ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.6

Intelligent Keyword Storage in Cloud Document System

A.Nirmal Kumar[1], S.Sageengrana[2], K.Karthick[3]
Assistant Professor[1, 2], PG Scholar[3]
Department of Information Technology and Engineering[1], Department of Computer Science and Engineering[2]
Master of Computer Applications[3]
Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Chennai, India

Abstract:
The Information sharing is that the key target of Cloud space for storing servers. It permits storage of automatic and huge volume of knowledge with restricted price and high access profit. Security got to be needed to be in given due importance for the cloud information with highest care to the knowledge and word to the knowledge owner. The projected approach similarity live of "coordinate matching" combined with "inner product similarity" quantitatively evaluates and matches all relevant information with search keyword to reach best results. This approach, each document is interpretation to a binary vector to characterize a keyword controlled at intervals the document. The search keyword is additionally diagrammatical as a binary vector, that the similarity is additionally specifically measured by the inner quite the question vector with the info vector. The inner choice computation and place along the 2 multi-keyword stratified search over encrypted knowledge (MRSE) schemes ensures knowledge privacy and provides elaborate information regarding the dynamic operation on the knowledge set and index and thence improves the search expertise of the user.

Keywords: Cloud storage servers, Coordinate matching, Inner product similarity, Binary vector, Query vector, Multi keyword ranked search (MRSE).

1. INTRODUCTION:

The term Cloud refers to a Network or web. In varied words, we'll about to say that Cloud may be a few things that square measure gift at remote location. Cloud will give services over network. Service Models unit of measurement the reference models on it the Cloud Computing depends. There part unit many alternative service models all of which may take the shape like as everything as a Service. The planned approach match live of "coordinate matching" combined with "inner product similarity" quantitatively evaluates and matches all relevant information with search keyword to make best results. Then that user can able to transfer identical document with changes in this document that document modified words unit of Measurement updated inside the individual page. These themes have introduced Multi-keyword ranked search and complex quantity computation for data privacy. These two schemes square measure answerable for providing elaborate information of the dynamic operation and automatically improves the search experience for the users. This methodology uses the crypto logic primitives applied to the matter of secure storage inside the presence of untrusted servers and wish for owner managed key distribution.

2. PROBLEM CLASSIFICATION:

Existing system will have a very simple document based search system. Users can search for the documents with the name of the document. For instance, typhoid.doc can be retrieved by searching for the keyword "typhoid". In case the word penicillin is available in the document typhoid.doc it won't be available in the search. Ranked keyword search enhances the feature of retrieving the keyword in the documents too but the documents were stored in the file system. Moving the documents and maintaining the documents becomes tougher in this case. Encrypted information be stored securely in cloud and we don't have an option of hash table to access the data faster and effectively. No clear process flow between the document owner and document accessor. In the existing system, Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Ranked Searchable encryption allows data owner to outsource his data in an encrypted manner while maintaining the selectively search capability over the encrypted data.
In the existing system, the documents and data be stored into secure cloud storage whereas the documents were scanned with the keywords and an index of information were stored for future searching options. The existing system have used symmetric encryption algorithm to store the data securely.The loss of points presented System: No semantic based retrieval is available, No Efficient keyword indexing and multi attribute searching is done, Multi optional key updates to the document requestor such as whatsapp and sms is not available.

Presented Algorithm:
Ranked searchable encryption system
The overall algorithm comprises of four basic steps of generating the key followed by building the index and creating search index for future reference. In the existing system, the author have used the Symmetric Encryption algorithm to maintain and preserve the order and for storing the data safely in a ciphered format.

3. RECOMMENDED SYSTEM:

Proposed systems emphasize and address most of drawbacks of the existing system. Below are the items, Storage Efficiency - > Instead of the file system, the data will be stored in the format of BLOB (Binary Large Objects) in the database. This enhances storage efficiency and maintainability too. Access Efficiency -> Data will be stored in the format of keywords by

automatic intelligent data retrieval process and the data will be indexed efficiently by hashing technique. Therefore, the data will be accessed at a very high speed. Process Efficiency – A clear process flow of document owner uploading the documents with the indexing keywords. Document viewer request for the keywords and document access. An automatic request will be initiated to the owner and in turn on confirming the document. Keys with RSA based encryption emphasized will be sent to the end user to access the documents. Technology Efficiency – key updating can be achieved via what's app or SMS to the end users. Intelligence Efficiency – Semantic information were stored and the data will be retrieved accordingly in a semantic manner. The planned uses of Multi-keyword ranked search over encrypted cloud data (MRSE), "Coordinate matching" by inner product similarity.

ALGORITHM FOR PROJECTED SYSTEM:

B Tree searchable MD5 encryption system

- In the proposed system, we are using Key word based/matching technique in identifying the keywords with the use of inherent efficient storage and searching B-Tree scheme.
- After retrieving the data from the B-Tree, the data will be stored in the format of MD5 encrypted data.
- The whole process involves scanning of the documents with the top-down keyword parsing technique in grabbing the specified keywords into the system to enable a giant knowledge repository.

4. PROPOSED SYSTEM ARCHITECTURE DIAGRAM

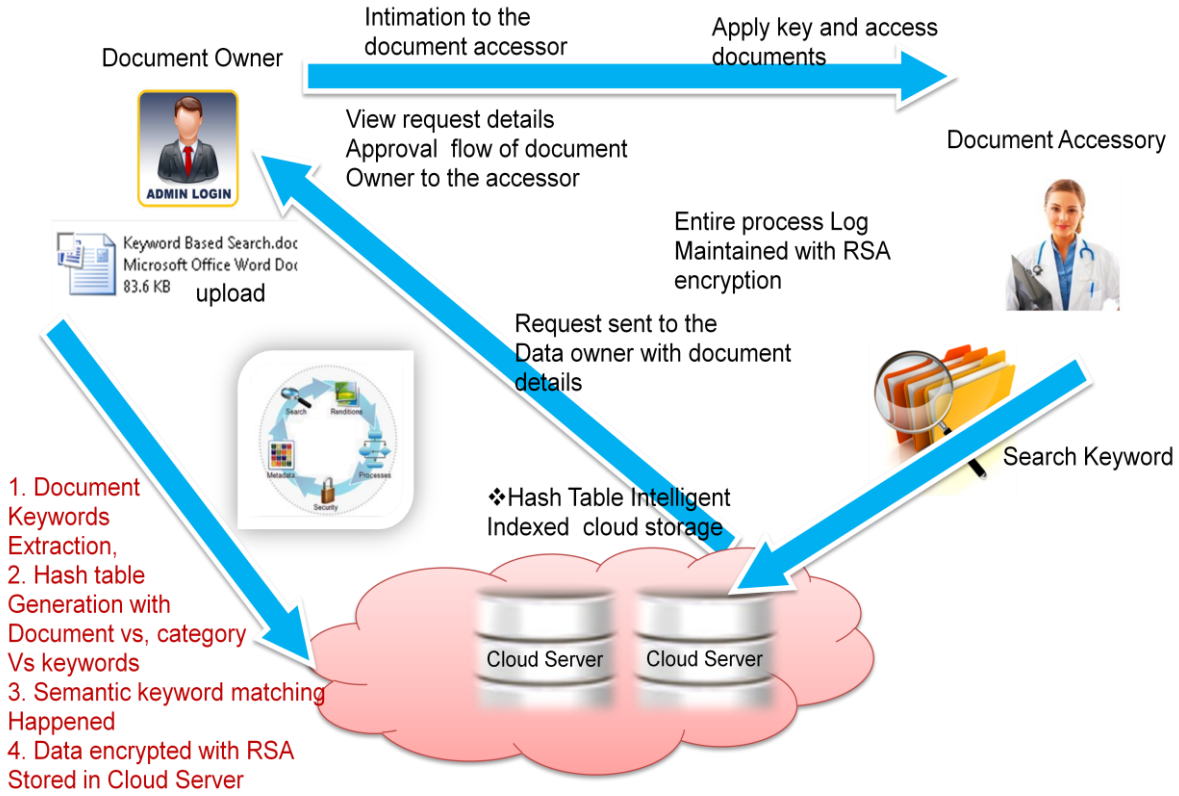

Figure.1. Intelligent keyword searching system

System Architecture: The architecture and entire processing of the proposed system is explained the architecture diagram.

Techniques: There are multiple techniques used for the implementation process.

User Authentication: User's information is stored in database to check whether the user is authenticated or unauthenticated user.

Name Search: After the verification of user's information in a Database. User is allowed to search a document using document name only. Content-wise search is not allowed in this Proposed System.

Stop Word Removal: Documents searched in a database by the order of document's name and its content. Document's content is sorted by using stop word removal technique in a database. Stemming technique used to list the words by the removal of stemming words in a document.

Multi-Keyword Search: Users search the document by using multiple keywords in a database. Removal of words in a database finally sorts out some multi-keywords for the document. These Multi-keywords are sorted by means of using priority basis .Multi-keyword ranked search over encrypted(MRSE) data is a technique involved and returning files in a ranked keyword order regarding to certain relevance criteria (eg: keyword frequency)

5. LITERATURE REVIEW:

Topic: **Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing**

It presents Concert, an approach for the automated analysis of workflows. If a workflow does not adhere to the given rules, re-usable rule patterns are used to pinpoint the workflow vulnerabilities. Auditors can check the rule adherence of workflows before workflow execution, and thanks to the rule patterns certification is open to scrutiny by customers.

A clear business flow is maintained in the system. The request will be directed to the appropriate work flow accordingly. No clear explanation is provided about the alternatives for the deviation in the system.

Topic: Ensuring Data Storage Security in Cloud Computing This project explains about the data blocks and the file encryption happening in the system. One of the important

concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of his data in the cloud. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). The data will be proofed based on the Service Level agreement specified by the stack holders. Practical implementation of this project is not focused clearly. Author provides a suggestion on the encryption process without any explanation on the steps to follow to incorporate this functionality in the cloud server.

6. METHODOLOGY

Algorithms: The existing system uses MD5 algorithm. Ranked keyword search enhances the feature of retrieving the keyword in the documents too but the documents were stored in the file system. The overall algorithm comprises of four basic steps of generating the key followed by building the index and creating search index for future reference. The proposed scheme involves Key word based/matching technique in identifying the keywords with the use of inherent efficient storage and searching B-Tree scheme. After retrieving the data from the B-Tree, the data will be stored in the format of MD5 encrypted data. The process of system initialization, user registration for all the user whom wants to access the document, file upload, user revocation process, registration of new user and file download.

System Initialization: The group manager does this operation by first generating a binary map grouping the system S= (q, G1, G2, e (.,.)), once after this processed the user is allowed to select any two elements randomly. Now the group manager publishes the parameters (S, P, W, Y, Z, f, f1, End ()). Here, the group manager keeps the parameters as the secret master key.

Registration for existing user: First step, the user sends the ID (the identity of the user id) as the request to the group manager and the public key which is distributed is in asymmetric encryption algorithm. Once the group manager receives the request the chooses a random number r belongs to Z*q and computes the request using R=e(P,P)r and finally the user gets a verification message.

File upload: Here the group members are allowed to choose an unique data file identity ID data (identity of data) and a random number k belongs to Z*q.

User revocation: This process is done by the group manager with the help of cloud. When a user i from the group in the local storage space and also updating the user group list which is stored in the cloud.

Registration of new user: For the registration of the new user the group manager performs the same operation by naming the identity as IDm+1.

File download: The operation is performed by the group members and the cloud, the group member encrypts the IDdata with their key and sends a request to the group manager. Once the request is received by the cloud it decrypts and compares the encryption key and selects the required document and sends it to the user in the encrypted format.

7. CONCULSION AND FUTUREWORK:

Main goal of this project is evaluation of storing data in cloud more secure. This section include various test conducted on data stored in cloud, these test are conducted on the basic of various parameters. Due to loss and spoil of Data communication, picture on the characteristics of human mind memory in organizing and exploiting episodic events and semantic words in information recall, this paper presents a personal web revalidation technique based on context and content keywords. In future enhancement, the clients hold a alert from cloud admin to approval the other user request. In future enhancement, the user gets an alert from cloud admin to approval the other user request. And also in future the system is used to store and view the file like Image, Video, Audio and etc.

8. REFERENCES:

[1] E. Adar, J. Teevan, and S. T. Dumais, “Large scale analysis of Web revisitation patterns,” in Proc. SIGCHI Conf. Human Factors Comput. Syst., 2008, pp. 1197–1206.

[2] D. Carmel, et al., “Personalized social search based on the user’s social network,” in Proc. 18th ACM Conf. Inf. Knowl. Manage., 2009, pp. 1227–1236.

[3] A. Cockburn, S. Greenberg, S. Jones, B. Mckenzie, and M. Moyle, “Improving Web page revisitation: Analysis, design and evaluation,” Inf. Technol. Soc., vol. 1, no. 3, pp. 159–183, 2003.

[4] T. Deng, L. Zhao, H. Wang, Q. Liu, and L. Feng, “ReFinder: A context-based information re-finding system,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 9, pp. 2119–2132, Sep. 2013.

[5] S.Tyler and J. Teevan, “Large scale query log analysis of refinding,” in Proc. 3rd ACM Int. Conf. Web Search Data Mining, 2010, pp. 191–200.

[6] Jung T, Li X, Wan Z and Wan M, Privacy preserving cloud data access with multi-authorities, in Proc. 32nd IEEE Int. Conf. Comput. Commun, 2013, 2625–2633.