Privacy Preservation in Multikey Word Based Text Retrieval in Cloud

[1]S.Balaji, [2]M.Komathi, [3]K.Karthika
[1]Assistant Professor, [23]UG Student,
Department of Computer Science and Engineering
GRT Institute of Engineering and Technology, TamilNadu, India.

Abstract

In recent times Mobile cloud computing has drawn attention of the researchers. In mobile cloud computing to attain efficient data storage it is possible to outsource the data. To ensure confidentiality and integrity of data we need to encrypt and send the data. While the data is secured it becomes complicated. To prevail over the challenge a multi keyword search in cloud is proposed. NLP based text mining process is involved to extract the multi keywords association and the encrypted data is split into blocks to store in the cloud storage. This increases the privacy of the out sourced data and implies unauthorized access is not viable.

Keywords: *Encryption, Natural Language Processing, Mining.*

I. INTRODUCTION

Research in cloud computing is receiving a lot of attention from both academic and industrial worlds. In cloud computing, users can outsource their computation and storage to servers using Internet. Clouds can provide several types of services like applications, infrastructures and platforms. Examples for application services are Google Apps, Microsoft online, infrastructures, Examples for infrastructure services Amazon’s EC2, Eucalyptus, Nimbus and Examples for platform services platform to help developers write application Amazon’s S3, Windows Azure. Much of the data stored in clouds is highly sensitive, for example, medical records and social networks. Security and privacy are thus very important issues in cloud computing. In one hand, the user should authenticate itself before initiating any transaction, and on the other hand, it must be ensured that the cloud does not tamper with the data that is outsourced.

In order to search in cloud, some requirements is needed, search over encrypted data should support the following three functions. First, the searchable encryption schemes should support multi-keyword search, and provide the same user experience as searching in Google search with different keywords; single-keyword search is far from satisfactory by only returning very limited and inaccurate search results. Second, to quickly identify most relevant results, the search user would typically prefer cloud servers to sort the returned search results in a relevance-based order ranked by the relevance of the search request to the documents.

Cloud storage systems driven by high speed networks and large data centre provide reliable storage services over the Internet. A user can store his files in a cloud storage service and user can access files via only the Internet .user is concerned about the security of the stored files. Data confidentiality and data robustness are the main security issues for cloud storage. For data confidentiality, the user can first encrypt files, and then store the encrypted files in cloud storage and the files in the cloud are confidential against not only outsiders, but also the cloud service provider .For data robustness, there are two concerns: service failure, and service corruption. Service failure is when the user does not get responses for his retrieval requests, and service corruption is when the user gets corrupted data.

In the existing system, the main disadvantage is that unauthorized person can also access the data. So, the data is not secured. In order to overcome the above disadvantage, we propose an authentication and access control. In authentication process, only the legitimate user or authorized person can access the data and in access control we use asymmetric kind of encryption (we use two keys i.e private key and public key) to protect the data more secured.

II. RELATED WORK

The proposed system is for efficient information retrieval over query ranked for cloud storage. The cloud computing having two primary issues: 1.privacy 2.efficiency. To overcome these issues they used ADL for private keyword-based file retrieval scheme, efficient information retrieval for ranked query (EIRQ) and reduce the query cost. The disadvantages of this paper is trust the third party. For our future enhancement, we will explore an extension of our solution that would apply to the case where we don’t need to trust the ADL.[1]

Cloud computing is used to store and access their data from the cloud services over the internet. The proposed system they implement the Single Keyword Search Over Encrypted Data And then Multi-keyword Ranked Search over Encrypted cloud data (MRSE).It is based on secure inner product computation and efficient similarity measure of coordinate matching and to achieve various stringent privacy requirements in two different threat models: Assignment of anonymous ID to the user to provide more security and improve the data search service experience. In Future work we will checking the whole integrity of the rank order in the search result assuming the cloud server is un trusted.[2]

As they are proposed on an e-health monitoring system with minimum service delay and privacy preservation by exploiting geo-distributed clouds in using a traffic-shaping algorithm. This algorithm is used converts the user health data traffic to the non health data traffic to reduce traffic analysis .The distributed cloud servers assign the servers to the requested users under the load balance condition is based on resource allocation .The service delay and privacy preservation is efficiently improved through the numerical analysis. Future work, studying a more general and complicated case where users have random medical requests and diverse privacy preservation requirements.[3]

The project defines and solves the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wise privacy in the cloud computing paradigm. The System "Multi-keyword Ranked Search with Fault Tolerance" Manager has been implemented which work on encrypted cloud data, which provide users many specialized services so that the system can satisfied all special requirements of users. The third party will be access the information without any permission to data owner System which gives alert sending message (SMS) from Mobile Server on user's mobile. . But the best part is only data which is visible to hacker on cloud is updated, original data is not updated. The disadvantages of this paper is low overhead on both computation and communication.[4]

As they are proposed on secure and efficient multi-keyword similarity searchable encryption (MKSim) that return matching document based on query search .These approach is highly efficient and ready to be deployed in the real-world cloud storage system. The solution is based on Term Frequency - Inverse Document Frequency(TF-IDF) measurement and ring-LWE technique As future work, To optimize the index construction algorithm and continue to research on secure mechanisms for the effective utilization over outsourced cloud data.[5]

Mobile cloud computing is a technique that shifts the data and computing service modules from individual devices to a geo graphically distributed cloud service architecture. They proposed a service decision making system for inter-domain service which is used to transfer and balance the computation which loads among multiple cloud domains. Because to efficiently manage the cloud resources across multiple cloud domains and it will be critical in providing continuous mobile services and to end that they formulated the service request decision making process as a semi-Markov decision process. They developed an SDMP-based computing model for inter domain services for communication cost and to control the expenses of computing resources. Future work is analyzing the optimal system that will maximize system rewards for large-scale system.[6]

To store a dynamic collection of encrypted documents Dynamic Searchable Symmetric Encryption is used. In this paper they presented a dynamic SSE scheme it is simpler and efficient than existing. They implemented a prototype for efficiency on datasets they demonstrated from the prior work. Their scheme is simpler especially it support only the upload and download operation to the server. In their server scheme based on cloud storage service rather cloud computation. In building dynamic SSE scheme they introduced a Blind Storage, while retrieving the files it doesn't reveal the files name and contents it just learns about its existence. Future work is to secure the scheme against the corrupted server using multiple keywords.[7]

To improve the search efficiency they proposed a tree based index structure and various adaptive methods for multi dimensional algorithm. Encryption helps to protect the user data, yet practically-efficient secure function over encrypted data is a challenging problem. So they presented a privacy-preserving multi keyword text search (MTS) with similarity based ranking to address this problem. For further enhancement of search privacy, two secure index schemes i.e., are cipher text model and background model must be introduced to boost the search privacy. [8]

Multi-Keyword fuzzy search is used to cover the possible keyword misspelling which lead to larger index file size and higher search complexity. They proposed a scheme which achieves fuzzy matching through algorithm design rather than expanding index file. Without increasing the index or search complexity this scheme supports effectively multiple keyword fuzzy search. This scheme is first mainly used to achieve the multi-keyword fuzzy search over encrypted cloud data. Disadvantage: In this paper they mentioned about only about the search process rather than the storage.[9]

In cloud computing search over encrypted data is very important technique. To protect user data privacy in untrusted cloud so they are using the secure search schemes. This scheme is mainly focusing on the single contributor scenario. In this paper they proposed on focusing the different scenario that can be contributed from multiple owners and are searchable by multiple users. Attribute-Based

Encryption (ABE) and attribute based keyword search user revocation (ABKS-UR) that enables the fine grained search authorization. This search authorization is also implemented by the owner-enforced access policy on index of each file.[10]

III. PROPOSED SYSTEM

In the proposed system, we introduce an authentication and access control to the users for privacy preserving and to secure the data. In existing system there is no authentication and access control scheme to provide secured data to the users. The outsourced data is directly stored in cloud hence the data may be used by the third party users and some issues like data confidentiality and lack of privacy will occur. These issues will be overcome in our proposed system by using Asymmetric Encryption algorithm.

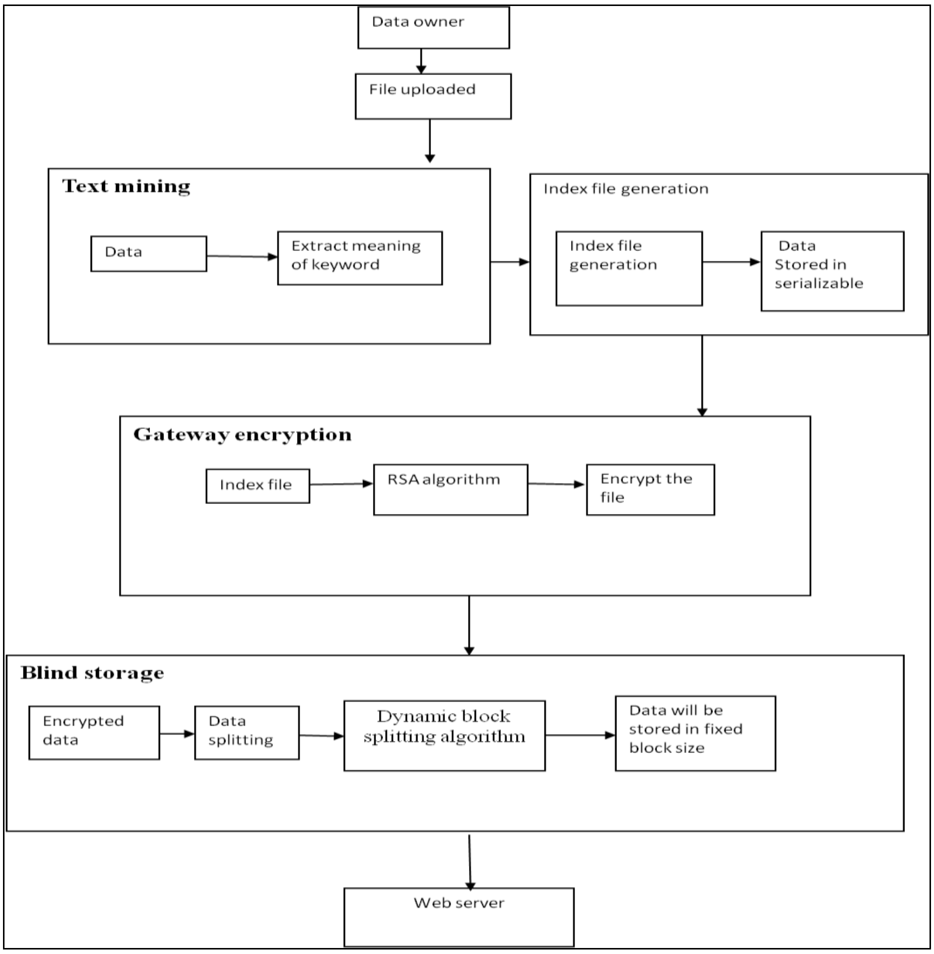

Fig 1.1 Architecture for Privacy Preserved Cloud Storage

The main aim of this project is to preserve the outsourced data in cloud through gateway encryption and blind storage, and to implement multi keyword ranked search over the encrypted data in a secure way by NLP process without downloading and decrypting the entire group member file contents.

Firstly, Data owner should register in cloud and creating a group for data users to maintain the privacy of user data. Data user should register and must give the request to the data owner. If the data users want to read the content of files, the data user should request to the cloud and then cloud will request to data owner. Data owner checks the data user attributes and status of access control. After this process the user can access the data in secured manner. Then the group owner must accept the request from user. After accepting the requests from the data users then they can able to access the documents, files, etc and data owner provide public key and private key to data user using RSA and BASE64 algorithm. Data user cannot access the data or information from the web page until the data owner accepts the request. Multiple groups can be created. Each group is having owner and users.

Data owner can upload the file using the text mining process. After uploading the file from the group owner then the index file will be generated and it will be saved as a serializable object in the cloud. In the text mining process there are two techniques: NLP technique and Word Net Tool. NLP technique is used to extract the content which has been uploaded. The extracted word from the uploaded file using the NLP technique and that word get synonyms form the

Word Net Tool. The In this process using the software requirements POS tagger is implemented in the keyword in files. All the communication to cloud server will be done through Web Service.

After completing the NLP process the uploaded files are encrypted in gateway and it will be stored in index file. The owner can give access control and privileges to the user while uploading the data. Access control refers to whether the user has permission to access the file or not. The privilege refers to how much extend that the user has rights over the data (read and write). The file will be split into blocks and its encrypted using RSA encryption algorithm and the encrypted blocks will be uploaded to the cloud service and stored in blind storage.

Blind storage is used to divide the uploaded file into fixed size blocks. These blocks are indexed by the sequence of random integers. Files content are stored in block randomly so the cloud can view encrypted content only. Encryption key only knows to data owner. By splitting the uploaded files into blocks then no one can view the files easily.

Data user will try to search a file in cloud server. The cloud servers map the keywords and search the related files. The cloud server gives the related filename to user. The user should click the filename to view the content. At that time user request to cloud server and server send the user details and filename to the data owner. Then data owner knows all public key of user so he encrypt the private key using data user public key and the encrypted key send to server and server send that key details to user, then user decrypt the key using our private key. After that the data user can get private key of data owner and then access the data through blind storage.

Hence we developed an efficient search in multi keyword through blind storage which enable accurate, efficient and secure search over encrypted data. Privacy is preserved for data in cloud while storing in blind storage and also achieved access control for each user.

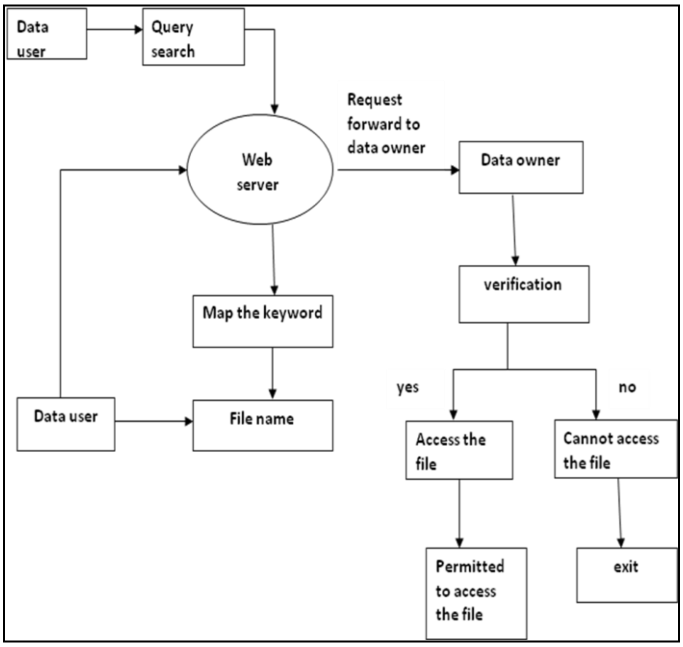

Fig 1.1 User Query Processing to Access File from Cloud

IV. CONCLUSION

Architecture for efficient data out sourcing in mobile data cloud is proposed. Encrypting the data ensures the security of the uploaded data at the same time multi keyword search is supported by text mining process. The relevance of text mining is planned to be extended with fuzzy logic. The data splitting ensures even if a attack is primed towards the data it is not possible to access the entire content since different data blocks are stored in various locations of cloud. Future work is intended on insisting a improvised homomorphic type encryption for efficient keyword search with high security.

REFERENCES

[1] Q. Liu, C. C. Tan, J. Wu, and G. Wang, ``Effcient information retrieval for ranked queries in cost-effective cloud environments," in Proc. IEEE INFOCOM, Mar. 2012, pp2581_2585.

[2] Madan S.A. ,B.M. Patil," Comparison of Privacy Preserving Single-Keyword Search and Multi-Keyword Ranked Search Techniques over Encrypted Cloud Data,"

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Y. Luo, ``Exploiting geo distributed clouds for a e-health monitoring system with minimum service delay and privacy preservation," IEEE J. Biomed. Health Inform., vol. 18, no. 2, pp. 430_439, Mar. 2014.

[4] Kalika Kambale1, Swati Barsagade2,Nilima Bhujbal3,Swati Khilare4,"Multi-Keyword Ranked Search over Encrypted Cloud Data Manager",vol.1,no.58,Mar.2015.

[5] W. Sun, et al., ``Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2013, pp. 71_82.

[6] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, ``An SMDP-based service model for inter domain resource allocation in mobile cloud networks," IEEE Trans. Veh. Technol., vol. 61, no. 5, pp. 2222_2232, Jun. 2012.

[7] M. Naveed, M. Prabhakaran, and C. A. Gunter, ``Dynamic searchable encryption via blind storage," in Proc. IEEE Symp. Secur. Privacy, May 2014, pp. 639_654.

[8] W. Sun, et al., ``Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2013, pp. 71_82.

[9] B.Wang, S.Yu,W. Lou, and Y. T. Hou, ``Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112_2120.

[10] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, ``Protecting your right: Attribute-based keyword search with _ne-grained owner-enforced search authorization in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 226_234.