Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/2333
DOI URL: http://dx.doi.org/10.21474/IJAR01/2333
ISSN NO. 2320-5407

RESEARCH ARTICLE

PRIVACY PRESERVING NATURAL LANGUAGE PROCESSING IN THE CLOUD SUPPORTING SIMILARITY BASED TEXT RETRIEVAL THROUGH BLIND STORAGE.

T. Thilagam.
Assistant Professor Deparment Of Computer Science Engineering Gojan School Of Business And Technology

Manuscript Info

Manuscript History

Received: 30 September 2016
Final Accepted: 30 October 2016
Published: November 2016

Key words:-
Cloud computing, searchable encryption, multi-keyword ranked search, blind storage, access pattern.

Abstract

In cloud computing, a fundamental application is to preserve 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 .By using NLP (Natural language processing) technique used to search multi keyword in cloud its extract the meaning from Word Net tool. In this paper, we develop the searchable encryption for multi-keyword ranked search over the storage data. Efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. Within this framework, we leverage an efficient index to further improve the search efficiency, and adopt the blind storage system to conceal access pattern of the search user. Security analysis demonstrates that our scheme can achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. Finally, using extensive simulations, we show that our proposal can achieve much improved efficiency in terms of authentication and access control compared with the existing proposals.

Introduction:-

Research in cloud computing is receiving a lot of attention from both academic and industrial worlds. In cloud computing, user can outsource their computation and storage to servers (also called clouds) using Internet. Clouds can provide serveral types of services like applications (e.g., Google Apps, Microsoft online), infrastuctures (e.g.,Amazon'sEC2,Euclyptus,Nimbus),and platforms to help developers write applications (e.g., 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 In order to 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. our main contributions can be summarized as follows:

Corresponding Author:- T. Thilagam..
Address:- Assistant Professor Deparment Of Computer Science Engineering Gojan School Of Business And Technology.

- We introduce a relevance score in searchable encryption to achieve multi-keyword ranked search over the encrypted cloud data. In addition to that, we construct an efficient index to improve the search sufficiency
- By modifying the blind storage system, solve the trapdoor unlinkability problem and solve access pattern of search user from the cloud server.
- we give thorough security analysis and high security level including confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability and access pattern of the search user. The terms are compared with existing proposals.

The remainder of this paper is organized as follows. In Section II, the system model, security requirments and design goal are formalized. In Section III, we recap Group creation, Text mining process, Blind storage, Query search .Its security analysis and performance evalution are presented in Section IV and Section V, respectively. In Section VI, we present related work. Finally, we conclude this paper in section VII

Systemmodel,security requirements and design goal:-

System model:-

As shown in fig.1, the system model Is consists of three data owner, search users and cloud server.

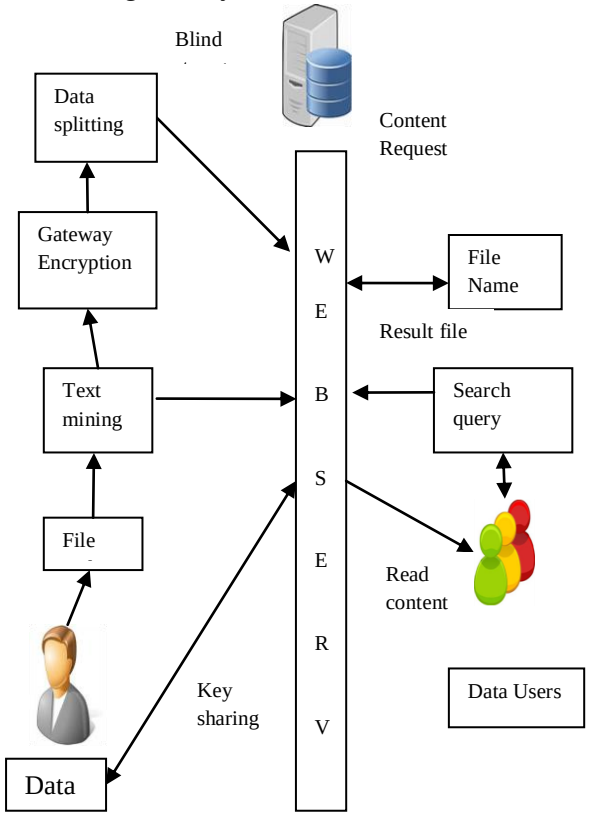

Figure 1:- System model

efficient and reliable methodology for search over encrypted data which is splited in to multiple blocks and then stored in blind storage. Here the encrypted multi keyword search pre computes the resulting search documents for the input query from users through Natural language processing Technique which is implemented on gateway (client side) on user file upload. Hence the matching documents which is pre compute the before searching the encrypted cloud contents are retrieved from cloud. Here we does not pull all the encrypted data's from cloud for searching, which is time consuming and ineffective. The matching documents memory locations on blind storage are retrieved from the serializable objects which is stored in the gateway. User can download the resulting documents after getting the keys from the group owner. Asymmetric kind of encryption for key re-encryption and is more secured.

Multiple groups can be created. Each group is having owner and users. User can upload the files in public and private mode. If user uploads files in public mode, the file is to encoded using Base64 algorithm. If user uploads in private mode the file content is encrypted using RSA algorithm and then can give access control for each group user. User search in cloud using keywords, cloud can send the related files to respective user. If data user wants to read the contents of files, data user should request to cloud and then cloud will request to data owner. Data owner checks the user attributes and access control, then the owner forward the private key and data's in secure manner.

Security requirements:-
In the ,we consider the cloud server to be curious but honest which means it executes the task assigned by the data owner and the search user correctly. However, it is curious about the data in its storage and the received the data Provide the following four security requirements:

- Confidentiality of document and index
- Privacy
- Trapdoor unlinkability
- Access pattern of the search user

Design goal;-
To enable efficient and efficient and privacy-preserving multi-keyword ranked search over encrypted cloud data via blind storage system

- Multi-Keyword Ranked Search
- Search Efficiency
- Confidentiality and privacy preservation

Preliminaries;-

Group creation:-
Data owner should be register in this environment and create a group. Data users also register and give request to group owner to add a group user. Data owner accept the request from the user. Multiple groups can be created. Each group is having owner and users. Data user only can access the respective data owner documents. Data user cannot access the webpage until the data owner accepts the request.

Text mining process;-
In this module the data owner can upload the document. Data owner can upload the files, the content of file is to be extracted using NLP technique and that words can get synonyms using Word Net tool. In first step of text mining process POS tagger is implemented to extract the keywords in files .NLP process is used to extract the literal meaning of keywords previously extracted. The Words are analyzed in Word Net API so that the related terms can be found for use in the index file. This index file will be generated for each upload from group owner and saved as a serializable object in cloud. All the communication to cloud server will be done through web service.

Blind storage:-
The uploaded data's are encrypted in gateway after Natural language Processing is done and stored as index file. The owner can give access control and privileges to 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 splited into blocks and its encrypted using RSA encrypting algorithm and the encrypted blocks will be uploaded to the cloud service and stored in blind storage. Blind storage all documents are divided into fixed size blocks. These blocks are indexed by 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.

Query search:-
Data user will try to search a query in cloud server. The cloud servers map the keywords and search the related files. The cloud server gives the related filename to user. To view the content the user should click the filename; 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.

Proposed scheme:-

In Proposed system, we introduced an efficient and reliable methodology for search over encrypted data which is splited in to multiple blocks and then stored in blind storage. Here the encrypted multi keyword search pre computes the resulting search documents for the input query from users through Natural language processing Technique which is implemented on gateway (client side) on user file upload. Hence the matching documents which is pre compute the before searching the encrypted cloud contents are retrieved from cloud. Here we does not pull all the encrypted data's from cloud for searching, which is time consuming and ineffective. The matching documents memory locations on blind storage are retrieved from the serializable objects which is stored in the gateway. User can download the resulting documents after getting the keys from the group owner. Asymmetric kind of encryption for key re-encryption and is more secured.

Multiple groups can be created. Each group is having owner and users. User can upload the files in public and private mode. If user uploads files in public mode, the file is to encoded using Base64 algorithm. If user uploads in private mode the file content is encrypted using RSA algorithm and then can give access control for each group user. User search in cloud using keywords, cloud can send the related files to respective user. If data user wants to read the contents of files, data user should request to cloud and then cloud will request to data owner. Data owner checks the user attributes and access control, then the owner forward the private key and data's in secure manner

Security analysis:-

Under the assumption presented in section II, we analyse the security properties in the term of confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability and access pattern of the search user.

Related work:-

In existing system encryption of the documents are done in cloud server. All the files uploaded by the user are encrypted in cloud and stored in static memory locations. Hence Multi keyword search is not possible on the encrypted cloud data. In order to make a search, the existing system downloads all the encrypted files and then decrypt for content based searching which is the traditional way to search. In searchable symmetric encryption (SSE) schemes, large number of documents, search results should be retrieved in an order of the relevancy with the searched keywords using TF-IDF method.

Problem Definition Outsourced encrypted Data are directly stored in cloud, which may lead to severe confidentiality and privacy issues. Searchable encryption schemes fail to offer sufficient insights towards the construction of full functioned search over encrypted cloud data. Server side encryption which is in secure. Bulk content retrieval for file searching, which is inefficient.

Group sharing with access control on encrypted data is not well studied yet. search over encrypted data which is splited in to multiple blocks and then stored in blind storage. Here the encrypted multi keyword search pre computes the resulting search documents for the input query from users through Natural language processing Technique which is implemented on gateway (client side) on user file upload. Hence the matching documents which is pre compute the before searching the encrypted cloud contents are retrieved from cloud. Here we does not pull all the encrypted data's from cloud for searching, which is time consuming and ineffective. The matching documents memory locations on blind storage are retrieved from the serializable objects which is stored in the gateway. User can download the resulting documents after getting the keys from the group owner. Asymmetric kind of encryption for key re-encryption and is more secured.

Multiple groups can be created. Each group is having owner and users. User can upload the files in public and private mode. If user uploads files in public mode, the file is to encoded using Base64 algorithm. If user uploads in private mode the file content is encrypted using RSA algorithm and then can give access control for each group user. User search in cloud using keywords, cloud can send the related files to respective user. If data user wants to read the contents of files, data user should request to cloud and then cloud will request to data owner. Data owner checks the user attributes and access control, then the owner forward the private key and data's in secure manner Multiple group creation, each group is having owner and multiple users.

We can give access control to each file for separate user. To encrypt data using Asymmetric algorithm (RSA) and key re-encryption. Using NLP technique and word net tool for text mining process. Index file generation on cloud.

Conclusion:-

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 .For the future work, we will investigate on the required number of files gives to searching query issues in searchable encryption technique

References:-

1. H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, ``An SMDP-basedservice model for interdomain resource allocation in mobile cloud networks,"*IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2222_2232,Jun. 2012.
2. M. M. E. A. Mahmoud and X. Shen, ``A cloud-based scheme for protectingsource-location privacy against hotspot-locating attack in wirelesssensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 10,pp. 1805_1818, Oct. 2012.
3. Q. Shen, X. Liang, X. Shen, X. Lin, and H. Y. Luo, ``Exploiting geodistributed
4. clouds for a e-health monitoring system with minimum servicedelay and privacy preservation," *IEEE J. Biomed. Health Inform.*, vol. 18,no. 2, pp. 430_439, Mar. 2014.
5. H. T. Dinh, C. Lee, D. Niyato, and P. Wang, ``A survey of mobilecloud computing: Architecture, applications, and approaches," *WirelessCommun. Mobile Comput.*, vol. 13, no. 18, pp. 1587_1611, Dec. 2013.
6. H. Li, Y. Dai, L. Tian, and H. Yang, ``Identity-based authenticationfor cloud computing," in *Cloud Computing*. Berlin, Germany:Springer-Verlag, 2009, pp. 157_166.
7. W. Sun, *et al.*, ``Privacy-preserving multi-keyword text search in the cloudsupporting similarity-based ranking," in *Proc. 8th ACM SIGSAC Symp.Inf., Comput. Commun. Secur.*, 2013, pp. 71_82.
8. B.Wang, S.Yu,W. Lou, andY. T. Hou, ``Privacy-preserving multi-keywordfuzzy search over encrypted data in the cloud," in *Proc. IEEE INFOCOM*,Apr./May 2014, pp. 2112_2120.
9. E. Stefanov, C. Papamanthou, and E. Shi, ``Practical dynamic searchableencryption with small leakage," in *Proc. NDSS*, Feb. 2014.
10. Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, ``Secure dynamic searchablesymmetric encryption with constant document update cost," in *Proc.GLOBECOM*, Anaheim, CA, USA, 2014.
11. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Ro³u, and M. Steiner,``Highly-scalable searchable symmetric encryption with support for Boolean queries," in *Proc. CRYPTO*, 2013, pp. 353_373.