Volume 6, Issue 4, April 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Novel and Advanced Mechanism for Keyword Query and Synonyms Search over Encrypted Cloud Data

Jay Kishan Vishwakarma
Student, M.Tech, Department of CSE
All Saints' College of Technology, Bhopal
Madhya Pradesh, India

Shamaila Khan
Associate Professor, Department of CSE
All Saints' College of Technology, Bhopal
Madhya Pradesh, India

Abstract— Adopting cloud computing, a person is able to accumulate their files and records on remote servers & let files accessed to community users through cloud servers. Because the outsourced records are probable to have confidential and personal data, before uploading data into the cloud they are characteristically encrypted. In cloud, complexity of searching over the encrypted record is much important because of the usability of outsourced records. We noticed the problem in this paper and build up a well grained n-keyword search system over encrypted cloud files. We found three techniques while we did complete review on several methods & instruments, Firstly we begin with the significant count & favourite issues winning keywords which allow the exact keyword search & privatise consumer experience. Secondly, we set up realistic and extreme well organize many keyword search system. This future methods will hold up complex reason search of the combine of OR & AND process of keywords. In the third technique, we additionally use the secret sub-dictionaries method to attain improved competence on directory structure, trapdoor produce and enquiry. Finally we analyze the performance of the future methods in the conditions of privacy of documents, confidential safeguard of directory and trapdoor, & unlink capacity of trapdoor. From end to end wide testing use & genuine word record set, we evaluate the presentation of the proposed method. Both the safety examination and trial outcome exhibit that the proposed method can attain the similar safety level compared to the existing ones and improved presentation in conditions of functionality, complexity and competence.

Index Terms— Keyword Search, Cloud Computing, Encrypted File Searching

I. INTRODUCTION

The cloud computing considers computing as usefulness & let out the calculating & storage capability to the community persons [1], [2], [3]. In such a structure, the person can slightly store their record on the cloud server that is information outsourcing & then mark the cloud information open for community right to use the cloud server. This stand for an additional scalable, less price & steady method for community information right to use because of the scalability & more competence of cloud servers, & as a result is Positive to small venture. Notice that the outsourced information may enclose responsive confidential information. It is frequent essential to encrypt the confidential information prior to broadcast the information to the cloud servers [4], [5]. The information encryption, though, would importantly inferior the utilization of information owing to the complexity of searching over the encrypted information [6]. Merely encrypting the information may motionless reason other safety anxiety. For example, Google search utilize Secure Sockets Layer (SSL) to encrypt the link among search user & Google server when confidential information, like email, documents, come into view in the search outcome [7]. But if the search user clicks into an additional website from the search outcome, page, that the website may be capable to notify the search condition that the consumer has used. On deal with above problems, the search able encryption (for example [8], [9], [10]) has been just set up as a primary move towards allow searching over encrypted cloud information, which process the below mentioned procedures, 1st the information holder required to create different keywords according to the outsourced information. These keywords are then encrypted & saved at the cloud server. When the search user requires to right to use the outsourced information, it able to choose few related keywords & send the code manuscript of the chosen keywords to the cloud server. The cloud server then utilize the code text to contest the outsourced encrypted keywords, & finally given back the related outcome to the search user. To attain the similar search competence & accuracy over encrypted information as that of plaintext keyword search, a wide range of dictionary should be provided to the system. In this research we develop a system which can easily handle the searching of keywords in encrypted documents by using proposed novel mechanism.

II. LITERATURE REVIEW

In the cloud systems, searchable encryption methods [11-13] are capable to give safe search over encrypted information for consumer. They set up a searchable upturned directory that saves a catalog of mapping from keywords to the matching bunch of files which enclosed this keyword. If information consumer contributes a keyword, a trapdoor is created for this keyword & then present to the cloud server. Upon receipts of the trapdoor, the cloud server carries contrast among the index & trapdoor, & lastly precedes the information users every files that consist this keyword. But, these techniques merely accept accurate single keyword search.

The issues on secure keyword search over outsourced cloud record will studied some researchers. The suggest a safe graded keyword search methods from Wang et al., [14] Their answer merge upturned index with order-preserving symmetric encryption (OPSE).In condition of graded search, the order of recover files is resolute by arithmetical significance keep count, Which able to be intended by TF x IDF. The related achieve is encrypted by OPSE to make sure safety.

It improves method utilize able & stores conversation slide. This explanation only hold up single keyword graded search. As per Cao et al., [15] suggest a technique that take on resemblances gauge of "Organize corresponding" to imprison the related of files to the enquiry. They utilize "internal creation resemblance" to gauge the hold of every folder. This explanation hold up accurate several keyword graded search. It is sensible & the search is supple. According to Sun et al., [16] future a MDB-tree stand system which ropes graded several -keyword search. This method is more competence, but the advanced competence will guide to inferior accuracy of the search out come in the methods. Added fuzzy keyword search [17-19] have been urbanized. These techniques use a spell identification device, like search for "wireless" in its place "Wireiess", or the information arrangement may not be the similar for example ; " record mining" "versus data mining",

According to Chuah et al., [17] suggest a solitude conscious bed-tree technique to bear fuzzy several keyword search. This will move towards consumer edit space to develop fuzzy keyword sets. Bloom sift are build for each keyword. Then, it builds the index tree for all folders where every leaf lump a confusion worth of keyword. Li et al., [18] use edit space to count keywords, resemblance & make storage – essential fuzzy keyword arranges. Particularly the wildcard – based fuzzy arrange build up move towards is intended to save storage slide. According to Wang et al., [19] use wildcard- based fuzzy arrange to make a confidential tire- cross search index.

In the searching stage, if the edit space among recovery keywords & ones from the fuzzy arranges is not as much of prearranged set of worth., It is measurable related and income the matching files. This fuzzy search will not bear semantic unclear search but this search techniques bear lenience of small typos & arrangement discrepancy. Allowing for the survival of polysemy and synonymy [20], the replica that ropes more keyword graded search and semantic search is more sensible.

In these documents, we will resolve the issue of multi- keyword dormant semantic graded search over encrypted cloud information & get back the for the most part related folders. Latent Semantic Analysis (LSA) is the fresh method that we define, based different keyword graded search which ropes multi-keyword dormant semantic graded search. To utilizing latent semantic Analysis, the planned system might arrival not merely the precise corresponding files, other than also the files as well as the conditions dormant semantically linked to the enquiry keyword. For e.g. when the customer enter the keyword "Car" to search the file, the planned techniques returns not only the file having "Car" but in addition the file as well as the word "automobile" we obtain great medium of word document organization information & build a semantic space in which provisions & documents are strongly connected are located close to one another. Without solitude break to meet the confront of supporting such different keyword semantic is to we suggest the plan: As per "Latent Semantic Analysis" the multi-keyword ranked search (MRSE) using. The flow of this paper is prepared as follows. We explain the system model and structured aims in Part III. Part IV explains about the proposed scheme representation. Part V details about the implementation, notations, solitude needs and algorithms involved. Part VI covers the results diagrams and efficiency analysis. Review and conclusion continues in Part VII.

III. SYSTEM MODEL

The system model is able to think of in 3 body, as shows in shape 1: the record holder, the record utilize and the cloud server.

D = { d1, d2, ...,dn }is the record gathered from record holder and a group of separate keywords W = {W1, W2, ...,Wn} is taken out from the record set D. From the record gathering D the data holder will initially build an encrypted search able index I. All files in D are encrypted & create a fresh file gathering, C. After that the record holder uploads together the encrypted record gathering C and encrypted index I to the cloud server. The record utilize gives keywords for the cloud server, and the matching trapdoor w T seek manage device is created. In these documents, we guess that the approval among the record utilizes & record holder roughly complete. The cloud server work out & gives back to the equivalent group of encrypted files before that the cloud server inward w T from the official operator. Additionally the cloud server calculates and retrieves documents that are highly related to the search query.

In case the information user may post an elective number 1 along with the trapdoor T so, theses all because of to decrees the message charges. The cloud server operator files encrypted file outsourced index, outsource search demand top-1 graded documents record holder numeral 1. Structural designed of graded search over encrypted cloud record B. The danger mock-up & planned aim the cloud server is measured as "truthful- but- interested" in our model. Mainly the cloud server together goes behind the chosen procedure requirements however simultaneously analyzed records in its storage & communication flows inwards throughout the procedure consequently as to study extra data [20]. In these documents, we have reason to attain safety & graded search beneath the above mentioned model.

A. The structured aims of our methods are following

As per Latent semantic Search : our main goal is to research the latent semantic bond among terms & documents and implement it in real time file system on server. The planned method attempt to place related substance close to each other in little distance in turn that it might return the information user the files obtain the state latent semantically connected with the enquiry keyword. Different keyword graded search, the cloud server support equally support outcome grading &

different keywords enquiry. Safety protecting: Our methods to get the solitude necessity and protect the cloud server from knowledge through extra data from index & trapdoor.

The directory privacy: The TF principles of keywords are saved in the directory. Therefore, directory saved in the cloud server needs to be encrypted.

Trapdoor Unlink Ability: The index examination over the search outcome also is done in cloud server in the mean time so that when user makes query two time then the similar query must create diverse trapdoors.

Keyword Solitude: cloud server cannot distinguish the keyword in search, index by studying the term frequency as TF.

IV. PROPOSED SCHEME

A. Synonym Expansion

With the intension of develop the correctness of searching outcome, the synonyms will be use. The synonyms are the language with the similar meaning will provide. If the keywords taken from outsourced text documents require being comprehensive by ordinary synonyms, As clod users seeks contribution might be the synonyms of the pre distinct keywords, not the clear or unclear similar keywords due to the probable synonym replacement and not have accurate education about the information. Then the keyword put is comprehensive by utilizing the build up synonym lexicon.

B. Rank Function

In data recovery, a graded purpose is regularly used to estimate pertinent achieve of identical files to a demand. Between plenty of graded purpose, the "TF X IDF" law is more extensively utilize.

Where Term Frequency (TF) give incidence of the term come into view in the file, and Inverse document frequency (IDF) is frequently got by separating the whole count of documents by the amount of files enclose the term. That means TF stand for the significance of the word in the document & IDF point out the weight or amount of difference in the entire text gathering.

Every File is matching to a directory vector D that saves regularized TF significant, and the enquiry vector Q saves regularized IDF importance. All measurement of D or Q is relevant to a keyword in W, and the order is similar with that in W, that is, D [i] d is matching to keyword i w in W.

.

V. IMPLEMENTATION

TABLE I. NOTATIONS

Notation	Meaning
$Dict$	Data Dictionary containing Keywords and Synonyms
$Admin$	Administrator
$User$	Application User
KW	Search Keyword
TF	Term Frequency of Keyword
$Rank$	Ranking
$File$	File
Syn	Synonym of Keyword

a. Notations for Algorithm

Fig. 1. Example of a figure caption. *(figure caption)*

A. Building Dictionary of Keywords & Synonyms

Administrator inserts the keywords and related numerous synonyms of it in dictionary while system setup.

$Admin \uparrow Dict\ where\ Dict = \{KW, |Syn|\}$

$|Syn_i| \in KW_i\ where\ |Syn_i| = \{Syn_1, Syn_2, \dots, Syn_n\}$

B. File Outsourcing

User has to logins and outsources word document plain text files along with the related keywords from the set of dictionary keywords. Document gets encrypted using predefined Secret key using AES algorithm.

$User \rightarrow login \rightarrow validate(User_{id}, User_{pwd})$

$if(true)$

$User \uparrow (File, |KW|)$

$end\ if$

C. File Keyword Rank

Server discovers frequency of keywords and its synonyms in the document and combines both rate of recurrence to assert TF for each KW. File is again operated for ciphering using predefined encryption key.

$server \rightarrow Operate(File, |KW|)$

$|Syn| = checkSynonym(|KW|)$

$parse(File)$

$\boldsymbol{While}\ (KW, Syn) \in |KW|\ \boldsymbol{do}$

$TF_{KW} = checkFrequency(KW)$
$TF_{Syn} = checkFrequency(Syn)$
$Rank_{KW} = TF_{KW} \cup TF_{Syn}$
$save(Rank_{KW})$
end loop
$'File = Enc(File, EKey)$
$save('File), Delete(SFile)$

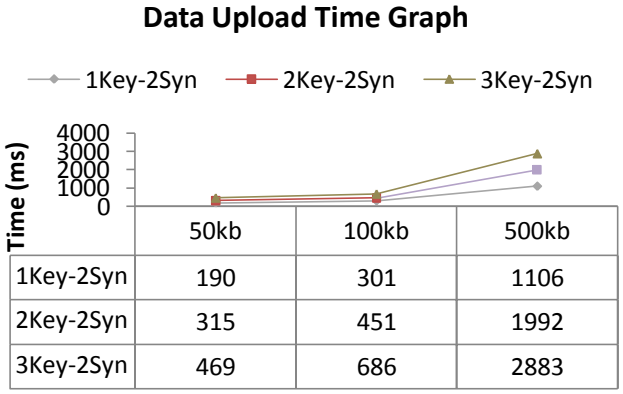
D. File Search

User makes a search for a keyword or multiple keywords within the cluster of files, ranking of every keyword is derived and an average is taken by total number of keywords to determine cumulative ranking of document for the searched keyword.

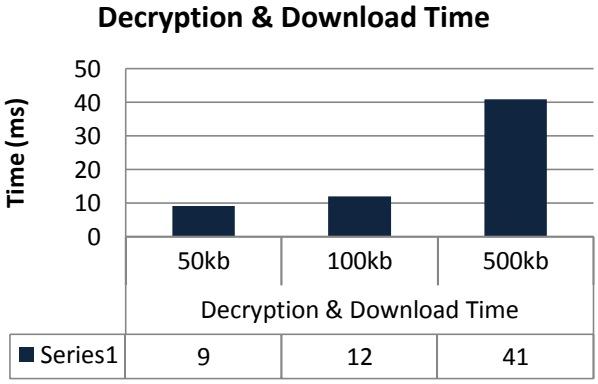
$User \rightarrow find(|KW|)\ where\ |KW| = \{KW_i, KW_{i+1}, \ldots., KW_n\}$
$\forall KW \in |KW|$
$Rank_{KW} = checkRank(KW)$
$cumulativeRank('File) = \frac{\{Rank_{KW_i}, Rank_{KW_{i+1}}, \ldots, Rank_{KW_n}\}}{n}$
$show(|'File|) \in cumulativeRank('File)$
$User \downarrow File \rightarrow\ File = Decrypt(selected('File), EKey)$

At the end of the search resulting file names are displayed and user is allowed to download document after decryption of the file using predefined key.

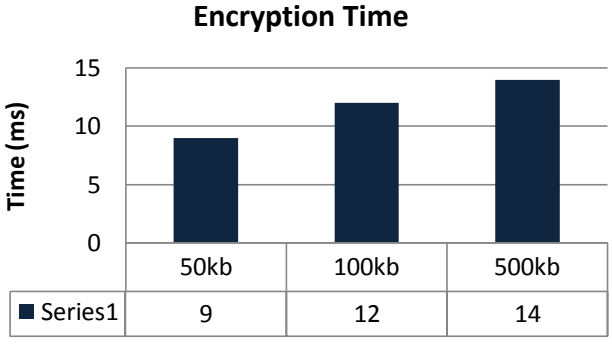
VI. EXPERIMENTAL RESULTS



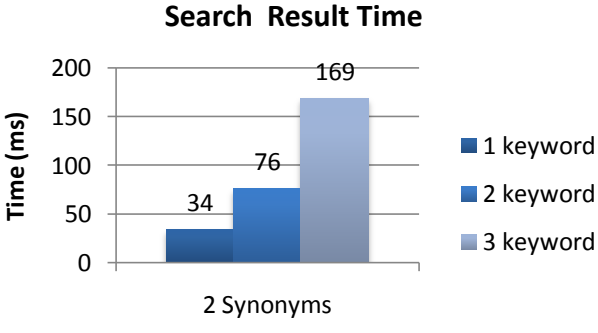
(VIa) Data Upload Time



(b) Decryption & Downloading Time



(c)Encryption Time



(d) Search Result Time

Above fig. (a) demonstrates the file upload time with different sizes and with different no. of keywords and synonyms. Fig. (b) demonstrates the decryption and download time with different sizes of files. Fig. (c) presents the encryption efficiency of AES algorithm in proposed scheme time with different sizes of files. Above fig.(d) demonstrates the search result time with different no. of keywords and synonyms in search query.

VII. CONCLUSION

This research recommends a narrative different keyword seek technique which is enhanced on an encrypted file over Cloud storage. This idea deems both major features which are significant for steady performance of system model without any drawbacks. First and foremost it gives guarantee of safety of the information saved on partial trustworthy server and next it retains the competence of whole methods by utilizing directory and book seeks methods. The searching technique

is just restricted to keyword enquiry seek but also look out of synonyms of the keywords though creating grading. Also as a additional improvement of the methods, trapdoor like, "or" and "and" are utilized which may additionally provide user a friendly atmosphere to create search. For cipher operations AES algorithm is good choice to use because it gives competence with safety.

Upcoming work can be passed on to searching techniques dependent on Natural Language Processing. Also the future structural design can be somewhat improved so that users do not required to identify the keyword but method should itself create relevant keywords and grading of the file on keywords & its synonym.

REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdpbased service model for interdomain resource allocation in mobile cloud networks," IEEE Transactions on Vehicular Technology, vol. 61, no. 5, pp. 2222–2232, 2012.

[2] M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting geodistributed clouds for e-health monitoring system with minimum service delay and privacy preservation," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 430–439, 2014.

[4] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacypreserving data aggregation without secure channel: multivariate polynomial evaluation," in Proceedings of INFOCOM. IEEE, 2013, pp. 2634–2642.

[5] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in Proceedings of GLOBCOM. IEEE, 2014, to appear.

[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.

[7] https://support.google.com/websearch/answer/173733?hl=en.

[8] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proceedings of S&P. IEEE, 2000, pp. 44–55.

[9] R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multikeyword ranked query over encrypted data in cloud computing," Future Generation Computer Systems, vol. 30, pp. 179–190, 2014.

[10] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, "Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage," IEEE Transactions on Emerging Topics in Computing, 2014, DOI10.1109/TETC.2014.2371239.

[11] D. Boneh, "Public key encryption with keyword search", Advances in Cryptology-Eurocrypt 2004, Springer, (2004).

[12] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions", Proceedings of the 13th ACM conference on Computer and communications security, ACM, (2006).

[13] D. X. Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000, Proceedings 2000 IEEE Symposium, IEEE, (2000).

[14] C. Wang, "Secure ranked keyword search over encrypted cloud data", Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference, IEEE, (2010).

[15] N. Cao, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", INFOCOM, 2011 Proceedings IEEE, IEEE, (2011).

[16] W. Sun, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking", Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ACM, (2013).

[17] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data", Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference, IEEE, (2011).

[18] S. Deshpande, "Fuzzy keyword search over encrypted data in cloud computing", World Journal of Science and Technology, vol. 2, no. 10, (2013).

[19] C. Wang, "Achieving usable and privacy-assured similarity search over outsourced cloud data", INFOCOM, 2012 Proceedings IEEE, IEEE, (2012).

[20] S. C. Deerwester, "Indexing by latent semantic analysis", JASIS, vol. 41, no. 6, (1990), pp. 391-407.

