Yahoo. However, search performance of Hikia, which is a semantic search engine is lowest

Figure.3




Visual Representation of Search Engines for first 20 Precision

## V. CONCLUSION

This paper gives brief literature survey on single keyword based searching, multi-keyword based searching and many other searching techniques in cloud based environment. Outcome of this survey is the multi-keyword search technique sounds to be more efficient than other available searching technique. Many search schemes over encrypted data, supports multi-keyword query and similarity ranking simultaneously for data retrieval in cloud computing.

## REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10, 2010
[2] Deepali D. Rane and Dr.V.R.Ghorpade "Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2015.
[3] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.
[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.
[5] 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., NO.1,2015.
[6[Bing Wang, Wei Song, Wenjing Lou, and Y. Thomas Hou "Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee" IEEE Conference on Computer Communications (INFOCOM), 2015.
[7]Yanzhi Ren, Yingying Chen, Jie Yang, Bin Xie " Privacy-preserving Ranked Multi-Keyword Search Leveraging Polynomial Function in Cloud Computing" Globecom Communication and Information System Security Symposium 2014.
[8] Mikhail Strizhov and Indrajit Ray "Multi-keyword Similarity Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2012.
[9] E.-J. Goh, " Bloom filters in order to construct the indexes for the data files" IEEE Conference on Computer Communications 2016.
[10] Jun Zhou, Zhenfu Cao, Xiaolei Dong and Xiaodong Lin "More Efficient Verifiable Outsourced Computation from Any One-way Trapdoor Function" IEEE ICC - Communication and Information Systems Security Symposium, 2015.
[12] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.