CHANDIGARH GROUP OF COLLEGES
Campus : Jhanjeri, Mohali

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

Study of various Multi-keyword Text search technique over Encrypted Data in cloud computing

Hitesh [1]
Ankita Puri[2]

Abstract

The promising benefit of cloud computing is outsourcing of data service, by which the data owners stores their data in the public data centers by economically saving their capital investment towards data management. Cloud Storage provides users with abundant storage space and makes it user friendly for immediate acquiring of data, which is the foundation of all kinds of cloud applications. To maintain the privacy of personal documents stored in cloud environment, it should get encrypted before outsourcing to the cloud. After placing the data on the cloud, retrieving the same is also a quiet tedious job. In order to retrieve the data, several approaches are available in which keyword enabled search of the encrypted data is one of the outstanding techniques. The majority of these approaches are limited to handle a single keyword search with its own limitation. To enhance searching in terms of efficiency and fastness, a multi-key word search technique can be adopted to retrieve a corresponding document from cloud. This paper proposes a survey on a secure search scheme supporting single-keyword or multi-keyword ranked search over encrypted cloud data.

CloudCloud Computing, Multi Keyword Search, Clustering in Cloud

I. INTRODUCTION

As Mobile cloud computing [1] [4] gets rid of the hardware limitation of mobile devices by exploring the scalable and virtualized cloud storage and computing resources, and accordingly is able to provide much more powerful and scalable mobile services to users. In mobile cloud computing, mobile users typically outsource their data to external cloud servers, e.g., iCloud, to enjoy a stable, lowcost and scalable way for data storage and access. However, as outsourced data typically contain sensitive privacy information, such as personal photos, emails, etc., which would lead to severe con dentiality and privacy violations [5], if without efficient protections. It is therefore necessary to encrypt the sensitive data before outsourcing them to the cloud. The data encryption, however, would result in salient difculties when other users need to access interested data with search, due to the dif culties of search over encrypted data. This fundamental issue in mobile cloud computing accordingly motivates an extensive body of research in the recent years on the investigation of search-able encryption technique to achieve efficient searching over outsourced encrypted data. Furthermore, after that in the wake of tolerating the trapdoor, the cloud server executes. To address the stresses of security and addition the choice of dispersed stockpiling, we fight for sketching out a virtual private stockpiling organization in perspective of new cryptographic procedures. Such an organization should hope to fulfill the "best of both universes" by giving the security of a private cloud and the helpfulness and cost speculation assets of an open cloud. In this wander, we will deal with the issue of multi-catchphrase situated look for over encoded cloud data and recuperate the most vital records.

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

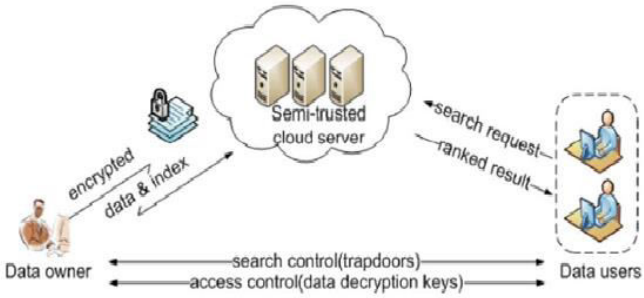

Fig 1: Privacy preserved Multi keyword Search[2]

II. MULTI-KEYWORD SEARCH

Now a day's cloud computing has become essential for many utilities, where cloud customers can slightly store their data into the cloud so as to benefit from on-demand high quality request and services from a shared pool of configurable computing resources. Its huge suppleness and financial savings are attracting both persons and enterprise to outsource their local complex data management system into the cloud. To safe guard data privacy and struggle unwanted accesses in the cloud and away from, sensitive data, for example, emails, personal health records, photo albums, videos, land documents, financial transactions, and so on, may have to be encrypted by data holder before outsourcing to the business public cloud; on the other hand, obsoletes the traditional data use service based on plaintext keyword search. The insignificant solution of downloading all the information and decrypting nearby is clearly impossible, due to the enormous amount of bandwidth cost in cloud scale systems. Furthermore, apart from eradicating the local storage management, storing data into the cloud supplies no purpose except they can be simply searched and operated. Thus, discovering privacy preserving and effective search service over encrypted cloud data is one of the supreme importance. In view of the potentially large number of on-demand data users and vast amount of outsourced data documents in the cloud, this difficulty is mostly demanding as it is really difficult to gather the requirements of performance, system usability, and scalability. On the one hand, to congregate the efficient data retrieval requirement, the huge amount of documents orders the cloud server to achieve result relevance ranking, as an alternative of returning undifferentiated results. Such ranked search system allows data users to discover the most appropriate information quickly, rather than burdensomely sorting during every match in the content group. Ranked search can also gracefully remove redundant network traffic by transferring the most relevant data, which is highly attractive in the "pay-as-you-use" cloud concept. For privacy protection, such ranking operation on the other hand, should not reveal any keyword to related information. To get better the search result exactness as well as to improve the user searching experience, it is also essential for such ranking system to support multiple keywords search, as single keyword search often give up far too common results. As a regular practice specifies by today's web search engines i,e Google search, data users may lean to offer a set of keywords as an alternative of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search demand is able to help narrow down the search result further. "Coordinate matching", as many matches as possible, is an efficient resemblance measure among such multi-keyword semantics to refine the result significance, and has been widely used in the plaintext information retrieval (IR) community. Though, the nature of applying encrypted cloud data search system remains a very demanding task in providing security and

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

maintaining privacy, like the data privacy, the index privacy, the keyword privacy, and many others. Encryption is a helpful method that treats encrypted data as documents and allows a user to securely search through a single keyword and get back documents of interest. On the other hand, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot put up such high service-level needs like system usability, user searching experience, and easy information discovery. Even though some modern plans have been proposed to carry Boolean keyword search as an effort to improve the search flexibility, they are still not sufficient to provide users with satisfactory result ranking functionality. The solution for this problem is to secure ranked search over encrypted data but only for queries consisting of a single keyword. The challenging issue here is how to propose an efficient encrypted data search method that supports multi-keyword semantics without privacy violation. In this paper, we describe and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving exact system wise privacy in the cloud computing concept. Along with various multikeyword semantics, select the efficient resemblance measure of "coordinate matching," it means that as various matches as possible, to confine the significance of data documents to the search query. Particularly, inner product similarity the numbers of query keywords show in a document, to quantitatively calculate such similarity assess of that document to the search query. For the period of the index construction, each document is associated with a binary vector as a sub-index where each bit signifies whether matching keyword is contained in the document. The search query is also illustrates as a binary vector where each bit means whether corresponding keyword appears in this search request, so the resemblance could be exactly calculated by the inner product of the query vector with the data vector. On the other hand, directly outsourcing the data vector or the query vector will go against the index privacy or the search privacy. To face the challenge of cooperating such multi keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is modified from a secure knearest neighbour (kNN) method, and then give two considerably improved MRSE method in a step-by-step way to accomplish different severe privacy needs in two risk models with enlarged attack competence

III. RELATED STUDY

Chi Chen et al. [1] developed the searchable encryption for multi-keyword ranked search over the storage data. Specifically, by taking the huge number of outsourced documents (data) on the cloud, algorithm utilize the k-nearest neighbor and relevance score techniques to design an efficient multi-keyword search scheme that can give back the ranked search results based on the accuracy. This system improve the search efficiency by supporting efficient index, and hide access pattern of the search user by adopt the blind storage system.

Keerthana G et al. [2]fabricated an application for enhancing cloud security utilizing partition and encryption technique which will enhance the cloud security. First of all record from client were taken and partition it into number of parts. After partition we encrypt the all record parts. At that point we send record parts to various cloud servers. At the point when client need that information back we take that information from cloud servers and decrypt that information. After decrypting, merging of that information is done and offer it to client. Our objective is that the application ought to have straight forward client interface for clients adaptability. The proficient method for giving security is the utilization of hybrid cryptography for more secured sending and receiving of information.

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

K.Ramadevi and L.Sunitha Rani [3]proposed schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors and systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel Additive Order and Privacy Preserving Function family. To prevent the attackers from eaves dropping secret keys and pretending to be legal data users submitting searches, we propose a novel dynamic secret key generation protocol and a new data user authentication protocol. Furthermore, PRMSM supports efficient data user revocation.

Khnd Sri Sandhya and K. Venkat Rao [4]scheme is based on multi-keyword ranked search which supports dynamic update operations. The data owner generates an exceptional tree-based catalog composition together with "Greedy Depth-first Search" criteria to make successful multi-keyword search. Achieving parallelism is the limitation of the existing system. The described technique extended the existing scheme with secure Dynamic Key generation along with the vector space model and it also includes TF_IDF model for index development as well for query generation. The dynamic key generation favors parallel search process by allowing multiple users to retrieve the same encrypted cloud data

Veerraju Gampala and Sreelatha Malempati [5]employed probabilistic public key encryption algorithm for encrypting the data and invoke ranked keyword search over the encrypted data to retrieve the files from the cloud. This technique aimed to achieve an efficient system for data encryption without sacrificing the privacy of data. Further, this ranked keyword search greatly improves the system usability by enabling ranking based on relevance score for search result, sends top most relevant files instead of sending all files back, and ensures the file retrieval accuracy. Thorough security and performance analysis, we prove that our approach is semantically secure and efficient.

Shreejit Pillai et al. [6]proposed schemes to deal with privacy preserving ranked multi keyword search in a multi owner environment to enable cloud servers to perform secure search without knowing the actual data of both keywords systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance course between keywords and files, and proposed a novel additive order and privacy preserving function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data owners submitting searches proposed a dynamic secret key generation key protocol and a new data user authentication protocol.

Shrilakshmi Prasad and B. S. Mamatha [7] defined and solve the problem of association attack by encrypting the index file using Paillier cryptographic algorithm. So cloud will have the challenge of searching the index file with the search query where both will be in an encrypted format. Hence privacy of the document will be preserved. Cosine similarity search is used to retrieve the top matching documents based on their relevance score and the beauty of the proposed system is the user can give multiple keywords in their search query. The enterprises are interested in storing their data in the public cloud. Before uploading the data on to the cloud, it needs to be encrypted to preserve privacy. In order to ease searching, the index file should be built for each document. The index file contains the keyword and its count in the particular document. The unencrypted index file leads to association attack since with the keywords and their count, the content of the document can be known.

Seema Ranga and Ajay Jangra [8]aimed of Intrusion detection System is to defend the security of the Computer system by a layer over the defense system. IDS systems sense the misuse, breach in the security system and also the malicious or unauthorized access to the system. Although Firewalls works

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

for the same reason but the major difference between firewalls and the IDS is IDS suspect the source of the attack and signals the alarm to the system but a firewall directly stops the communication without informing the system. These attacks requires true concerns as they harm the data stored in system and also effect the network traffic, data packet etc.

Wei Zhang et al. [9]proposed scheme to deal with privacy preserving ranked multi-keyword search in a multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, algo systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files and proposed a novel additive order and privacy preserving function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data users submitting searches, this approachdeveloped a novel dynamic secret key generation protocol and a new data user authentication protocol.

Izzat Alsmadi and Ikdam Alhami [10] defined that information users depend heavily on emails' system as one of the major sources of communication. Its importance and usage are continuously growing despite the evolution of mobile applications, social networks, etc. Emails are used on both the personal and professional levels. They can be considered as official documents in communication among users. Emails' data mining and analysis can be conducted for several purposes such as: Spam detection and classification, subject classification, etc. In this paper, a large set of personal emails is used for the purpose of folder and subject classifications. Algorithms are developed to perform clustering and classification for this large text collection. Classification based on NGram is shown to be the best for such large text collection especially as text is Bi-language

IV. COMPARATIVE STUDY FOR DATA SEARCHING AND ENCRYPTION TECHNIQUES[12]

Searchable Encryption: It allows users to securely search complete encrypted data through keywords. This method support only Boolean search, without capturing any relevant data. This approach suffers from two main drawbacks when directly applied in the context of Cloud Computing. First one, users who do not necessarily have pre-knowledge of the encrypted cloud data, have to post process every file got, in order, to find ones most matching their interest; another drawback, regularly getting all files containing the queried keyword further incurs unnecessary network traffic, when retrieve more than one files.

Single Keyword Searchable Encryption: A single keyword searchable encryption schemes usually builds an encrypted searchable index such that, it's content is hidden to the server, unless it is given appropriate trapdoors generated via secret key(s). Early work solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search. Where anyone with public key can write to the data stored on server, but only authorized users with private key can search. Traditional single keyword searchable encryption schemes are usually built in a way by creating an encrypted searchable index. Such indexes content will be hidden to the server. The information will be revealed only when the server gives the correct trapdoors that are generated via a secret key(s). The main drawback of single keyword-based search is that it is not comfortable enough to express complex information needs.

Ranked Keyword Search: Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (eg. keyword frequency) thus, making one

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

step closer toward practical deployment of privacy-preserving data hosting services in the context of cloud computing. To the best of knowledge it gives a legal status for the first time the problem of effective ranked keyword search over encrypted cloud data. Ranked keyword search strongly provides system usability by returning the matching files in ranked order concerning to certain relevance criteria, thus moving close towards the practical action of privacy preserving data presenting services in cloud computing. To achieve design goals investigate the statistical measure approach from Information retrieval (IR) and text removal to insert relevance score of each file during the establishment of searchable index before outsourcing the encrypted file collection. An IR system allocates a relevance score to each and every document and ranks those documents by this score. Relevance score is used to build a secure searchable index to properly protect the sensitive information. This technique enables data users to find the most related information rapidly, rather than burdensome sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. For privacy protection, such ranking function, however, should not leak any keyword relevant information. Another one, to improve search result accuracy as well as enhance user searching experience, it is also essential for such ranking system to support multiple keywords search.

Multi-keyword Searchable encryption: Over the years, various searchable encryption approaches have been developed to provide the ability for selectively retrieving the encrypted documents through a keyword search. Typically, these systems build a secure index structure and outsource it along with the encrypted documents to the remote server. Authorized users submit their requests as secret trapdoors that are integrated properly with the stored indexing information. The server uses the received trapdoor to search over the stored index, and retrieves the matching encrypted documents. However, the previous searchable encryption schemes are impractical for real world cloud computing scenarios because these systems are designed to handle either a single keyword search or a Boolean search.

Fuzzy Keyword Searchable Encryption: Fuzzy keyword search greatly enhances system usability by returning the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. More specifically, it uses edit distance to quantify keywords similarity and develop a novel technique that is a wildcard-based technique, for the construction of fuzzy keyword sets. This technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzy keyword sets is significantly reduced.

Plaintext Fuzzy Keyword Search: The importance of fuzzy search has received attention in the context of plaintext searching in information retrieval community. The problem is addressed in the traditional information access paradigm by allowing user to search without using try-and-see approach for finding relevant information based on approximate string matching. At the first glance, it seems possible for one to directly apply these string matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within an alphabet. This trivial construction suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

Boolean Keyword Search: Boolean systems allowed customers to specify their information need using a combination of Boolean operators AND, OR and NOT. Boolean systems have several disadvantages, for example there are no any features of document ranking, and it is very difficult for a customer to make a good search request. Thus, the drawback of existing system specifies the important need for new techniques that support searching flexibility.

International journal Of Business Management
ISSN NO. 2349-3402
Available at www.ijbm.co.in
VOL. 3 Issue 1, 2017

V. CONCLUSION

This paper concentrated on various information seek strategies crosswise over cloud servers. There are many techniques which attempt to minimize the searching time and providing the best accurate results. Finally we have discussed about EMRS technique to enable accurate, efficient and secure search over encrypted mobile cloud data. which outputs the accurate result . We have categorized the papers and there is a need to compare techniques in each category to understand the strengths and weaknesses