ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

Privacy Preserving Similarity-Based Text Retrieval through Blind Storage

[1]B.VIDHIYALAKSHMI

[1]M.E-CSE(Final Year Student)-T.J.S. ENGINEERING COLLEGE,Peruvoyal, Thiruvallur Dt ,Tamil Nadu.
vidit.56@gmail.com

ABSTRACT

The fundamental application of mobile cloud computing is outsourcing the data to external cloud server for system usability, resource utilization, pay as per use for cloud users, after this outsourced data need to be encrypted because privacy and confidentiality concerns of the owner. Due to these difficulties accurate search over encrypted mobile data not get. Consider searchable encryption for multi-keyword ranked search over the storage data. The large number of outsourced documents (data) in the cloud, utilize the relevance score and *k*-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked searchresults based on the accuracy. Efficient index used to improve the search efficiency and adopt the blind storage system to conceal access pattern of the search user. Security analysis achieves confidentiality of documents and index, trapdoor privacy, trapdoor unlink ability, and concealing access pattern of the search user. Searchable encryption scheme which achieves high efficiency for large databases with modest scarification on security guarantees. Improved efficiency in terms of search functionality and search time also achieves.

Key Terms: cloud computing, multi keyword ranked search, blind storage, access point

I.INTRODUCTION

Mobile cloud computing at its simplest refers to an infrastructure where both the data storage and data processing happen outside of the mobile device. Mobile cloud applications move the computing power and data storage away from mobile phones and into the cloud, bringing applications and mobile cloud to not just Smartphone users but a much broader range of mobile subscribers.

In cloud computing, users can outsource their computation and storage to servers (also called clouds) using Internet. Clouds can provide several types of services like applications (e.g., Google Apps, Microsoft online), infrastructures (e.g., Amazon's EC2, Eucalyptus, 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

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

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.

Cloud computing is a large-scale distributed network system implemented based on a number of servers in datacenters. The cloud services are generally classified based on a layer concept. In the upper layers of this paradigm, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) are stacked.Extending battery lifetime, battery is one of the main concerns for mobile devices.Improving data storage capacity and processingpower.

II.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.

III. EXISTING SYSTEM

3.1 Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud

It enhances multikeyword exact search that does not tolerate keyword spelling error, or single keyword fuzzy search that tolerates types to certain extent and keyword search solutions which require expanded storage for wild-card based fuzzy keyword set. The current fuzzy search schemes rely on building an expanded index that covers possible keyword misspelling, which lead to significantly larger index file size and higher search complexity and locality sensitive hashing to provide efficient fuzzy search with constant size index regardless the number of keywords associated with the file. Multikeyword fuzzy search scheme by exploiting the locality-sensitive hashing technique. The fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need of a predefined dictionary and effectively supports multiple keyword fuzzy searches without increasing the index or search complexity. Known cipher text model the cloud server can only access the encrypted files, the secure indexes and the submitted trapdoors. The cloud server can also know and record the search results. Known Background Model the cloud server knows additional background information in this model. The background refers to the information which cans be learned from a comparable dataset.

3.2 Enabling secure and efficient ranked keyword search over outsourced cloud data

This describes Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud. Using relevance score from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

resulting design is able to server-side ranking without losing keyword privacy. Together the advance of both crypto and IR community to design the ranked searchable symmetric encryption scheme .a ranking function is used to calculate relevance scores of matching files to a given search request. evaluating relevance score in the information retrieval community uses the TF_ IDF rule, where TF (term frequency) is simply the number of times a given term or keyword appears within a file (to measure the importance of the term within the particular file), and IDF (inverse document frequency) is obtained by dividing the number of files in the whole collection by the number of files containing the term.

3.3 Toward secure multikeyword top-k retrieval over encrypted cloud data

This describes Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud. In this paper, we focus on addressing data privacy issues using SSE. Formulate the privacy issue from the aspect of similarity relevance and scheme robustness. We observe that server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. To eliminate the leakage, we propose a two-round searchable encryption (TRSE) scheme that supports top-$(k) $ multikeyword retrieval. In TRSE, we employ a vector space model and homomorphic encryption. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on ciphertext. As a result, information leakage can be eliminated and data security is ensured. Thorough security and performance analysis show that the proposed scheme guarantees high security and practical efficiency.

IV. SYSTEM DESIGN

4.1 System Architecture

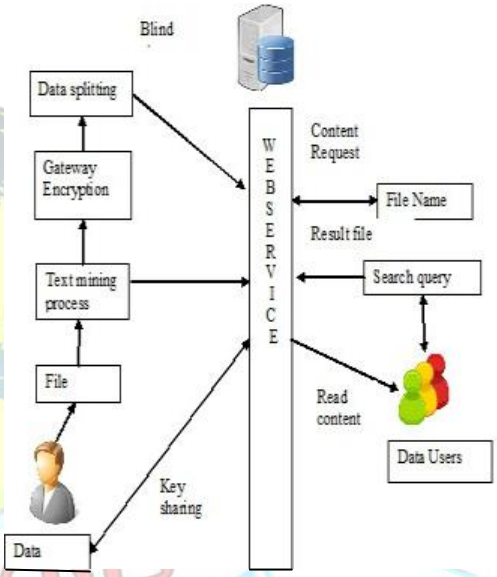

Fig.4.1 System Architecture

4.2. MODULES DESCRIPTION

Modules

- **Group creation**
- **Text Mining**
- **Blind storage**
- **Query search**

4.2.1 Module 1: 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

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

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.

4.2.2 Module 2: 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.

4.2.3 Module 3: 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.Files content are stored in block randomly so the cloud can view encrypted content only. Encryption key only knows to data owner.

4.2.4 Module 4: 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 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.

Algorithm Used

- Dynamic Blocks splitting algorithm
- RSA algorithm
- Base64
- Bubble sort algorithm
- Relevance semantics on corpus
- NLP technique

8. CONCLUSION

I 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 multi-keyword ranked search scheme to enable accurate, efficient and secure search over encrypted mobile cloud data. Security analysis have demonstrated that proposed scheme can effectively achieve confidentiality of documents

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

and index, trapdoor privacy, Trapdoor unlink ability, and concealing access pattern of the search user. Extensive performance evaluations have shown that the proposed scheme can achieve better efficiency in terms of the functionality and computation overhead compared with existing ones, I will investigate on the authentication and access control issues in searchable encryption technique. It can mainly be classified into two types: Searchable Public-key Encryption (SPE) and Searchable Symmetric Encryption (SSE).Although SPE can achieve above rich search functionalities.SPE are not efficient since SPE involves a good many asymmetric cryptography operations. This motivates the research on SSE mechanisms.

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
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.
4. 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.
5. 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.
6. 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.
7. 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.
8. E. Stefanov, C. Papamanthou, and E. Shi, ``Practical dynamic searchableencryption with small leakage," in *Proc. NDSS*, Feb. 2014.
9. 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.
10. D. Cash *et al.*, ``Dynamic searchable encryption in very-large databases:Data structures and implementation," in *Proc. NDSS*, Feb. 2014.