Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data

Siva Sankar.J[1], Naryanana.S[2]

[1]Student, M.Tech CSE, Gudlavalleru Engineering College
[2]Associate Professor, Dept of CSE, Gudlavalleru Engineering College

Abstract - To protect data privacy, sensitive cloud data has to be encrypted before outsourced to commercial public .Traditional technique follows Boolean search which is not yet sufficient to meet the demands by large number of users and huge amount of data files in cloud. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy.
Keywords – confidential information, cloud computing, Ranked keyword search, searchable encryption, order-preserving mapping.

I. INTRODUCTION

The cloud computing idea is not new. The scale of the cloud has an emotional change extraordinarily from what it was from the earliest starting point. As the engineering and business situations had advanced the status of cloud processing has changed. What was known as distributed computing long prior, however the uses in data today have changed by a tremendous degree.The ascent of the Internet starting in the mid-90s changed how machines could[1] be utilized and how data could be spread. With the thought of utility registering long gone, organizations such as Amazon had started to saddle the power of server ranches to offer a gaggle of items to would-be purchasers.

1.1 Open source incloud

Cloud offers an open-source Environment called Cloud Foundry, a Platform-as-a-Service that ought to hit fear in the hearts of its contenders, particularly any semblance of Salesforce.The nature's turf will offers programming designersand developers the application to fabricate apparatuses on open mists, private mists and wherever else, whether the basic server runs.

Developers, programmers venture Information Technology shops will soon have an alternate alternative for Paas (stage as-an administration) regarding Cloud[13] Swing, an approaching offering from Open Logic that fortifies on its center business environment of giving specialized client aid to open source programming stage. Cloud Swing clients can utilize the stage to gathering programming pieces of both open source and business items for utilization on cloud base administrations, for example, Elastic Compute Cloud, Amazon Ec2.

1.2 Cloud computing with its security

The security in cloud is amazing that is to incorporate consistently with the IT security in your server farm. On the other hand, the cloud administration supplier actualizes IT security associates.

- To spare and secure clients, clients, vendors from outside dangers.
- To guarantee that each individual client actualizing situations are diverse from each other.
- For every single kind of cloud environment administration, the supplier conveys a decent understanding arrangement of the IT security.
- IT security programming, for example, firewalls, interruption discovery frameworks, virtual private systems (PNs), and secure associations and equipment that the cloud supplier has set up.
- Came to know how the cloud suppliers are ensuring the general processing environment.

1.3 Basic model of system

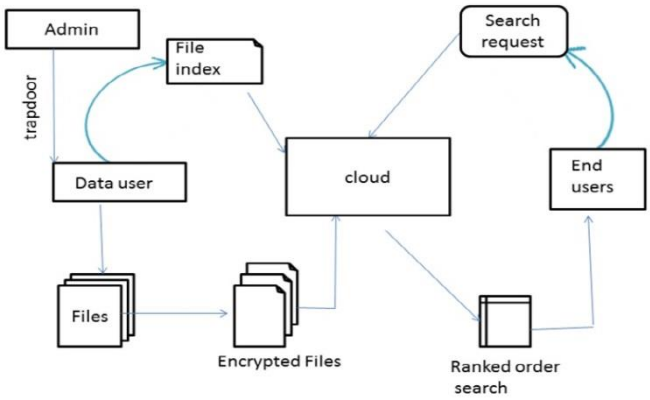

Fig 1: Basic model of system

1) Admin, who is the genuine holder of the database. It will create trapdoors for the verified clients.
2) Users are the parts in a gathering who are qualified for access (a piece of) the data of the database. They will transfer documents and recover records focused around hunt demand.
3) Server is an expert substance (e.g. cloud) to offer data administrations to approved clients. It is frequently obliged that the server is absent to substance of the database it keeps up, the inquiry terms in questions and reports recovered.

In fig 1.1 chart information client will transfer records and make record document for each of the record and after that will encode the document utilizing Data Encryption Standard (DES) and afterward store the information in cloud. In flip side client will hunt down information in cloud, the looked substance will be encoded configuration and it have to unscramble the information and after that will show the query item in a positioned arrangement[11]. Positioned arrangements of results are acquired utilizing the quantity of times the essential word is rehashed in each one record and by the client decision of documents, which incredibly enhances the productivity of positioned catchphrase seek. At last the yield via query output will contain pertinent information and also positioning of the statement and recurrence of the expression will be shown in a positioned arrangement.

1.4 Efficient Ranked Keyword Framework

The client gathers the information records and encodes into obscure arrangement by some security key the documents utilizing DES encryption and creates a mystery key. At that point information client creates the searchable record terms from the extraordinary words which was concentrated from document gathering. The underneath table 1 contains the example words and record terms which was

concentrated from document gathering. At that point the file terms are distributed on cloud server with encoded document.

Word	File Index
Soap	ravi.txt(1,1000)
Software	file3.txt(7,1002)
Protocol	ravi.txt(4,1001)
Dot net	file3.txt(2,1001)

Table 1: Index of words

Calculation of Rank: The minute record indexing is over, next rank is computed. For ascertaining the rank for each one document, term recurrence, record recurrence, the length of documents and the quantity of archives that the information client has in his accumulation needs to be know. The term recurrence computed focused around how frequently the watchwords happens in the same archive , and for each one record , and for each one term this needs to be figured. The report recurrence figured focused around how often a specific pivotal word exists in the diverse records. Given beneath Table 2 demonstrates the positions got for diverse decisive words in distinctive records, for instance cleanser catchphrase is rehashed in record Ravi 3 times and in pooji it is rehashed 2 times so rank 1 is given to ravi and second rank is given to pooji document.

Keywords	File1	File 2
Soap	ravi.txt(3,1000)	pooji.txt(2,1001)
Software	myfile1.txt(5,1000)	lucky.txt(2,1001)
Protocol	Pooji.txt(4,1000)	Ravi.txt(2,1001)
Programming	Lucky.txt(7,1000)	Pooji.txt(2,1001)

Table 2:Rank calculation

II. LITERATURE SURVEY

Writing review is the most paramount venture in programming improvement process. Before creating the device it is important to focus the time variable, economy and organization quality. Once these things are fulfilled, then next step is to figure out which working framework and dialect can be utilized for creating the apparatus. Once the developers begin building the apparatus the software engineers need part of outer backing. This backing can be acquired from senior developers, from book or from sites. Before building the framework the above attention are considered for creating the proposed framework.

Boneh.d, Crescenzo G. D.,ostrovsky.r and Persiano.g[1] portray the idea of open key encryption with magic word pursuit .Consider client Bob who sends email to client Alice encoded under Alice's open key. An email passage needs to test whether the email contains the magic word "critical" with the goal that it could course the email appropriately. Alice, then again does not wish to give the portal the capacity to unscramble all her messages. A component that empowers Alice to give a key to the door that empowers the passage to test whether the statement "critical" is a pivotal word in the email without

learning whatever else might be available about the email is demonstrated. This component is spoken to as Public Key Encryption with catchphrase Search. As an alternate sample, consider a mail server that stores different messages freely encoded for Alice by others. Utilizing this Alice can send the mail server a key that will empower the server to recognize all messages containing some particular pivotal word, yet learn nothing else.

Curtmola.r, Garayj.a , Kamara.s, and Ostrovsky.r [2] portray about the searchable symmetric encryption method for different clients. This encryption (SSE) permits a gathering to outsource the capacity of its information to an alternate gathering (a server) in a private way, while keeping up the capacity to specifically seek over it. This issue has been the center of dynamic research lately. They given two answers for SSE that all the while appreciate the accompanying properties.

Singhal.a[3] characterizes how to relegate a similitude measure to each one archive that demonstrates how nearly it matches a question. Boolean inquiries are by all account not the only system for scanning for data .If some accurate subset of the report being looked for is known, then they are absolutely proper, which is the reason they have been so fruitful in ranges, for example, business databases and bibliographic recovery frameworks .Often, in any case, the data prerequisite is less decisively known. Therefore, it is off and on again helpful to have the capacity to tag a rundown of terms that give a decent sign of which archives are pertinent, however they won't fundamentally all be available in the reports looked for. The framework ought to rank the whole gathering regarding the inquiry, so the main 100, say, positioned reports can be inspected for significance and those that constitute the answer set concentrated.

Song.d, wagner.d, and perrig.a [4] propose cryptographic plans for the issue of seeking on encoded information and give evidences of security to the ensuing such crypto frameworks. It is attractive to store information on information stockpiling servers, for example, mail servers and record servers in encoded structure to decrease security and protection dangers. Anyhow this typically intimates that one needs to give up usefulness for security. Case in point, if a customer wishes to recover just archives containing certain words, it was not formerly known how to let the information stockpiling server perform the pursuit and answer the inquiry
without loss of information secrecy.. The procedures have various pivotal focal points. They are provably secure.

III. PROPOSED WORK

3.1 Ranked based index system: The reason for investigation of the rank based record framework is to make complete data about the idea, conduct and alternate stipulations like execution measure and the framework enhancement. The fundamental objective of this investigation is to totally tag the specialized subtle elements for the rank based record in a succinct and unambiguous way.
The initial phase in creating anything is to express the prerequisites. This applies the same amount of to heading edge explore as to basic projects and to individual projects, and additionally to vast collaborations. Being dubious about your target just defers choices to a later stage where changes are considerably all the more excessive.
The issue explanation ought to state what could possibly be done not how it is to be carried out. It ought to be an announcement of necessities, not a proposal for an answer. A client manual for the coveted framework is a decent issue articulation. The requestor ought to show which gimmicks are compulsory

and which are discretionary, to stay away from excessively obliging outline choices. The requestor ought to abstain from depicting framework internals, as this confines execution adaptability. Execution particulars and conventions for communication with outer frameworks are authentic necessities. Programming designing measures, for example, measured development, plan for testability, and procurement for future augmentations, are additionally legitimate.

Positioned hunt extraordinarily upgrades framework ease of use by giving back where its due documents in a positioned request with respect to certain importance criteria (e.g., pivotal word recurrence), subsequently making one stage closer to reasonable sending of protection safeguarding information facilitating administrations in the connection of Distributed computing.

In the factual measure approach administrator verify clients by issuing trapdoors so clients can recover just that specific information agreeing what they have spared in cloud .In Distributed computing, outsourced document accumulation may be gotten to as well as overhauled regularly for different application purposes. The score elements is utilized as a part of the searchable file for a progressed framework, which is reflected from the comparing record accumulation upgrades. It is outlined as including recently encoded scores for recently made documents, or adjusting old scrambled scores for change of existing records in the record accumulation.

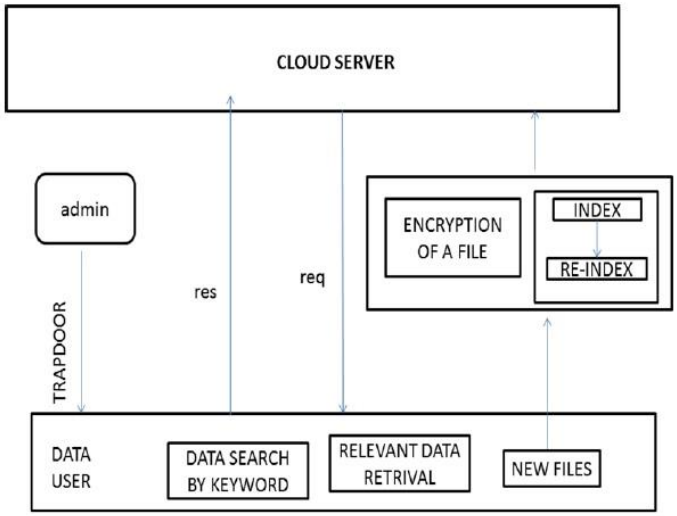

Fig 2:Architecture

3.2 Module Description

3.2.1 Providing a secure data transfer between cloud and user: User has to upload files in the cloud for that in order to provide Security files must be encrypted and then saved in the cloud.

3.2.2Maintenance of index file with relevant keywords: Based on the file which is saved in the cloud it will take word by word from file and build index.

3.2.3 Updating the index file at every updated data in the cloud: Each time when user updated with new file we will found new keywords . These keywords are also appended in the index file.

3.2.4 search and retrieval of files: User search his own files in cloud by using keyword search based on ranking it will retrieve the files in ranked order. Fig 3 demonstrates the pursuit and recovery process,

where the clients can send the inquiry decisive word solicitation to the cloud yet the cloud will give the positioning request of the documents recovery reaction to just approved client.

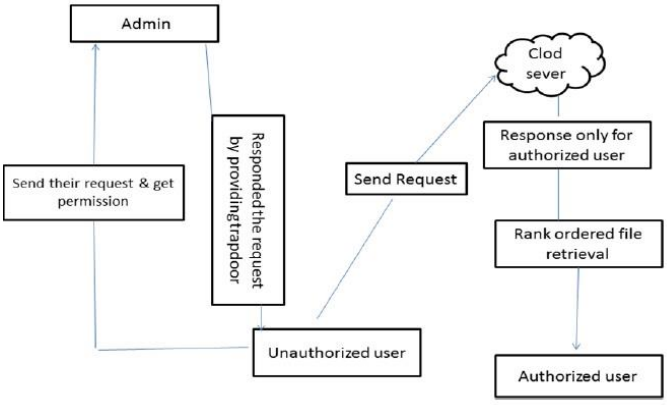

Fig 3: Method of search and retrieval

3.2.5 Ranking of search results: User can choose any of the files which is rankly retrieved form cloud and download it again when he search the same word the ranking of files will change according to user preference.

3.3 Algorithm for ranking

step 1: Read word by word from the file which is to be uploaded.
step 2: count number of time the word repeated in that file
step 3: check for that word whether present in index file or not step 4: if found, re arrange the order of previously saved file according to their word counts. Else append the word with its file name and count at the end of index file
step 5: upload the file to cloud
step 6: close

Table 3: Algorithm for ranking

3.4 Algorithm for search

step 1: Enter keyword to search
step 2: verify whether keyword is present in index file or not
step 3: if not present , display error message " word not found" . Else display the order of filenames , as placed in the index file(ranked order)
step 4: If updated by user, re arrange the order, placing the selected file 1st , and add to the index file in place of previous order
step 5: save index file,
step 6: close

Table 4: Algorithm for search

IV. CONCLUSION AND FUTURE SCOPE

Ranked keyword search on remotely stored information is done by saving files in cloud and retrieve the files by searching through the keywords. recovered files are demonstrated in ranked order which is done by using ranking algorithm in the index page. Security for data stored in cloud is done trough saving encrypted files and privacy of data is maintained by providing different trapdoors to dissimilar users. Ranked analysis is done by score dynamics i.e taking the user choices into consideration and giving highest rank to user chosen file so that user can get more efficient results.

As for the future work, the effectiveness of ranked keyword search is increased by concepts public-key systems that support comparison queries on encrypted data as well as more general queries such as subset queries and has to support arbitrary conjunctive queries (P1, P2..... Pn) without leaking information on individual conjuncts.

REFERENCES

[1] [1] P. Mell and T. Grance, ―Draft nist working definition of cloud computing,‖ Referenced on Jan. 23rd, 2010 Online at http://csrc. nist.gov/groups/SNS/cloud-computing/index.html, 2010.
[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, ―Above the clouds: A berkeley view of Utiliy computing,‖ University of California, Berkeley, Tech. Rep. UCBEECS- 2009-28, Feb 2009.
[3] Cloud Security Alliance, ―Security guidance for critical areas of focus in cloud computing,‖ 2009, http://www. cloudsecurityalliance.org.
[4] Z. Slocum, ―Your google docs: Soon in search results?‖ http:// news.cnet.com/8301-17939 109-10357137-2.html, 2009.
[5] B. Krebs, ―Payment Processor Breach May Be Largest Ever, Online at http://voices.washingtonpost.com/securityfix/2009/01/ payment processor breach may b.html, Jan. 2009.
[6] I. H. Witten, A. Moffat, and T. C. Bell, ―Managing gigabytes: Compressing and indexing documents and images,‖ Morgan Kaufmann Publishing, San Francisco, May 1999.
[7] D. Song, D. Wagner, and A. Perrig, ―Practical techniques for searches on encrypted data,‖ in Proc. of IEEE Symposium on Security and Privacy‘00, 2000.
[8] E.-J. Goh, ―Secure indexes,‖ Cryptology ePrint Archive, Report 2003/216, 2003, http://eprint.iacr.org/.
[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, ―Public key encryption with keyword search,‖ in Proc. of EUROCRYP‘04, volume 3027 of LNCS. Springer, 2004.
[10] Y.-C. Chang and M. Mitzenmacher, ―Privacy preserving keyword searches on remote encrypted data,‖ in Proc. of ACNS‘05, 2005.
[11] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, ―Searchable symmetric encryption: improved definitions and efficient constructions,‖ in Proc. of ACM CCS‘06, 2006.
[12] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, ―Searchablesymmetric encryption: improved definitions and efficientconstructions,‖ in Proc. of ACM CCS‘06, 2006.
[13] A. Singhal, ―Modern information retrieval: A brief overview,‖ IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.

