Secured Data Processing over Encrypted Cloud by Ranked Multi Keyword Search

R. S. Karthiga, U. V. Anbazhagu, J. Senthilkumar
Department of CSE, VISTAS, Chennai, India
rskarthigacse@gmail.com
anbazhagu.se@velsuniv.ac.in
skumar.se@velsuniv.ac.in

Abstract: Cloud computing is the usage of remote servers on the web to collect, organize and process document rather than a local server or our individual system. In cloud computing enterprise pay to access the web application and/or server service rather than buying or installing software or hardware. Large numbers of document proprietors are outsource their documents into cloud system. Aimed at privacy concern, delicate documents such as (email, password, health records, government officials and so on) are encoded in advance of outsourcing. Old-fashioned search encryption procedures are used only single keyword or Boolean keyword for searching. In the suggested method, we are implementing Multi keyword ranked search over encoded cloud document techniques to access the document efficiently by greedy search algorithm then preserve privacy using encryption technique AA-BSE. In this paper, documents are divided into more chunks, then encoded it and stored the data into different cloud server.

Keywords: Multi keyword search, Indexing, SHA512, AABSE, Encrypted cloud data.

I.INTRODUCTION

Cloud computing is the usage of computing reservoir that are provided as a facility over a web. Old fashioned document consumption facility centred on plaintext keyword search. The slight result of transferring the entire document and decoded locally is undoubtedly impossible because of enormous quantity of data transfer rate in the cloud system. There are no purposes of storage management and loading documents into cloud server unless the documents can be effectively searched and used efficiently. On account of huge amount of on demand document consumers and enormous quantity of data in the cloud, this problematic is main challenging as it is very tuff to achieve also the requirement of performance, system usability and scalability.

Thus, discovering confidentiality maintaining and efficient search service over in the cloud and beyond, delicate document, for example, mails, individual fitness histories, pictures, tax papers, business dealings, and so on, may have to be encoded by document proprietors in advance of outsourcing to the delicate documents in the cloud. To achieve the efficient data access need, the enormous quantity of files claims the cloud system to accomplish outcome relevance ranking, rather than giving undistinguishable outcomes. Such ranked search method allows document consumers to catch the utmost related data faster. Ranked search can also smartly reject unwanted network movement by returning only utmost appropriate document. For confidentiality security, such ranking process, nevertheless, never disclosure any keyword associated data.

In spite of the many benefits in cloud services, data owners outsource their sensitive document to cloud server for confidentiality. Only authorized user can access the sensitive documents. A common method for confidentiality security, data owners encrypt their document before outsourcing into cloud server. In this analysis, we statement the difficult of confidentiality preservation using multiple keywords search over encrypted cloud data. Our methodology offers multi-keyword search based on synchronize matching.

If document is not encrypted, it can be easily accessed by unapproved user. By encryption technique, we can hide the contents so it can be easily avoid unauthorized user. Encryption is the method of converting original information into coded format. The encrypted data can be accessed only by the user those has

accessed key. So, it can avoid unauthorized user. Single-Keyword search allow the user to search only actual single word not its variants.

In multi-keyword search, User search and lists many variants of same keyword. We select the effective principle of synchronize matching among the multi keyword semantics i.e. to find and catch similarities between query search and documents.

II. OBJECTIVES

- To facilitate effective and protected multi keyword search above encoded cloud document
- To achieve the efficient document access need, the enormous quantity of files claims the cloud system
- To accomplish outcome related ranking, rather than giving undistinguishable outcomes

III. SCOPE

- It helps in multi keyword ranked search
- It helps in searching efficiency
- It is used for confidential and privacy preserving

IV. SUMMARY FROM LITERATURE SURVEY

A. Search Encryption Techniques

It is a cryptographic method that permits the consumer to retrieve the document in encoded form efficiently.

B. Boolean Symmetric Search Encryption Techniques

This method (BSSE) utilizes the Boolean query search to accomplish requesting information; they are constituted of combination, disjunction and negotiation of keywords. In this technique, the major operation is the Gram Schmidt orthogonalization process. First operation in this method, query keywords are encrypted and labels and inside content to execute the searching of documents.

BSSE is completely design less, it allow linear search. So computation stage for every document Increases due to growth of label size. But it is the only effort focus on simple keyword matching and used only for Boolean searching queries.

C. Term Frequency and its Inverse Document Frequency Model

The author proposed an approach consolidate inverted index with novel privacy preserving searching. This model is used to calculate the score of numerical relevance to retrieve the documents. It saves communication overhead and system usability. This result only provides single keyword ranked search.

The privacy-preserving solutions offer benefits to mutually edge devices and facility suppliers; on the other hand they bring together computational and communicational overhead. Hence, suggested privacy-preserving methods should encounter both confidentiality necessities as well as performance.

D. Searchable Symmetric Encryption Techniques

Data owners are searching and retrieving their outsource data by SSE. It is used for efficient and security searching of cloud data. The main disadvantages of SSE techniques are large quantity of document and cannot accommodate high level requirements.

E. Fuzzy Keyword Search Techniques

Another name of Fuzzy keyword searching techniques is known as wild card based technique. When data owners searching the data accurately match the predefined keywords gives the corresponding data or the nearest probable identical data based on similarity keyword semantics, when accurate match fail. To minimize the storage capacity and representing overhead, this technique develop an effective method to build fuzzy keyword collections and edit space to Quantity keyword resemblance. The main disadvantages are enormous loading is difficult, Provision only Boolean keyword search and not provide the ranked search problem.

F. Community Key Encryption with Keyword Search

In this system, cloud system holds encrypted file and keyword indexes. By using its private key user create trap door. The cloud server checks the trap door

with existing encrypted keyword and retrieve encrypted file that match it.

G. Order Preserving Mapping Techniques

Order preserving mapping technique is used to search encrypted multi cloud data. It is used to preserve sensitive information such as email, government official records, health records, password, personal photos etc. It is more efficient techniques. But this technique causes lots of collisions in the networks. After main searching of data, encrypted files are processed. Many numbers of encrypted files are post processed.

H. One Many Order Preserving Mapping

One many orders preserving mapping is mainly used for multi ranked search of data, preserve security and upgrade the performances. The most benefits are less communication, decrease computation overhead, avoids network traffic and undesirable data retrieval. It has some disadvantages does not provide multi keyword. It increase the searching time and cost.

I. Pseudo Random Generator and Sequential Scan Technique

This technique provides high security for searching data. It is very secure, not complex, and practical and speed. No space and communication overhead. It is adaptable. The main disadvantage of sequential scan is inefficient for huge amount data. It is not fast in searching huge amount of data.

V. PROBLEM STATEMENT

Cloud computing supports cloud clients to remotely accumulate their document into the cloud so as to enjoy the on demand data user request and facilities. Cloud services permit the customer to retrieve data from any system, as long as it is linked to the web.

On account of one data owner in existing model, by denoting this result data proprietor and data consumer can communicate easily and interchange sensitive data. When many data proprietors are participate in the cloud, exchange of sensitive data will reason for huge communication overhead. We discover the difficult of secure multi keyword search for various data proprietors and many data consumer in cloud system.

VI.PROPOSED SYSTEM

We propose to enable multi keyword search for an encrypted document by two step process. Document will be segregated as multi part and multi keyword hash values will be added to individual document hash values. Multi Keyword search will improve accuracy of search retrievals. Whole and partial retrieval is possible. Unnecessary high network traffic will be avoided.

VII. LIST OF MODULE

- Literature Review
- Document Segregation to multipart
- Hash Automatic multi keyword
- Attribute Based Encryption
- Document Search
- Partial Document Retrieval
- Full Document Retrieval
- Conclusion

VIII. M ODULE DESCRIPTION

System Architecture is given in the flowchart 1

A. Document Segregation to Multi Part

Multi part document is majorly used in health care where patient's only partial record may be needed for administration. Hence multi part document was introduced. Major advantage of multi part document is increased security of document.

Along with this enables option to retrieve partial document. It is observed that partial document retrieval is very fast. In our case document will be split into multiple parts based on the size of the document. In document based size method, for segregate documents into multiple parts data owners can set the specified file size. Then documents can divided into multi parts it's not higher than a definite document size .Later keywords that are generated automatically will be assigned for the

respective parts. Few users may have hard to downloading very huge amount of documents from cloud.

B. Hash automatic multi keyword

Generally authors provide keyword for the whole document. Along with splitting the document into multi parts, we provide an additional feature of extracting keywords of the individual part automatically. This will help in enabling search facility for author missed keywords. We extract the keywords automatically and hash the keywords using SHA512 algorithm.

Data owners can create an index for each document. Documents linked with the index values or keyword is to be accessed. Index key for every document is more efficient to search all the documents until the corresponding document is found.

Creating hash value for retrieving documents or for preserving privacy. For privacy preserving, confirm that sending information have not been modified.

C. Attribute Based Encryption

In this encryption method, User identification is depending upon collection of descriptive attributes. User allows decrypting the data; only their identity attributes must satisfy any one of the specified attributes in the cipher text.

Attributes of the user are collected during registration. These attributes will be used for encryption and decryption mechanism. It is a kind of public key encryption. The cipher text and user secret key depend upon the attributes. A fundamental security feature of this method is collusion-resistance: An opponent that detains multiple keys should only be able to retrieve data if minimum one single key allow access.

C. User Access Check

User access check is primary for the document search. User's attribute is validated for search function. Enabling multi part document's major advantage is reducing un-necessary data download in case of very large documents. This will help in downloading requiring data and downloading the data very fast.

Data user asks permission to retrieve required documents in the cloud. Permission accepted or disapproved depends upon attributes of the user, access policy set by the data owner. Only if the user is allowed to search for documents, he can proceed further.

D. Partial Document Search

Since our documents are split and saved, we provide user an option to search for complete or partial document search. It permit data users to define a particular part of the whole document they are needed for.

When the user searches for multi keywords, all the keywords will be split hashed and validated against our indexes. When matched respective document will be retrieved.

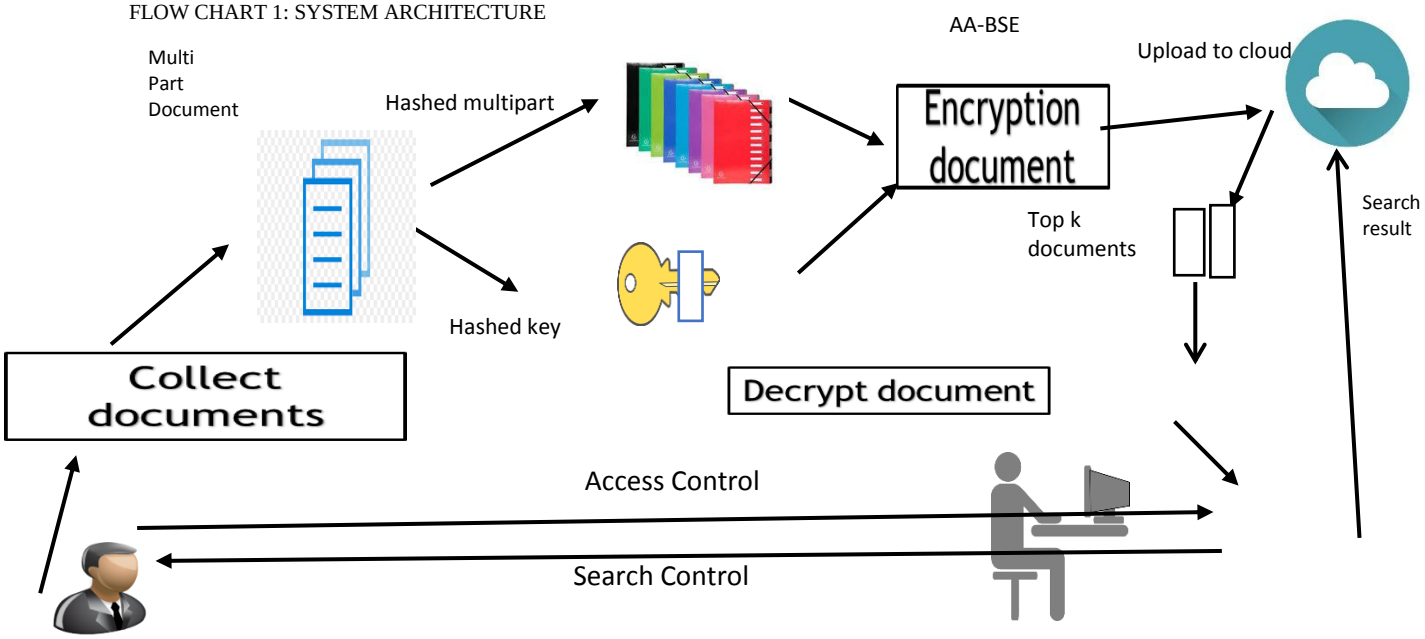
E. Whole Document Search

If the user decides to go with whole document search, he can still search the whole document. In this case the advantage we get is from automated key word generation process. This enables document retrieval from keywords what the author has missed.

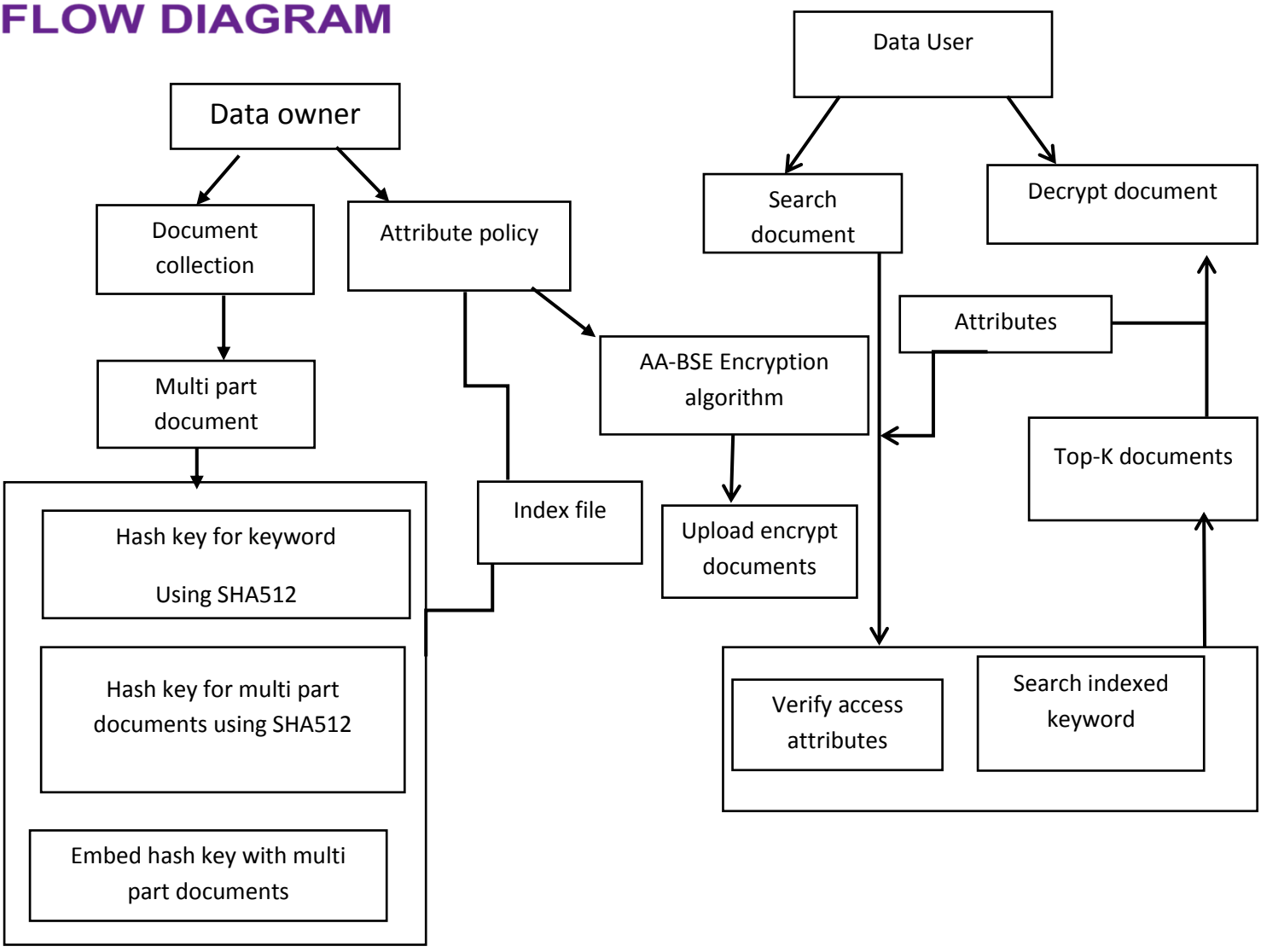
IX.CONCLUSION

In the suggested system data proprietors accumulate their encrypted data on the cloud. By using multiple keyword, authorized data user permit to search requested data. Document is split into multi part and automatic keywords are generated to improve accuracy and quicker document retrieval. TheSuggested multi keyword search over encrypted cloud data allows customers to attain protected and effective searches over multiple data owner's document.

FLOW CHART 1: SYSTEM ARCHITECTURE



FLOW DIAGRAM



REFERENCES

[1] TarikMoataz, AbdullatifShikfa, "Boolean Symmetric Searchable Encryption", ASIA CCS '13 Proceedings of the 8th ACM SIGSAC symposium on Information computer and communications security, .pp. 265-276, NY, USA , 2013.

[2] TianyuePeng, Student Member, IEEE, Yaping Lin, Member, IEEE, Xin Yao, Student Member, IEEE, and Wei Zhang

[3] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions an efficient constructions" ACM CCS 06[20] conference 2006.

[4] Jin Li,Qian Wang ; Cong Wang , Ning Cao , KuiRen , Wenjing Lou "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing" INFOCOM, 2010 Proceedings IEEE March 2010.

[5] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data" journal IEEE VOL:23 ISSUE:8 2012

[6] Cong Wang, Ning Cao, Jin Li, KuiRen, Wenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data" Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference 2010.

[7] Bing Wang, Wei Song, Wenjing Lou, Y. Thomas Hou, "Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee", 2015 IEEE Conference on Computer Communications(INFOCOM).

[8] DawXiaodingSong , Wagner, D. , Perrig A. "Practical Techniques for Searches on Encrypted Data" Security and Privacy, 2000. S&P 2000.Proceedings. 2000 IEEE Symposium May 2000

[9] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, "Verifiable PrivacyPreserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 11, November 2014.