iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4047

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Efficient File Retrieval from Cloud Servers Using Multi Keyword Sets

Ms. Aishvarya. S[1], Ms. Kirubha. C. V. N[2], Asst. Prof. Mrs. Boomija. M. D [3]
[1,2,3] *Department of information technology Prathyusha engineering college, Chennai - India*

Abstract: ***A Huge number of information proprietors have moved our information into cloud servers. Cloud information proprietors like to outsource archives in an encoded shape with the end goal of protection safeguarding. Hence it is vital to creating proficient and dependable cipher text seeks procedures. One test is that the relationship between archives will be ordinarily hidden during the time spent encryption, which will prompt critical hunt exactness execution corruption. They all get to the information from cloud utilized the catchphrase based inquiry. Approach bunches the records Based on the base importance edge, and after that parcels, the subsequent groups into sub-groups until the imperative on the most extreme size of the bunch is come to. Here we proposed the safe multi catchphrase positioned look from the encoded information from the cloud. It opens operations like an upgrade, erase, and the addition of archives. Here utilizing tree structure and shapeless scan strategy for recover the information from the cloud. These sorts of the strategy used to take care of the issue of watchword speculating assault. Here we proposed the Blowfish system for the encryption procedure. Here to diminish measurable assaults, apparition terms are added to the record vector for blinding list items. The proposed plan can accomplish linear search, semantic search, K gram1 and K gram2 searches and the query item like a number of record recovery additionally manages erasure and inclusion of reports adaptable.***

Keywords: Cloud search, Multi – Keyword, cipher text search, ranked search

I. INTRODUCTION

As Cloud Computing gets to be predominant, more touchy data are being concentrated into the cloud, for example, messages, individual wellbeing records, government archives, and so forth. By putting away their information into the cloud, the information proprietors can be calmed from the weight of information stockpiling and upkeep in order to appreciate the on-request fantastic information stockpiling administration. Nonetheless, the way that information proprietors and cloud server are not in similar trusted area may put the outsourced information at hazard, as the cloud server may never again be completely trusted. It takes after that touchy information typically ought to be scrambled preceding outsourcing for information security and battling spontaneous gets to. In any case, information encryption makes successful information usage an extremely difficult errand given that there could be a lot of outsourced information records. Also, in Cloud Computing, information proprietors may impart their outsourced information to a substantial number of clients. We show useful procedures for legitimate combination of pertinence scoring strategies and cryptographic systems, for example, arrange saving encryption, to ensure information accumulations and records and give efficient and exact inquiry capacities to safely rank-arrange archives in light of a question[1]. As an underlying endeavor, we persuade and take care of the issue of supporting productive positioned watchword hunt down accomplishing powerful use of remotely put away encoded information in Cloud Computing[2]. . They give question seclusion to inquiries, implying that the untrusted server can't learn much else about the plaintext than the query output; they give controlled looking, so that the untrusted server can't hunt down a subjective word without the client's approval[3]. Utilizing our component Alice can send the mail server a key that will empower the server to recognize all messages containing some specific watchword, yet learn nothing else. We define the idea of open key encryption with catchphrase hunt and give a few developments[4]. we motivate and solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored encrypted data in Cloud Computing[5].The individual customers may need to simply recoup certain specific data archives they are involved with in the midst of a given session. A champion among the most understood courses is to explicitly recoup reports through catchphrase based chase instead of recuperating all the mixed records back which is absolutely unfeasible in dispersed registering circumstances. Such watchword based pursuit procedure permits clients to specifically recover documents of intrigue and has been generally connected in plaintext seek situations, for example, Google look. Sadly, information encryption limits client's capacity to perform watchword pursuit and in this manner makes the customary plaintext hunt strategies inadmissible down Cloud Computing. Other than this, information encryption

likewise requests the assurance of catchphrase security since watchwords typically contain imperative data identified with the information documents. In spite of the fact that encryption of catchphrases can ensure watchword security, it assists renders the customary plaintext seek strategies pointless in this situation.

II. RELATED WORKS

A. Swaminathan,Y. Mao, G.-M. Su,H. Gou,A. Varna,S. He,M. Wu,and D. Oard in 2007, To build up a structure for confidentiality protecting rank-requested hunt in substantial scale archive accumulations. We investigate methods to safely rank-arrange the records and concentrate the most significant document(s) from a scrambled gathering in light of the encoded seek inquiries. We show useful procedures for legitimate combination of pertinence scoring strategies and cryptographic systems, for example, arrange saving encryption, to ensure information accumulations and records and give efficient and exact inquiry capacities to safely rank-arrange archives in light of a question. Exploratory results on the W3C accumulation demonstrate that these strategies have similar execution to customary scan frameworks intended for non-scrambled information regarding seek exactness.

2.k.sateesh kumar reddy, m.Hymavathi, b.Babu,Syed Abdul Haq in 2012, To characterize and take care of the issue of secure positioned watchword look over scrambled cloud information. Positioned seek incredibly upgrades framework convenience by empowering query item significance positioning as opposed to sending undifferentiated results, and further guarantees the record recovery exactness. As an underlying endeavor, we persuade and take care of the issue of supporting productive positioned watchword hunt down accomplishing powerful use of remotely put away encoded information in Cloud Computing. To secure information protection, delicate cloud information must be scrambled before outsourced to the business open cloud, which makes compelling information usage benefit an extremely difficult assignment. Albeit customary searchable encryption strategies permit clients to safely look over scrambled information through catchphrases, they bolster just Boolean inquiry and are not yet adequate to meet the viable information use require that is inalienably requested by extensive number of clients and gigantic measure of informtion documents in cloud

3.Dawn Xiaodong Song David Wagner Adrian Perrig in 2000, Our procedures have various essential preferences. They are provably secure: they give provable mystery to encryption, as in the untrusted individual from staff serving at table can't take in regardless of which about the plaintext when just given the ciphertext. They give question seclusion to inquiries, implying that the untrusted server can't learn much else about the plaintext than the query output; they give controlled looking, so that the untrusted server can't hunt down a subjective word without the client's approval. They likewise bolster shrouded questions, so that the client may approach the untrusted server to scan for a mystery word without uncovering the word to the server. It is appealing to store in succession on data collection servers, for instance, mail servers and file servers fit as a fiddle to decrease security and assurance threats. But this usually implies that one has to sacrifice functionality for security .

4.Dan Boneh,Giovanni Di Crescenzo, Rafail Ostrovsky,Giuseppe Persiano in 2004, A mail server that stores different messages freely scrambled for Alice by others. Utilizing our component Alice can send the mail server a key that will empower the server to recognize all messages containing some specific watchword, yet learn nothing else. We define the idea of open key encryption with catchphrase hunt and give a few developments. The issue of looking on data that is mixed using an open key system. Consider customer Bob who sends email to customer Alice encoded under Alice's open key. An email entryway needs to test whether the email contains the catchphrase "squeezing" with the target that it perhaps will course the electronic message in like way. Alice, on the other hand does not wish to give the entryway the ability to unscramble each one of her messages.

5.Cong Wang, Ning Cao, Jin Li, Kui Ren , and Wenjing Lou in 2010, To define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria. we motivate and solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. This approach suffers from two main drawbacks when directly applied in the context of Cloud Computing. On the one hand, users, who do not necessarily have pre-knowledge of the encrypted cloud data, have to postprocess every retrieved file in order to find ones most matching their interest; On the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic,which is absolutely undesirable in today’s pay-as-you-use cloud paradigm.

III. PROBLEM STATEMENT

A large number of data owners have moved our data into cloud servers for our convenience for multiprocessing in anywhere can access and reduce our work. In that, secret and sensitive data must encrypt before storing into the cloud to protect our privacy. Currently, they

search the data from cloud used the keyword based search. The dissertation is to provide various types of search options for the users to retrieve the maximum number of file search from the encrypted data in the cloud. Here we can generate lots of keywords for each file using fuzzy search algorithms on uploading the file. While searching files, retrieve the maximum additional files by matching the corresponding generated fuzzy keywords with the file name of all files available in the cloud server.

IV. EXISTING SYSTEM

The existing techniques for keyword-based information retrieval, which are wide, used on the plaintext data to the search from the cloud server. A traditional way to reduce information seepage is data encryption. However, this will make server-side data operation, such as penetrating on encrypted data, become a very challenging task. In the recent years, researchers have proposed many symbols text search scheme by incorporate the cryptography technique. These methods have been proven with provable security, but their methods need massive operations and have high time complexity. In this system have a lot of security issues are there Keyword Guessing Attack will happen the hackers can easily guess the keyword then they can easily hack our content from the cloud server. Existing search system will provide the result only based on the Boolean keyword matching system, it means whether it will find the exact file name same as the keyword than the file will be retrieved from the server, it won't provide any search result for misspelled keywords. And also the existing search system never provide the result based on similar keyword

A. Disadvantage of Existing System

1) Cloud server computes the relevance score between documents and the query.
2) Only exact keyword search is achievable in the application.
3) Searching an encrypted file over the cloud is complex

V. SYSTEM IMPLEMENTATION

We make available the competent search scheme to search the documents from the cloud server using multi-keyword. We contain a server to produce the nebulous keyword set from the file name Here we using the nebulous keyword set it will create the all feasible misspell keywords. Search keyword get encrypt and it will check with the collection of original encrypted the file name in the cloud server if the keyword will get matched then we connect the nebulous keyword set for that particular keyword and we doing to search the file list based on that nebulous keywords also then we retrieve the files from the cloud server and here we consider the searching performance also. Like how much time it will take to complete the task and many files it will retrieve.

A. User Interface
B. File Upload
C. Keyset Generation

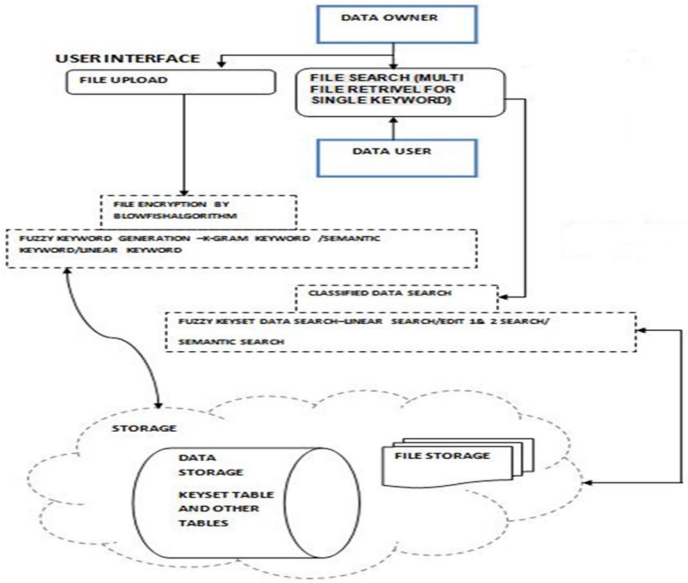


Fig.1 *System* Architecture Design

D. Multi Search.

1) Linear search
2) K gram search
3) Semantic search

E. File Downloads with OTP

1) Login/New User

In this module, the login development itself has lots of security. Usually, the user account name and appropriate password of that account are sufficient to do the justification and login process, but here some more actions are given to make more

F. Upload File

In this module, we want to load the input document then read the input document file and want to implement the preprocessing to that input file. So that the file attached can be processed to the next phases.

G. Search

1) Frequent Search: In this module, we get the non-stop words as input and calculate the count of words and find the repeated occurrence of each and every word from the non-stop words.
2) Similarity Search: From the maximum frequents word we find the weight age of the each and every word than from the weight age value to going to calculate the similarity between the words, based on the similarity we going to group the words into clusters.
3) Linear Search: In this module we are going to create search regarding the keywords, each cluster has n number of similar words as keywords this words we going to find the file for that cluster with the help of lexical analysis tool.

H. File Downloading Process

Document downloading procedure is to get the relating mystery key to the comparing record to the client mail id and afterward unscramble the record information. The record downloading process decoding key to capacity servers with the end goal that capacity servers play out the unscrambling Operation. What's more, the document is downloaded.

I. Algorithms

1) Blowfish Algorithm: An encryption algorithm the stage an significant role in protected the data in accumulate or transferring it. The encryption algorithms are categorized into Symmetric (secret) and Asymmetric (public) keys encryption. In Symmetric key encryption or secret key encryption, only one key is used for both encryption and decryption of data. Eg: Data encryption standard(DES), Triple DES, Advanced Encryption Standard(AES) and Blowfish Encryption Algorithm. In asymmetric key encryption or public key encryption uses two keys, one for encryption and other for decryption. Eg: RSA
2) K Gram Algorithm: We will utilize the k-gram list to recover vocabulary terms that have numerous k-grams in the same way as the question. We will contend that for sensible meanings of numerous k-grams in like manner," the recovery procedure is basically that of a solitary look over the postings for the k-grams in the inquiry string –q. When we recover such terms, we can then locate the ones of slightest alter remove from –q.
3) K Gram: Enumerate all k-grams in the query term. Example: bigram index, misspelled word boardroom._Bigrams: bo, or, rd, dr, ro, oo, om._Use the k-gram index to retrieve “correct” words that match query term kgrams._Threshold by number of matching k-grams.
4) Key-Expansion: It will change over a key of at most 448 bits into a few sub-key exhibits totaling 4168 bytes. Blowfish utilizes extensive number of sub-keys.These keys are producing prior to any information encryption or decoding. The p-cluster comprises of 18, 32-bit sub keys: P1,P2,…….,P18

Four 32-bit S-Boxes consist of 256 entries each.

VI. CONCLUSION

The first occasion when we formalize and tackle the issue of supporting productive yet protection saving fluffy hunt down accomplishing powerful usage of remotely put away, scrambled information in Cloud Computing. We plan a propelled procedure (i.e., trump card based system) to develop the capacity proficient fluffy catchphrase sets by abusing a critical perception on the

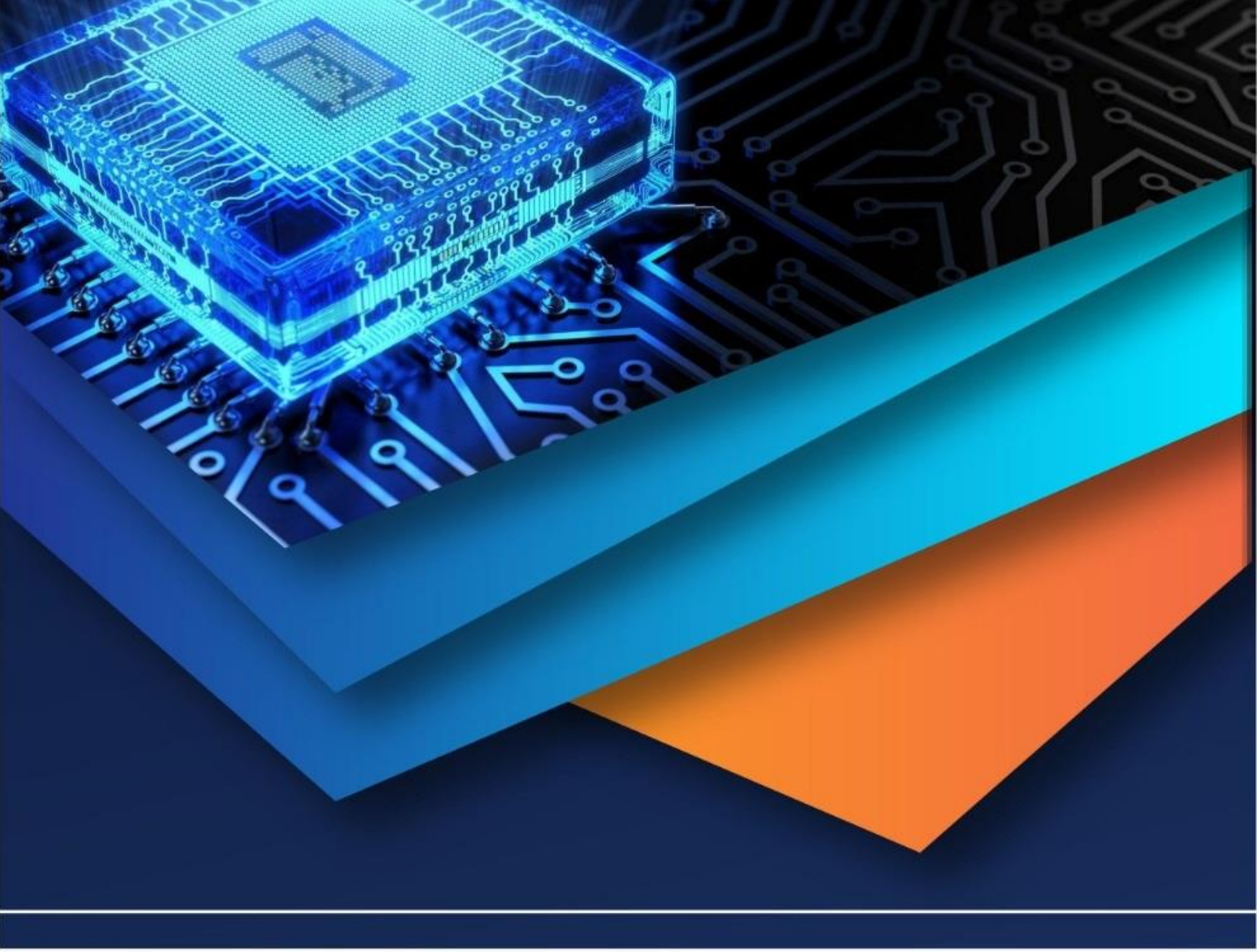
closeness metric of altering separation. In view of the built fluffy watchword sets, we promote propose a productive fluffy catchphrase seek plot. Through thorough security investigation, we demonstrate that our proposed arrangement is secure and protection saving, while precisely understanding the target of feathery catchphrase look.

VII. FUTURE SCOPE

In this scope, we focus on enabling intense yet security protecting soft catchphrase look for in Cloud Computing. To the best of our understanding, we formalize inquisitively the issue of possible delicate watchword search for over encoded cloud information while keeping up catchphrase security. Delicate watchword look in a general sense improves framework ease of use by giving back the arranging records when clients' searching for wellsprings of data definitively sort out the predefined catchphrases or the nearest conceivable arranging chronicles in light of watchword comparability semantics when correct match falls flat.

REFERENCES

[1] A.Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, “Confidentiality-preserving rank-ordered search,” in Proc. ACM ACM Workshop Storage Security Survivability, Alexandria, VA, 2007, pp. 7–12.
[2] C. Wang, N. Cao, K. Ren, and W. J. Lou, “Enabling secure and efficient ranked keyword search over outsourced cloud data,” IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.
[3] D. X. D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44–55.
[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Proc. EUROCRYPT, Interlaken, SWITZERLAND, 2004, pp. 506–522.
[5] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, “Secure ranked keyword search over encrypted cloud data,” in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 253–262.
[6] C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, “A hierarchical clustering method For big data oriented ciphertext search,” in Proc. IEEE INFOCOM, Workshop on Security and Privacy in Big Data, Toronto, Canada, 2014, pp. 559–564.
[7] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, “Achieving secure, scalable, and fine-grained data access control in cloud computing,” in Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1–9.
[8] D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M. C. Rosu, and M. Steiner, “Dynamic searchable encryption in very large databases: Data structures and implementation,” in Proc. Netw. Distrib. Syst. Security Symp., vol. 14, 2014, Doi: http://dx.doi.org/ 10.14722/ndss.2014.23264
[9] S. Grzonkowski, P. M. Corcoran, and T.Coughlin, “Security analysis of authentication protocols for next-generation mobile and CE cloud services,” in Proc. IEEE Int. Conf. Consumer Electron.2011, Berlin, Germany, 2011, pp. 83–87.
[10] M. Naor and K. Nissim, “Certificate revocation and certificate update,” IEEE J. Sel. Areas Commun., vol. 18, no. 4, pp. 561–570, Apr. 2000.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY