Intelligent Keyword using Binary Vector in Cloud Storage System

R. Anto Rose*, S. Vinitha, G. Keerthana, M. Kavipriya
Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, avadi, Tamilnadu
***Corresponding author: E-Mail: antorose@velhightech.com**

ABSTRACT

Information sharing is the main goal of Cloud Storage servers. This sharing scheme allows sharing and storage of sensitive and large volume of data with limited cost and high access benefits. All this information sharing and retrieval should be highly secured with utmost care to the data and should provide confidence to the data owner. But this gives rise to the limitation of the utilization of data through plain text search. Hence an excellent methodology is required to match the keywords with encrypted cloud data. The proposed approach similarity measure of "coordinate matching" combined with "inner product similarity" quantitatively evaluates and matches all relevant data with search keyword to arrive at best results. This approach, each document is associated with a binary vector to represent a keyword contained in the document. The search keyword is also described as a binary vector, so the similarity could be exactly measured by the inner product of the query vector with the data vector. The inner product computation and the two multi-keyword ranked search over encrypted data (MRSE) schemes ensures data privacy and provides detailed information about the dynamic operation on the data set and index and hence improves the search experience of the user.

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

1. INTRODUCTION

Cloud computing, which holds the characteristics of intrinsic data sharing and low maintenance, provides a better utilization of resources. (Wan, 2012) The term Cloud refers to a Network or net (Jung, 2013). In alternative words, Cloud is some things that are gift at remote location (Zhongma Zhu, 2016). Cloud will give services over network. Service Models are the reference models on that the Cloud Computing relies (Yang, 2013). These may be classified into three basic service models as listed below: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) (Ostrovsky, 2014). There is a unit several alternative service models all of which may take the shape like anything as a Service. (Kamara, 2010) This can be Network as a Service, Business as a Service, Identity as a Service, information as a Service or Strategy as a Service (Waters, 2008). The Infrastructure as a Service (IaaS) is the simplest level of service (Yang, 2013). In cloud computing, cloud service providers offer an abstraction of infinite storage space for clients for hosting required data. (Armbrust, 2010). It may help clients reduce their financial overhead of data managements by migrating the local managements system into cloud servers. (Delerablee, 2007). The planned approach similarity live of "coordinate matching" combined with "inner product similarity" quantitatively evaluates and matches all relevant information with search keyword to make best results (Delerablee, 2007). Then that user can able to upload the same document with changes in that document that document modified words are updated in the individual page (Ren, 2013). But obviously security constraints must be most importantly concentrated because as we now outsource the storage of data, which is possibly sensitive, to cloud providers (Kamara, 2010). Hence to preserve data privacy, a common known approach is to encrypt data files on or before the clients upload the encrypted data into the cloud (Ateniese, 2005). Unfortunately, it is very difficult to design a secure and efficient data sharing scheme and especially for dynamic groups in the cloud (Goyal, 2006). Hence, in this scheme we have introduced Multi-keyword ranked search and inner product computation for data privacy (Boneh, 2004). These two schemes are also responsible for providing detailed information of the dynamic operation and automatically improves the search experience for the users (Boneh, 2005). Achieving a secure, scalable and fine grained data access control in cloud computing is the main goal and which is achieved by exploiting and uniquely combining techniques of attribute based decryption(ABE), proxy re-encryption and lazy re-encryption. This system uses the cryptographic primitives applied to the problem of secure storage in the presence of untrusted servers and desire for owner managed key distribution. Eliminating almost key distribution in the hands of individual data owners provides a basis for a secure storage system. A secure file system was designed to be layered over

An insecure network and P2P file systems such as CIFS, NFS, Yahoo! And ocean store. Key management and revocation is simple with the minimal out of band communication. A traitor tracing mechanism that can be integrated with any form of subset cover revocation scheme that satisfies a bifurcation property. This mechanism does not need a prior bound on the number of traitors and does not expand the message length compared to the revocation of the same set of traitors.

In the existing system, the documents and data is stored into a secure cloud storage whereas the documents were scanned with the keywords and an index of information were stored for future searching options. Existing system will have a very simple document based search system.The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results. Symmetric encryption algorithm by

stop word concept the unwanted keywords will be removed. The document search by name not by content. So we get relevant information and irrelevant information. Encrypted information were stored securely in cloud and we don't have an option of hash table to access the data faster and effectively. Ranked search ultimately enhances system usability by enabling the 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. This existing system consists of some demerits, they are no semantic based retrieval is available (i.e.,) the word with the related meaning cannot be retrieved. Then there is 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.This system does not consists of Single-keyword search with ranking and Boolean- keyword search with ranking.

Proposed Scheme: The system is defined to solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and tries to establish a set of privacy requirements for such a secure cloud data utilization system to become to reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching". In proposed system define public or private page and will be stored. Individual page updation is in this system. The document (abc.doc) is ranked by multi key word concept. Checksum value for each page. 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 huge knowledge repos tire.

Design Goals:

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. So that, the data will be accessed at a very high speed (On a whole a high end intelligent data system is built in cloud).

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 updation can be achieved via Whatsapp or SMS to the end users.

Intelligence Efficiency- Semantic information were stored and the data will be retrieved accordingly in a semantic manner.

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, f_1, End ()). Here, the group manager keeps the parameters as the secret master key.

Registration for existing user: First step, the user sends the ID_i (the identity of the user i) 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 $ID_{m+1.}$

File download: The operation is performed by the group members and the cloud, the group member encrypts the ID_{data} 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. On receiving the document the user decrypts and gets the original file.

System Architecture: The architecture and entire processing of the proposed system is explained in the Figure.1.

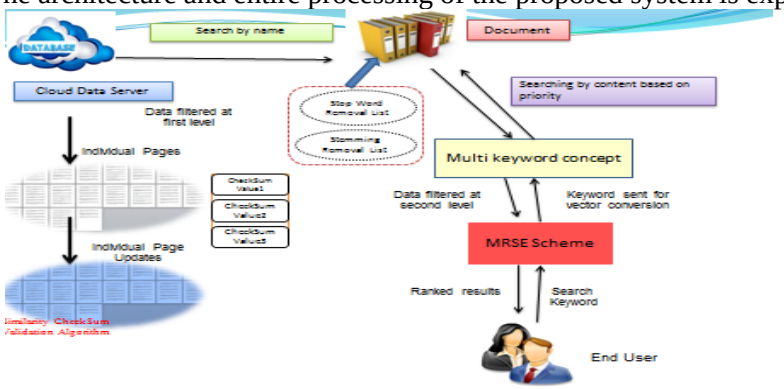

Figure.1. Intelligent keyword searching system

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. When the user is authorized person means document is searched .User's information is searched by using name and password in a database. Both the name and password is matching with the database then only documents are searched in a database. User is unauthorized person, documents not searched and it will produce errors.

Name Search: After the verification of user's information in a Database. User is allowed to search a document using document name only. Each and every word is sorted and produce output based on name search in a database. 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)

Individual Page Updation: Once the user checks about documents they are getting contents of various pages as a output .User wants to do the updation in particular page in a document, each page searched in a database and particular page is sorted out first and it is modified by the user. Those modification of particular page is done by using check sum technique in a database. It checks the pages and giving priority to the content.

Priority Based Search: User searched the document by using multi-keywords and finally outputs produced based on large number contents. Each and every content in a page is analyzed in a document and priority given to huge pages in a database. Priority wise pages are sorted in this module and cost is reduced because of using this technique in a cloud server. Document is searched by name of document. And outputs produced based on the multi-keyword search in a document. Finally, documents saved in a database using public or private methodology. When the document is saved as public, it's related to everyone's view. Second method, it is saved as private produces output related to authorized user view only.

Performance Evaluation: The performance evaluation process is done with the help of NS2 and compared with the original dynamic broadcast encryption (ODBE). With this features we can also measure the size of the file uploaded or downloaded. In figure.2 and figure.3, consists of the comparison between computation cost of members for uploading with proposed scheme, MONA, ODBE and RBAC.

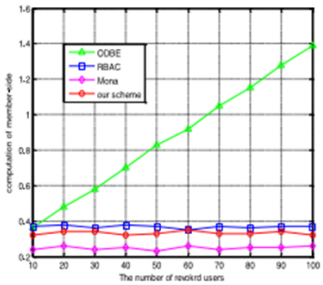

Figure.2. Uploading 10 Mb File

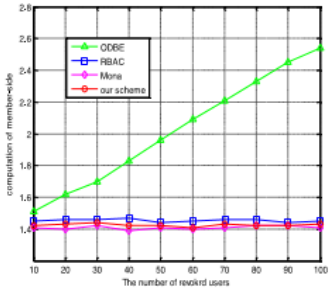

Figure.3. Uploading A 100 Mb File

In figure.4 and figure.5, consists of the comparison between computation costs of members for file downloading with the proposed scheme, MONA, ODBE AND RBAC.

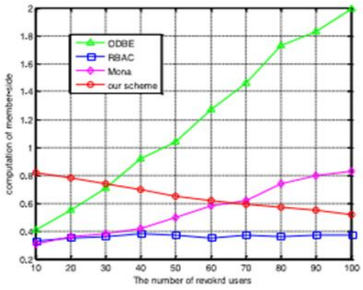

Figure.4. Downloading 10 Mb File

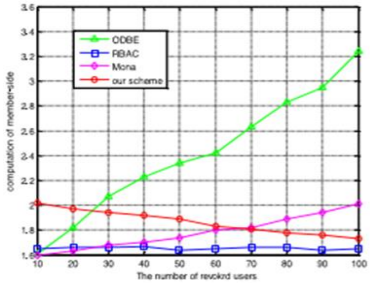

Figure.5. Downloading 100 Mb File

2. CONCLUSION

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 Damage of Data Transmission, We proposed one Concept. In order to overcome that, We Proposed a Technique to Transferring the Image or Video form Source to Destination Without any Loss of Data and Leakage of Data.

Future Enhancement: In future enhancement, the user get a 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.

REFERENCS

Armbrust M, Fox A, Griffith R, Joseph A.D, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I and Zaharia M, A view of cloud computing, Commun. ACM, 53 (4), 2010, 50–58.

Ateniese G, Fu K, Green M and Hohenberger S, Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage, Proc. Network and Distributed Systems Security Symp. (NDSS), 2005, 29-43.

Boneh D, Boyen X and Goh E, Hierarchical identity based encryption with constant size ciphertext, in Proc. Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2005, 440–456.

Boneh D, Boyen X and Shacham H, Short group signature, in Proc. Int. Cryptology Conf. Adv. Cryptology, 2004, 41–55.

Delerablee C, Paillier P and Pointcheval D, Fully collusion secure dynamic broadcast encryption with constant-size Ci-pher- texts or decryption keys, in Proc. 1st Int. Conf. Pairing-Based Cryp- tography, 2007, 39–59.

Goh E, Shacham H, Modadugu N and Boneh D, Sirius, Securing Remote Untrusted Storage, Proc. Network and Distributed Systems Security Symp. (NDSS), 2003, 131-145.

Goyal V, Pandey O, Sahai A and Waters B, Attribute-based encryption for fine-grained access control of encrypted data, in Proc. ACM Conf. Comput. Commun. Security, 2006, 89–98.

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.

Kallahalla M, Riedel E, Swaminathan R, Wang Q and Fu K, Plutus: Scalable Secure File Sharing on Untrusted Storage, Proc. USENIX Conf. File and Storage Technologies, 2003, 29-42.

Kamara S and Lauter K, Cryptographic cloud storage, in Proc. Int. Conf. Financial Cryptography Data Security, 2010, 136–149.

Kamara S and Lauter K, Cryptographic cloud storage, in Proc.14th Financial Cryptography Data Security, 2010, 136–149.

Naor D, Naor M and Lotspiech J.B, Revocation and Tracing Schemes for Stateless Receivers, Proc. Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO), 2001, 41- 62.

Ostrovsky R, Sahai A and Waters B, Attribute-based encryp- tion with non-monotonic access structures, in Proc. 14th ACM Conf. Comput. Commun. Security, 2014, 195–203.

Pointcheval D and Stern J, Security Arguments for Digital Signatures and Blind Signatures, J. Cryptology, 13 (3), 2000, 361-396.

Ren K, Wang C and Wang Q, Security challenges for the public cloud, IEEE Internet Comput., 16 (1), 2012, 69–73.

Wan Z, Liu J and Deng R, Hasbe, A hierarchical attribute- based solution for flexible and scalable access control in cloud computing," IEEE Trans. Inf. Forensics Security, 7 (2), 2012, 743–754.

Waters B, Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization, in Proc. Int. Conf. Practice Theory Public Key Cryptography Conf. Public Key Cryptography, 2008, 53–70.

Yang K and Jia X, Expressive, efficient and revocable data access control for multi-authority cloud storage, IEEE Trans. Parallel Distrib. Syst., 25 (7), 2013, 1735–1744.

Yang K, Jia X and Ren K, DAC-MACS, Effective data access control for multi-authority cloud storage systems, in Proc. 32nd IEEE Int. Conf. Comput. Commun, 2013, 2895–2903.

Zhongma Zhu, Rui Jiang, A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud, IEEE transactions on parallel and distributed systems, 27 (1), 2016.