The Multi Keyword Search over the Encrypted Data in Cloud Storage

[1]Sindhuja Somisetty, [2]Dasari Ravi Kumar
[1,2]Dept. of Computer Science & Engineering, QIS Institute of Technology, Ongole, AP, India

Abstract

The significance of keywords will change a great dispense of exact results, and in this manner the inclination variables of keywords state to the significance of keywords inside the inquiry keyword set nominative via search users and correspondingly allows customized hunt to oblige particular client inclinations. Individual will remotely store her insight on the cloud server, particularly information outsourcing, thus make the cloud computing open for free through the cloud server. It contain delicate protection data, they're typically encrypted before transferred to the cloud. Be that as it may, significantly restrains the ease of use of outsourced information on account of the issue of watching out over the encrypted computing. Amid this paper, we tend to address this issue by building up the fine-grained multi-keyword search conspires over encrypted cloud information. The cloud server then uses the figure to coordinate the outsourced encrypted keywords, and in conclusion gives back the coordinating results to the pursuit client. to achieve the comparable inquiry intensity and exactitude over encrypted information as that of plaintext catchphrase seek, a top to bottom collection of examination has been produced in writing. an investigation client questions the outsourced reports from the cloud server with taking after 3 stages. To start with, the pursuit client gets each the key and symmetrical key from the data proprietor. Second, in accordance with the hunt keywords, the pursuit client utilizes the key to think of trapdoor and sends it to the cloud server. Last, she gets the coordinating record arrangement from the cloud server and unscrambles them with the symmetrical key.

Keywords

Cloud Computing, Encryption, Multi-Keyword Search, Coordinate Matching, Searchable Encryption, Keyword Search Ranked.

I. Introduction

In Cloud computing where cloud users can remotely store their information into the cloud in order to appreciate the on-request amazing applications and administrations from a mutual pool of configurable figuring assets. Its extraordinary adaptability and financial reserve funds are persuading both people and undertakings to outsource their neighbourhood complex information administration framework into the cloud. To secure information protection and battle spontaneous gets to in the cloud and past, touchy information, for instance, messages, individual wellbeing records, photograph collections, impose reports, budgetary exchanges, et cetera, may must be encrypted by information proprietors before outsourcing to the business open cloud; this, in any case, obsoletes the customary information usage benefit in view of plaintext catchphrase search. The paltry arrangement of downloading every one of the information and decoding locally is plainly unreasonable, because of the immense measure of data transmission cost in cloud scale frameworks. In addition, besides dispensing with the nearby stockpiling administration, putting away information into the cloud fills no need unless they can be effectively searched and used. Hence, investigating security saving and compelling inquiry benefit over encrypted cloud information is of vital significance. Considering the possibly substantial number of on-request information users and enormous measure of outsourced information reports in the cloud, this issue is especially testing as it is amazingly hard to meet likewise the prerequisites of execution, framework ease of use, and versatility. From one viewpoint, to meet the viable information recovery require, the extensive measure of records request the cloud server to perform result importance positioning, rather than returning undifferentiated results. Such positioned seek framework empowers information users to locate the most pertinent data rapidly, instead of burdensomely dealing with each match in the substance accumulation. Positioned hunt can likewise carefully dispose of superfluous system movement by sending back just the most pertinent information, which is exceptionally alluring in the "pay-as-you-utilize" cloud worldview. For security assurance, such positioning operation, be that as it may, ought not release any keyword related data. Portable cloud computing can adequately address the asset constraints of cell phones, and is along these lines basic to empower broad asset devouring versatile registering and correspondence applications. Of all the portable cloud computing applications, information outsourcing, for example, iCloud, is essential, which outsources a versatile client's information to outside cloud servers and as needs be gives an adaptable and "dependably on" approach for open information get to. With the security and protection issues identified with outsourced information turning into a rising concern, encryption on outsourced information is frequently necessary [2]. With the coming of cloud computing, information proprietors are roused to outsource their unpredictable information administration frameworks from neighbourhood locales to the business open cloud for extraordinary adaptability and financial funds. Be that as it may, for securing information protection, touchy information must be encrypted before outsourcing, which obsoletes conventional information use in light of plaintext keyword seek. In this way, empowering an encrypted cloud information search administration is of vital significance. Considering the vast number of information users and records in the cloud, it is important to permit various keywords in the pursuit demand and return archives in the request of their significance to these keywords. Related chips away at searchable encryption concentrate on single keyword inquiry or Boolean catchphrase search, and seldom sort the list items [3].

II. Related work

Stateful Anonymous Credentials: "Controlling Access to relate Oblivious information exploitation Crateful Anonymous Credentials" anticipated that, n this work, we have a tendency to ponder the undertaking of allowing a substance provider to authorize propelled get to administration arrangements on neglectful conventions led with unknown users. As our essential application, we tend to demonstrate the best approach to build security safeguarding databases by joining unmindful exchange with partner expanded unknown archive framework. this permits a data administrator to farthest point that things each client may get to, while not computing something concerning users' personalities or thing choices. This durable security ensure holds even once users are doled out totally unique get to administration strategies and are permitted to adaptively construct a few questions. To do as such, we tend to demonstrate the best approach to expand existing unknown record frameworks all together that, moreover to affirming a client's characteristics, they conjointly store state concerning the client's data get to history. Our development underpins an extensive change of get to administration approaches, together with practical and individual acknowledge of the Brewer-Nash (Chinese Wall) and Bell-LaPadula (Multilevel Security) arrangements, that are utilized for money related and barrier applications. Moreover, our framework is predicated on standard suppositions inside the standard model and, when partner starting setup part, every managing needs exclusively steady time. Security expanded Access Control: Outsourced information Sharing" anticipated that, antiquated get to administration models typically accept that the element actualizing access administration strategies is moreover the proprietor of computing and assets. This supposition not holds once information is outsourced to an outsider stockpiling provider, similar to the cloud. Existing access administration arrangements primarily work in defensive secrecy of keep information from unapproved get to and hence the capacity provider. In any case, amid this setting, get to administration arrangements still as users' get to designs conjointly move toward becoming protection delicate data that should be protected from the cloud. we have a tendency to propose a two-level get to administration topic that blends coarse-grained get to administration upheld at the cloud, that licenses to ask satisfactory correspondence overhead and at a comparative due dates the information that the cloud gains from his fractional read of the get to rules and in this manner the get to examples, and fine-grained logical teach get to administration authorized at the client's angle, that gives the predefined nature of the get to administration arrangements. Our answers handles each output and compose get to administration.

III. Problem Statement

Computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud. Single-keyword search without ranking, Boolean-keyword search without ranking Single-keyword search with ranking

IV. Trapdoor Generation

Users register their identity tokens so as to get secrets to rewrite the info that they're allowed to access. Users register solely those identity tokens associated with the Owner's sub ACPs and register the remaining identity tokens with the cloud in a very privacy conserving manner. It ought to be noted that the cloud doesn't learn the identity attributes of Users throughout this part. Once Users register with the Owner, the Owner problems them 2 set of secrets for the attribute conditions in command that are gift within the sub ACPs in ACPB cloud. The Owner keeps one set and offers the opposite set to the cloud. 2 totally different sets are employed in order to forestall the cloud from decrypting the Owner encrypted knowledge.

V. Ranked Keyword Searching

As cloud computing has become an integral part of IT industry, data owners share their outsourced data. Due to these vast amounts of information available on WWW, large number of users attempts to retrieve certain specific data files they are interested in. One of the most popular ways to do so is through keyword based search. Keyword searches are done to utilize cloud data for a certain query. Such keyword search techniques allow users to selectively retrieve files of interest and have been widely applied in plain text search scenarios (C.wang). Great efforts have been made for facilitating users via keywords search. However, there are few researchers about entertaining the exact user query and presenting a ranked URL list according to it. Keywords searchers are typically done in such a way that users can utilize clouds to query a collection (7). To eliminate unnecessarily network traffic by not sending back the irrelevant data, ranked keyword search is used. This technique is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation should not leak any keyword related information. To improve the search result accuracy as well as to enhance the user searching experience, it is necessary for such ranking system to support multi-keyword search, as single keyword search often yields far too coarse results (5). The information is retrieved from the matching files to calculate the relevance scores of given request. If ranking system supports multiple keyword search then, it is possible to improve the search result accuracy as well as user searching experience can be enhanced. In all web search engines, users provide a set of keywords instead of only one keyword to indicate that they are interested in a particular area. Each keyword in the user query is used to narrow down the search process.

VI. Proposed System

There are three main actors present in these activities: cloud server, data owner, and data user. Data owner have her own sets of documents, to maintain these documents locally is become difficult task. Maintain and stored the documents locally are expensive for storage and it arises computational overhead. Hence data owner motivate to outsource their sets of documents on cloud to get more flexibility.

But before migration process, the data privacy issue is arises in front of owner, hence to maintain the security and privacy she used encryption methods and outsource the data in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. Information leakage would affect the data privacy which is unacceptable to data owner. The data user is sanctioned to process multi keyword retrieval over the outsourced data. The data user encrypts the query and sends it to the cloud server that returns the pertinent files to the data user. Afterward, the data user can decrypt and make use of the files.

A. Vector Space Model

It is used for accurate ranking. TF-IDF rule is used to find the accurate ranking and similarity measures. Where TF denotes occurrence count of term within a document and IDF is obtained by dividing the total number of documents in collection by number of document containing the term. It gives the top-k retrieval result. IDF =total number of documents in collection/ number of documents containing the term.

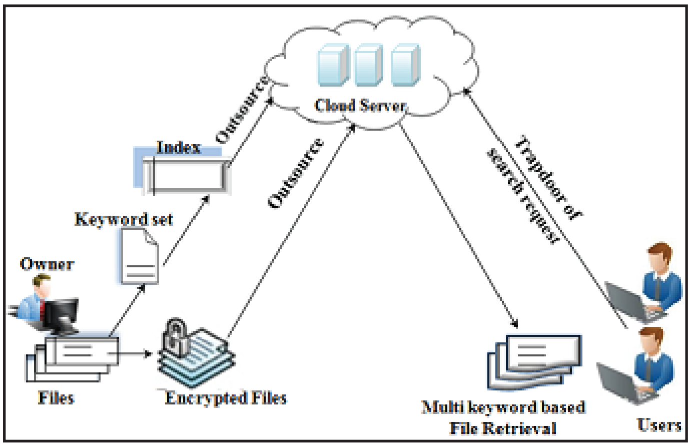

Fig. 1: Architecture of Proposed System

B. Enhance Secure Index Scheme

To achieve accurate multi-keyword ranked search, we adopt the cosine measure to evaluate similarity scores. In particular, we divide the original long document index vector into multiple sub index vectors such that each sub index represent subset of keyword and becomes a part of ith level of index tree as shown in proposed system. The query vector is divided in same way as document index vector. The final similarity score for document 'd' can be obtain by summing up the score of each level. Based on these similarity score, the cloud server determine the relevance document d to query Q and send top most relevant document to user. By using level wise secure inner product scheme, the document index vector and query index vector are both well protected.

C. MD Algorithm

MD algorithm is used to find k-best match in database that is structure as MDB-tree. MDB tree represents by attribute domain and each attribute in that domain has attribute value.

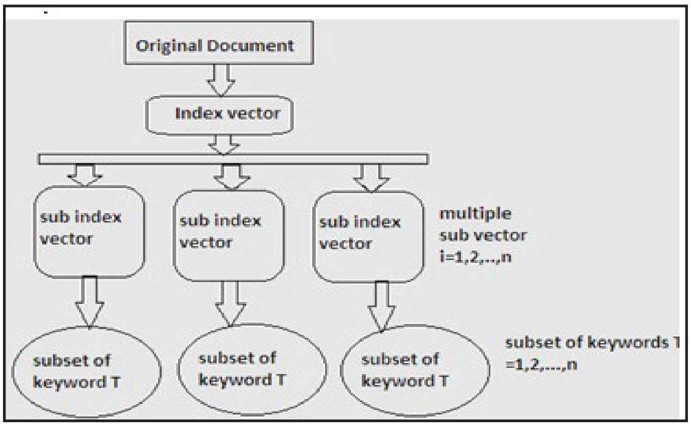

Fig. 2: Mechanism of Document Index Formation

D. Check File Status

Proposed system announces a third party auditor to audit user file request for checking integrity of corresponding file. Audit result from third party would be helpful for cloud service provider to enhance cloud based service platform.

Proposed Algorithms

1. Algorithm for Top Result selection:

```
    (i). Input
Take variable 'k' like a number and list source of selected item
    (ii). Initialization:
Set pointers tk&tid as a null
    (iii). Iteration phase
        (a). For all i ϵ source do Insert(tk,( i, index))
        (b). End for
        (c). For all tuple ϵ tk do tid.append(tuple[1])
        (d). End for

    (iv) Output:
    tid
```

2. Algorithm for Insertion

```
    (i). Input
Take list tk to stored the top scoring items
Tuple(i,index)
    (ii). Iteration
    (a). If length(tk) < k then
Insert(i, index) into tk in ascending order of items
    (b). Else
For all element ϵ tk do
If i<element[0] then Continue
Else Discard tk[0], insert (i, index) into tk in ascending order of item
EndIf
EndFor
EndIf
```

VII. Experimental Result

Some outcomes are resulting from this scheme:

A. Response Time

Fig. 3 shows a graph in which time require to get search result after adding number of documents in database. If database size increases then time require to get result increases.

Results must require less time for MD search as compare to MRSE technique.

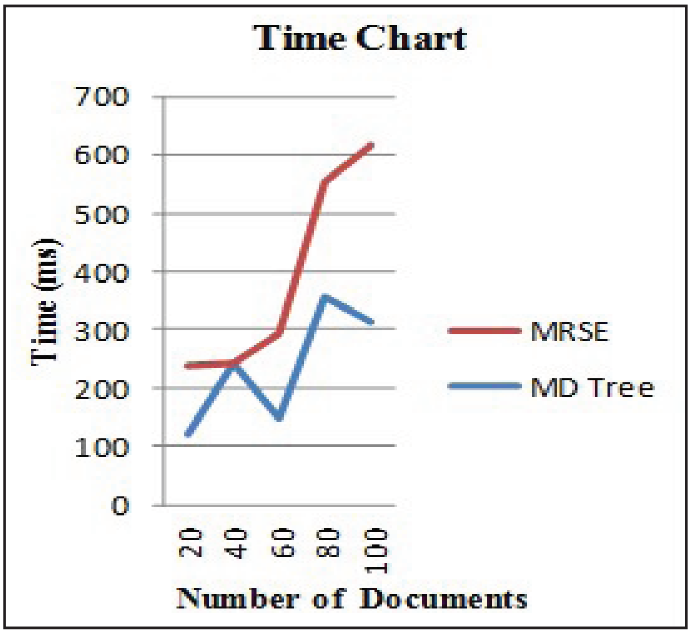

Fig. 3: Response Time

B. Encryption time

Fig.4 shows a graph in which graph shows the expected comparative analysis for time requires to encrypt keywords using both techniques.

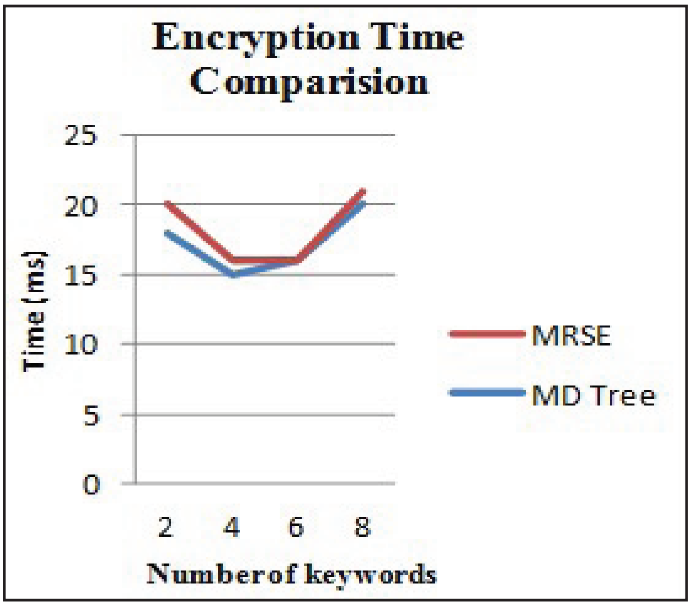

Fig. 4: Encryption Time Comparison

VIII. Conclusion & Future Scope

The previous work mainly focused on providing privacy to the data on cloud in which using multi-keyword ranked search was provided over encrypted cloud data using efficient similarity measure of co-ordinate matching. The previous work also proposed a basic idea of MRSE using secure inner product computation. There was a need to provide more real privacy which this paper presents. In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the CSP and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained. In this paper, the asserting problem of searching encrypted cloud data using ranked multi-keyword (MRSE) is defined and solved. Out of distinct multi-keyword semantics, the adequate similarity measuring of "coordinates matching" and "inner product similarity, i.e., possibilities of many matches for capture the documents from query search perceptible evaluations for similarity measures. Adopting the basic idea for the MRSE based on secure inner product computation and archive privacy requirements in two distant thread models. Experiments based on the real-world data further showing an indeed advent of low overhead on computation and communication. In future, the cloud server is treated as entrusted state, the integrity checking of the rank order in search results analyze.

References

[1] Wei Zhang, Yaping Lin, Siwang Zhou,"Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing".

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia,"A view of cloud computing".

[3] C. Wang, S. S. Chow, Q. Wang, K. Ren, W. Lou, "Privacy preserving public auditing for secure cloud storage," Computers, IEEE Transactions on, Vol. 62, No. 2, pp. 362–375, 2013.

[4] D.Song, D.Wagner, A.Perrig,"Practical techniques for searches on encrypted data", In Proc. IEEE International Symposium on Security and Privacy (S&P'00), Nagoya, Japan, Jan. 2000, pp. 44–55.

[5] R. Curtmola, J. Garay, S. Kamara, R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," In Proc. ACM CCS'06, VA, USA, pp. 79–88, 2006.

[6] P. Golle, J. Staddon, B. Waters,"Secure conjunctive keyword search over encrypted data", In Proc. Applied Cryptography and Network Security (ACNS'04), Yellow Mountain, China, pp. 31–45, 2004.

[7] C. Wang, N. Cao, J. Li, K. Ren, W. Lou,"Secure ranked keyword search over encrypted cloud data," In Proc. IEEE Distributed Computing Systems (ICDCS'10), Genoa, Italy, Jun. 2010, pp. 253–262.

[8] N. Cao, C. Wang, M. Li, K. Ren, W. Lou,"Privacypreserving multi-keyword ranked search over encrypted cloud data", In Proc. IEEE INFOCOM'11, Shanghai, China, Apr. 2011, pp. 829–837.

[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. IEEE INFOCOM'10, San Diego, CA, Mar. 2010, pp. 1–5.

[10] P. Xu, H. Jin, Q. Wu, W. Wang,"Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack", Computers, IEEE Transactions on, Vol. 62, No. 11, pp. 2266–2277, 2013.

Sindhuja Somisetty is Pursuing M.Tech (Computer Science and Engineering) in QIS Institute of Technology, Ongole, Prakasam Dist, Andhra Pradesh, India.

Dasari Ravi Kumar is currently working as Asst. Professor in QIS Institute of Technology, in the Department of Computer Science & Engineering, Ongole, Prakasam Dist, Andhra Pradesh, India.