Novel Indexed based approach for Searching over encrypted cloud data

Richa Singh Patel
M.Tech student, Takshshila Institute of Engineering & Technology,
Jabalpur (M.P.) [India]
Email: richa_ssp@yahoo.com

Chandni Nagvanshi
Asst. Professor, Takshshila Institute of Engineering & Technology,
Jabalpur (M.P.) [India]

Abstract — As the quality of cloud computing is increasing; it has revamped the read of contemporary info technology which is motivating the information owners to source their data to the general public cloud server like Amazon Drive, Microsoft Azure, Google Drive, etc.

With the appearance of cloud computing, knowledge house owners are driven to source their advanced knowledge management systems from native sites to the industrial public cloud for excellent flexibility and economic savings. Except for protective knowledge privacy, sensitive knowledge got to be encrypted be-fore outsourcing, that obsoletes ancient knowledge utilization supported plaintext keyword search. Thus, facultative an encrypted cloud knowledge search service is of overriding importance. Considering the big variety of information users and documents within the cloud, it's necessary to permit multiple keywords within the search request and come documents within the order of their connection to those keywords (MRSE – Multiple Keyword graded Search).

Existing techniques specialise in single keyword search and infrequently type the search results. In our projected mechanism, we tend to outline and specialise in the matter of graded search over encrypted knowledge. We tend to capture the connection of information documents to the search question. Graded/Rank search also can elegantly eliminate spare network traffic by causing back solely the foremost relevant knowledge that is very fascinating. Hence, exploring privacy-preserving and effective search service over encrypted knowledge is of nice importance.

Index Terms— Cloud computing, privacy-preserving, keyword search, ranked search

1 INTRODUCTION

Cloud Computers that do work for you, but that are stored somewhere else and maintained by other compnies known as Cloud Computing which is the long dreamed vision of calculating (computing) as a utility. Cloud customers remotely store their data into the cloud to enjoy the on-demand high-quality applications and services from a shared pool of configurable figuring out/calculating useful things supplies. Its great flexibility and money-based savings are giving a reason to do something for both people and businesses to pay someone else to do something for their local complex data management system. To protect privacy of data and without being requested accesses in the cloud and beyond it, sensitive data, for instance, e-mails, personal health records, collections of songs/books for inserting pictures, tax documents, and so on, may have to be turned into secret code by data owners before paying someone else to do something to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The unimportant solution of downloading all the data and changing secret codes into readable messages locally is clearly not having way too full of problems, due to the large amount of radio frequency cost. More than that, aside from eliminating the local storage management, storing data into the paid someone else to do something storage doesn't serve any purpose unless they can be easily searched and used..

Traditional searching methods provide Boolean search to search over unreadable data, which is not related when the number of users and the number of data files stored in the cloud is large. These also cause an inconvenient situation involving two major issues, one being the after-processing that has to be done by the users to find the relevant document in need and the other is the network traffic that is undesirable in present situation when all the files matching with keywords is retrieved. But this paper proposes Ranked keyword search that overcomes these issues.

There are also other techniques like Ranked search which can elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable. Hence, exploring privacy-preserving and effective search service over encrypted data is of great importance. To enhance the search result, we need efficient methods to perform similarity search over large amount of encrypted data. Existing methods are a lot used for fast search of thing that's almost the same as compared to another on plain data in information retrieval community. In our scheme, we propose to use it in the big picture of the secret/unreadable data.

2 Related Study

Even though there are various systems existing, this literature survey mainly concentrates on the single keyword based encryption and multi-keyword based encryption and also included other searching techniques due to it known advantages.

2.1 Searching Techniques

Search over secret data is a way of doing things of great interest in the Cloud Computing time in history, because many believe that sensitive data has to be turned into secret code before paying someone else to do something to the cloud servers in order to secure user data privacy. In the survey many privacy preserving multi-keyword search needed things are defined in the cloud. The privacy preserving needed things are as follows:

2.1.1 Multi-keyword text search similarity based ranking schemes

As popularity of Cloud in increasing, to avail benefits like reduced management cost and ease of accessing huge amount of documents are put on to the cloud. Although turning data into secret code helps in protecting user data, it leaves the well-functioning yet practically secure search functions over secret data a challenging problem. This paper presents a privacy-preserving multi-keyword text search (MTS) scheme with Similarity-based ranking to deal with this problem. Keyword searching looks for words anywhere in the record. Searching by keyword are a good alternative for a subject search in absence of approved subject heading form. Keyword may also be used as an alternative for a title or author search when you have incomplete title or author information.

2.1.2 Public key encryption with keyword search

Many businesses are moving their valuable data to the cloud due the advantage of storage as a service, since it can be accessed from anywhere any time, costs less. The trust between cloud user and provider is most important. We use security as a limit/guideline to establish trust. The science of making secret codes is one way of beginning and building on trust. Searchable cryptography is a method to provide security. In books lot of work is being done by researchers on developing efficient searchable schemes.

The problem of searching on data that is secret using a public key system is studied here. Let us consider an example where user A sends email to user Data Owner Encrypted under Data Owner's public key. An email gateway wants to test whether the email contains the keyword "extremely important" so that it could route the email in the same way. Data Owner, on the other hand does not wish to give the gateway the ability to change secret codes into readable messages. This defines and constructs a way that permits Owner of Data to calculate a key to the entrance that permits the entrance to check whether or not the word "extremely important" could be a keyword within the email while not learning anything concerning the e-mail. The paper refers to the way as Public Key encryption with keyword Search. As another example, think about a mail server that stores different messages publicly turned into secret code for Data Owner by others. Using the method Data Owner can send the mail server a key that will enable the server to identify all messages containing some clearly stated keyword, but learn nothing else. The paper defines the idea of public key Encryption with keyword search and gives a lot of constructions.

The general term used for a spread of mechanisms that share the principle of looking (typically, terribly large) areas of objects is Similarity search wherever the sole obtainable comparator is that the similarity between any combine of objects. This is often changing into more {and more} necessary in an age of enormous data repositories wherever the objects contained don't possess any universe, as an example giant collections of pictures, sounds and different subtle digital objects. Nearest neighbor search and vary queries are necessary subclasses of similarity search, and variety of solutions exist. Analysis in Similarity Search is dominated by the inherent issues of looking over advanced objects. Sadly, in several cases wherever similarity search is critical, the objects are inherently advanced. The foremost general approach to similarity search that permits construction of economical index structures use the mathematical notion of topological space.

2.1.3 Approximate Keyword-based Search over Encrypted Cloud Data

To protect the privacy, users ought to encipher their sensitive knowledge before outsourcing it to the cloud. However, the standard cryptography schemes square measure inadequate since they create the appliance of categorization and looking out operations tougher tasks. Consequently, searchable cryptography systems square measure developed to conduct search operations over a group of encrypted knowledge. These systems unfortunately enable their purchasers to perform a precise search however not search approximate, a very important would like for all the present info retrieval systems. Associate exaggerated attention has been paid recently to the approximate searchable secret writing systems to seek out keywords that match the submitted queries just about. This work focuses on constructing a versatile secure index that permits the cloud server to perform the approximate search operations while not revealing the content of the question trapdoor or the index content. A Keyword looking out appearance for words anyplace within the record. Keyword searches square measure a decent substitute for a subject matter search after you don't apprehend the approved subject heading kind. Keyword might also be used as a substitute for a title or author search after you have incomplete title or author info. You will conjointly use the Guided Keyword search choice to mix search parts, cluster terms, or choose indexes or fields to be searched.

2.1.4 Secure Conjunctive Keyword Search Over Encrypted Data

The security model for conjunctive keyword search over encrypted information is the outcome of some analysis and gifts the primary schemes for conducting such searches firmly. During this proposal 1st a theme that the communication price is linear within the range of documents, however that price will be incurred “offline” before the conjunctive question is asked. The safety of this theme depends on the Decisional Diffie-Hellman (DDH) assumption. A second theme whose communication price is on the order of the quantity of keyword fields and whose security depends on a brand new hardness assumption.

There are measure that give conjunctive keyword queries on encrypted information. Though such conjunctive searches definitely don't comprehend all attainable search criteria, we have a tendency to believe that they're a vital building block as indicated by the reliance of today’s net search engines on conjunctive search. To inspire the matter of conjunctive search additional, and illustrate the difficulties it raises.

2.1.5 Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data

Although ancient searchable secret writing techniques enable users to firmly search over encrypted information through keywords, they support solely mathematician search and aren't nevertheless sufficient to fulfill the effective information utilization would like that's inherently demanded by quantity} of users and large amount of knowledge files in cloud. During this paper, we have a tendency to outline and solve the matter of secure hierarchic keyword search over encrypted cloud information. Hierarchic search greatly enhances system usability by sanctioning search result connectedness ranking rather than causing undifferentiated results, and any ensures the file retrieval accuracy. To explore the applied math live approach, i.e. connectedness score is a general tendency, from developing a one-to-many order-preserving mapping technique to properly defend those sensitive score info and information retrieval to make a secure searchable index. Thorough analysis shows that our projected answer enjoys “as-strong-as possible” security guarantee compared to previous searchable secret writing schemes, whereas properly realizing the goal of hierarchic keyword search. Intensive experimental results demonstrate the potency of the projected answer.

2.1.6 Similarity Based Search Schemes

Zhihua Xia et.al,[5] proposed a secure, efficient and dynamic search scheme, which 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. In addition, the parallel search process can be carried out to further reduce the time cost. The security of the scheme is protected against two threat models by using the secure KNN algorithm. Experimental results demonstrate the efficiency of proposed scheme. 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 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 k-nearest neighbor queries (kNN) were studied in an outsourced setting. These techniques are based on a notion of distance between data points and the query point, and approximation of such distances. The comparison operation required by range query search does not have a natural notion of distance or approximation.

2.2 Survey of searchable encryption

In the literature, searchable coding may be a useful technique that treats encrypted information as documents and permits a user to firmly search through one keyword and retrieve documents of interest. However, direct application of those approaches to the secure massive scale cloud information utilization system wouldn't be essentially appropriate, as they're developed as crypto primitives and can't accommodate such high service-level needs like system usability, user looking expertise, and straightforward data discovery. though some recent styles are planned to support Boolean keyword search as an endeavor to counterpoint the search flexibility, they're still not up to give users with acceptable result ranking practicality .Our early works are conscious of this drawback, and supply solutions to the secure graded search over encrypted information drawback however just for queries consisting of one keyword. a way to style associate degree economical encrypted information search mechanism that supports multi-keyword search while not privacy breaches still remains a difficult open drawback.

The proposed mechanism focuses on using uploaded files in form of clusters or in groups. Similarity of keywords is analyzed to identify similar set of files in groups. Keywords are retrieved from files for identifying indexes for each group. Search mechanism is dependent on index values generated during processing. The indexes are identified and files similar to searched keywords are identified from selected clusters of similar indexes. The ranking process displays the final output in organized format based on relation of searched keywords with corresponding files.

2.2.1 Overview

Searchable Searchable cryptography has been an energetic analysis space and lots of quality works. ancient searchable cryptography schemes sometimes build AN encrypted searchable index such its content is hidden to the server; but it still permits activity document looking with given search question. the primary to analyze the techniques for keyword search over encrypted and outsourced information. The authors begin with plan to store a group of plaintext documents on information storage server like mail servers and file servers in encrypted kind to scale back security and privacy risks. The work presents a scientific discipline theme that permits indexed search on encrypted information while not unseaworthy any sensitive data to the untrusted remote server.

The current mechanism focuses on analyzing input file in sort of clusters for segregation in separate teams. Similarity of keywords is analyzed to spot similar set of files in teams. Keywords also are analyzed for distinctive index values for every cluster. Search mechanism is associated to index values generated throughout process. The index values area unit known and files the same as searched keywords area unit known from hand-picked clusters victimization assortment. The ranking method arranges the ultimate output supported relation of searched keywords with corresponding files.

3 Proposed Model

We have planned associate in nursing economical theme that permits the Cloud Service supplier (CSP) to see the files that area unit associated with the keywords searched by the user, rank them and send the foremost relevant files while not knowing any info concerning the cloud. This planned technique has outlined and resolved the matter of effective however safe and sound rank keyword search over Encrypted cloud information. Hierarchical search greatly enhances system usability by returning the matching files in an exceedingly hierarchical order relating to sure vital criteria (e.g. keyword frequency) therefore creating one step nearer towards wise consumption of secure information hosting services in Cloud Computing. These papers has outlined and resolved the difficult downside of privacy protective and economical multi keyword hierarchical search over encrypted cloud information storage (MRSE), and establish a group of strict privacy.

Advantages of Proposed System

1) Proposed schemes indeed introduce low overhead on computation and communication.
2) It uses ranked search mechanism to support more search semantics and dynamic data operations.
3) It is more secure and efficient.

System model

To enable efficient similarity search, data owner builds a secure index and outsources it to the cloud server along with the encrypted data items. Server performs search on the index according to the queries of the data users without learning anything about the data other than what data owner allows an adversary to learn.

Data Owner's Modules:

1. Login Module
2. Data Upload
3. View Files(Search)

Data Users Modules:

4. Login/Registration
5. View Files(Search)
6. Download Files

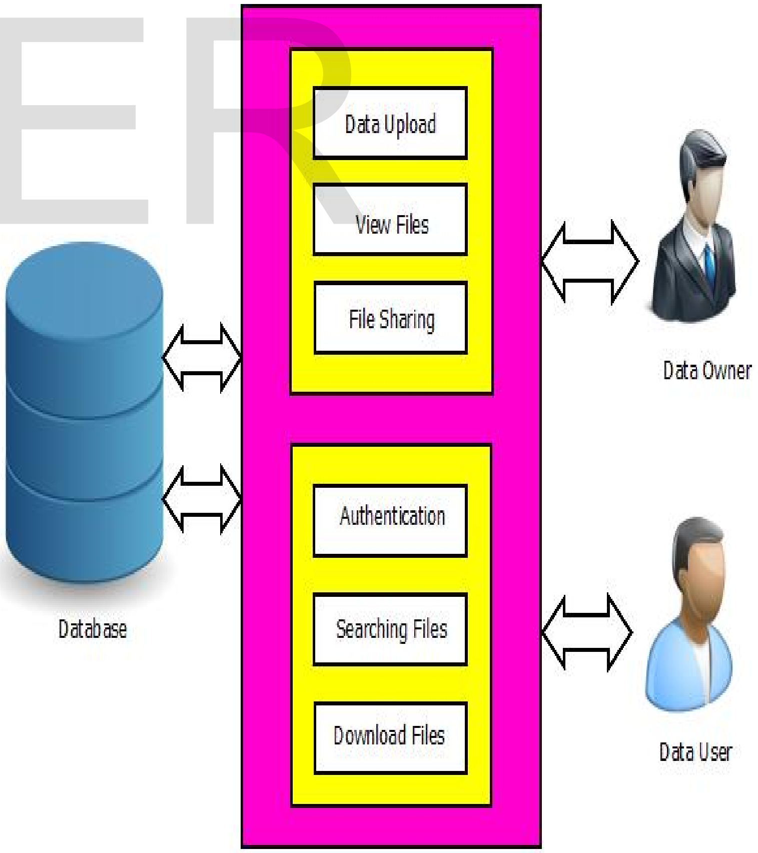

Figure 1: System Model

Implementation:

This section will show the implementation of proposed mechanisms in which we will going to implement Secure Search Algorithm to provide privacy in multi keyword search over encrypted cloud data using AES (Advanced Encryption Standard) Algorithm. Secure Search Algorithm is a method which is used for determining which items in a given set are similar. Rather than using the approach of comparing all pairs of items within a set, items are grouped into buckets, such that similar items will be more likely to hash into the same buckets. As a result, the number of comparisons needed will be reduced; only the items within any one bucket will be compared. Indexing is often used when there exists an extremely large amount of data items that must be identified. In these cases, it may also be that the data items themselves will be too large, and as such will have their dimensionality reduced by a feature extraction technique beforehand.

The main application of proposed system is to provide a method for efficient search through indexing of data into clusters. This process is done through keyword extraction and indexing. The general idea is to index clusters in such a way that similar documents are more likely to be grouped into the same bucket. It focuses on analyzing files in form of clusters for arranging in separate groups. Similarity of keywords is analyzed to identify similar group of files. Keywords are also processed for identifying indexes for each cluster. Search mechanism is based to indexes generated during processing. The indexes are identified and files similar to searched keywords are selected from identified clusters using indexing. The ranking process organizes the final output based on relation of searched keywords with corresponding files.

4 CONCLUSION

The cloud computing paradigm has recently revolutionized the organization's approach of operational their information, significantly within the approach they store, access, and method information. As Associate in Nursing rising computing paradigm, cloud computing attracts several organizations to think about a cloud's potential in terms of its cost-efficiency, flexibility, and offload of body overhead. Organizations usually delegate their machine operations, additionally to their information, to a cloud; otherwise, there would be no purpose in outsourcing the info at the primary place. Privacy and security problems within the cloud are preventing firms from utilizing those benefits despite the tremendous benefits that the cloud offers. Therefore, owing to the increase of assorted privacy problems, sensitive information ought to be encrypted before being outsourced to the cloud. Exploitation secret writing as how to realize information confidentiality could cause another issue at the cloud throughout the question analysis. In general, it's terribly tough to method encrypted information during a privacy-preserving manner while not ever having to rewrite it. The question here is however the cloud will perform computations over encrypted information whereas the info keep within the cloud are encrypted in the least times. Together with this direction, this study proposes Associate in nursing economical

5 ACKNOWLEDGEMENTS

We would like to acknowledge the Management and Executive Director at TAKSHILA INSTITUTE OF ENGINEERING & TECHNOLOGY, JABALPUR (M.P.) for their continuous encouragement, support and providing the necessary resources that made this work possible.

REFERENCES

[1] J.S. Ning Cao, Cong Wang, Ming Li, KuiRen, and Wenjing Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS VOL:25 NO:1 YEAR 2014

[2] D. X. Song, D.Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In Proceedings of the 2000 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 2000.

[3] D. Boneh and B. Waters. Conjunctive, subset, and range queries on encrypted data. In Proceedings of the 4th IACR Theory of Cryptography Conference, Amsterdam, The Netherlands, Feb. 2007.

[4] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou. Privacy-preserving multi-keyword ranked search over encrypted cloud data. In Proceedings of the 30th IEEE International Conference on Computer Communications, Shanghai, China, Apr. 2011.

[5] C. Wang, N. Cao, K. Ren, and W. Lou. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Transactions on Parallel and Systems, 23(8):1467–1479, 2012.

[6] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS '10), 2010.

[7] Ru ihui Zhao, Hongwei Li, Yi Yang, Yu Liang, "Privacy-preserving Personalized Search over Encrypted Cloud Data Supporting Multi-keyword Ranking", 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)