International Journal of Advanced Engineering Research and Science (IJAERS) *Special Issue (NCCSIGMA-16)*
https://dx.doi.org/10.22161/ijaers/si.11 *ISSN: 2349-6495(P) | 2456-1908(O)*

Confidentiality Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data

V Naresh Kumar, Kranthi Kiran G

Department of CSE, CMR Technical Campus, Hyderabad, India

Abstract*—With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search and rarely sort the search results. This paper defines and focuses on the problem of ranked search over encrypted data. It captures the relevance of data documents to the search query.*

Keywords— Cloud Computing, privacy, encryption, ultiple keywords, ranked search.

I. INTRODUCTION

Cloud Computing is computing utility where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from shared pool of configurable computing resources. It is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In simple terms, cloud computing means storing and accessing data and programs over the internet instead of a computer's hard drive.The importance of cloud computing is increasing and receiving growing attention in the scientific and industrial communities [3]. It is a flexible, cost-effective and proven delivery platform providing business or consumer IT services over the internet. However cloud computing has an added risk as the essential services are often outsourced to a third party, which causes data security and privacy, support data and service availability.Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system. To protect privacy of data and oppose unsolicited accesses in the cloud and beyond it, sensitive data, for instance, e-mails, personal health records, photo albums, tax documents, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search.

The insignificant solution of downloading all the data and decrypting locally is clearly impractical, due to the large amount of bandwidth cost. Moreover, aside from eliminating the local storage management, storing data into the outsourced storage doesn't serve any purpose unless they can be easily searched and utilized.Ranked search [1] can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable. Hence, exploring privacy-preserving and effective search service over encrypted data [2] is of great importance. To enhance the search result, we need efficient methods to perform similarity search over large amount of encrypted data.LSH (locality sensitive hashing) is extensively used for fast similarity search on plain data in information retrieval community. In our scheme, we propose to utilize it in the context of the encrypted data.

II. DETAILS EXPERIM ENTAL

2.1. Problem Statement:

The great flexibility and economic savings of outsourced storage are motivating both individuals and enterprises to outsource their local complex data management system. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, may have to be encrypted by data owners before outsourcing.

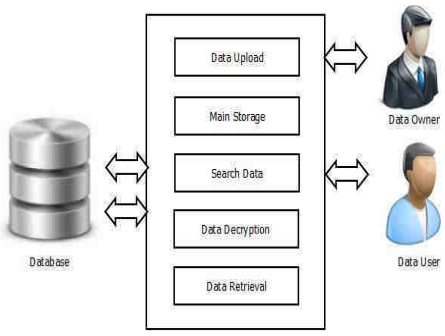

Fig.1: Architecture Diagram

2.2 Proposed Work:

To enable efficient similarity search, data owner builds a

https://dx.doi.org/10.22161/ijaers/si.11 *ISSN: 2349-6495(P) | 2456-1908(O)*

secure index and outsources it to the cloud server along with the encrypted data items. Serverperforms search on the index according to the queries of the data users without learning anything about the data other than what data owner allows an adversary to learn.

2.2.1Data Owner's Modules:

1. Login Module
2. Data Upload
3. View Files(Search)
4. File Share

1. Login Module:

Data Owner need to login on the portal in order to upload the data on the cloud.

2. Data Upload:

This module is dedicated for the process of uploading the content on the cloud server. Data owner have to select any data file (text) from the local machine in order to upload it on cloud server.

3. View Files(Search):

If the data owner wants to search any specific file, this module provides the option. The output of search operation will provide the list of the files; the module will also have the option for deletion of specific file.

4. File Share:

This module provides the option for file sharing. The module is somehow depends on the View file module. After searching files, the data owner can select specific file(s) and select specific user(s) with whom he want to share with.

2.2.2Data Users Modules Description:

1. Login/Registration
2. View Files(Search)
3. Download Files

1. Login/Registration:

In order to be a part of domain, user needs to register on the portal. In order to register on the portal, user need to provide some personal information which will be stored at back-end. At the time of login the credentials Username and password will be used to authenticate the user.

2. View Files(Search):

It is the core module for the Data user. In this module user can search specific files (desired). User can search specific file by providing a tag (can be part of content or name of the file) in the searching bar, if the file exists for the particular tag value, the file title will be shown on the dashboard. In the case of wrong tag insertion for search, the returned output will be null.

3. Download Files:

This module depends on the Search module. After searching files if some files list returned, User gets option for downloading the files. User can selectspecific file and download it by providing the respected private key (which is mailed him/her by the data owner)

III. RESULTS AND DISCUSSION

3.1 Algorithm

In our paper the following algorithm is used:

Algorithm Steps

During File Uploading following Steps will execute.

1. Reads file data and split into word by word (based on space).
2. Removal of Stopwords(eg:- and,or,anetc)
3. Performs Stemming(e.g.;- suppose two words print,printed.removed from second word.)

These three steps are called preprocessing steps.

4. Calculate tf, tfidf value for each keywords in the file and stores in the database.
5. For first 5 uploads,performs only this steps.
6. For the next uploads,along with these steps performs following steps also
7. Creates a document vector [array of tfidf value all the keywords of uploading file].
8. Creates document vector for all other uploaded files also.
9. Performs cosine similarity calculation of uploading doc with all other docs in the database.

So for vector A = (a1, a2) and B = (b1, b2), the
cosine similarity is given as:
(a1 b1 + a2 b2) / sqrt(a1^2 + a2^2) sqrt(b1^2 + b2^2)
Where a1, a2, b1, b2 are tfidf values.

Eg:-If 6^{th} file is uploading,will get 5 cosine similarity.(how much similar the 6^{th} file with all 5 uploaded doc.)

Calculates Avg of all 5 cosine similarity.then takes cosine similarities which are greater than this avg.that files and this uploading files will comes under one cluster.

10. Finds high frequency 10 keywords in each doc and stores in files table.

Index Construction

For each cluster, constructs an index.

1. Finds the files in that cluster.
2. Retrieves high frequency keywords in all files in that cluster.
3. Finds common characters in all the keywords, maximum of 4. Eg:-ACDF.
4. Finds TFIDF of top 10 high frequency keywords in that cluster.

Eg:- TFIDF*1000 → for all top 10 keywords (bucket)→ xi → x1 * x2 * x3 .... x10 → 13467

5. Concatenates ACDF13467->This is the index for that cluster.

Searching

https://dx.doi.org/10.22161/ijaers/si.11 *ISSN: 2349-6495(P) | 2456-1908(O)*

Following steps will perform during searching.
1. User inputs search keywords in textbox.
2. Retrieves tfidf of keywords from table.
3. Finds cluster [9] which contain the keyword and retrieves index.
4. Splits the index into two part.
5. e.g. search → hello → 123, world → 1452
13467 % 123 == 0 && 13467 % 145 == 0
6. Performs character analysis based on otherpart of index.
7. Finds cluster which having more character similar to search term.
8. Select indexes similar to tfidf and character analysis.
9. Identify keywords associated to index and similar search keywords.
10. Display docs related to this index.

Ranking

Eg: - apple, fruit, tree, plant- these are the search keywords.
TFIdf are
apple-10
Fruit-8
Tree-11
Plant-4
Doc2 contain apple and fruit->10+8=18 doc1 → fruit, tree → 8 + 11 = 19 doc3 → plant → 4.
Order of Display will be, doc1doc2doc3

IV. CONCLUSION

In this paper, we proposed an efficient similarity searchable symmetric encryption scheme. To do so, we utilized locality sensitive hashing which is widely used for fast similarity search in high dimensional spaces for plain data. We proposed LSH based secure index and a search scheme to enable fast similarity search in the context of encrypted data. In such a context, it is very critical not to sacrifice the confidentiality of the sensitive data while providing functionality. We provided a rigorous security and proved the security of the proposed scheme under the provided definition to ensure the confidentiality.

REFERENCES

[1] Chi Chen, Member, IEEE, Xiaojie Zhu, Student Member,IEEE, Peisong Shen, Student Member, IEEE,Jiankun Hu, Member, IEEE, Song Guo, Senior Member, IEEE, Zahir Tari, Senior Member, IEEE, and Albert Y. Zomaya, Fellow, IEEE, " An Efficient Privacy-Preserving Ranked Keyword Search Method" , IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 4, APRIL 2016

[2] Ning Cao†, Cong Wang‡, Ming Li†, Kui Ren‡, and Wenjing Lou††Department of ECE, Worcester Polytechnic Institute, Email: {ncao, mingli, wjlou}@ece.wpi.edu,‡Department of ECE, Illinois Institute of Technology, "Privacy Preserving Multi Keyword Ranked Search Over Encrypted Cloud Data.

[3] Keiko Hashizume, David G Rosado2, Eduardo Fernández-Medina and Edu ardo B Fernandez, "An analysis of security issues for cloud computing", Hashizume et al. Journal of Internet Services and Applications 2013, 4:5

[4] Cong Wang†, Ning Cao‡, Jin Li†, Kui Ren†, and Wenjing Lou‡†Department of ECE, Illinois Institute of Technology, Chicago, IL 60616 Email: {cong, jli, kren}@ece.iit.edu‡Department of ECE, Worcester Polytechnic Institute, Worcester, MA 01609 Email: {ncao, wjlou}@ece.wpi.edu,

[5] "Secured ranked Search over Encrypted Data" 2010 International Conference on Distributed Computing Systems

[6] W. K. Wong,The University of Hong Kong wkwong2@cs.hku.hk ,David W. Cheung The University of Hong Kong dcheung@cs.hku.hk Ben Kao The University of Hong Kong kao@cs.hku.hk Nikos Mamoulis The University of Hong Kong nikos@cs.hku.hk, "Secure kNN Computation on Encrypted Databases"

[7] Cong Wang, Student Member, IEEE, Qian Wang, Student Member, IEEE,Kui Ren, Senior Member, IEEE, Ning Cao, and Wenjing Lou, Senior Member, IEEE, "Toward Secure and Dependable Storage Services in Cloud Computing", 220 IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 5, NO. 2, APRIL-JUNE 2012

[8] Keiko Hashizume1*, David G Rosado2, Eduardo Fernández-Medina2 and Eduardo B Fernandez1," An analysis of security issues for cloud computing, Hashizume et al. Journal of Internet Services and Applications 2013, 4:5

[9] T. Jothi Neela1* and N. Saravanan2," Privacy Preserving Approaches in Cloud", Vol 6 (5) | May 2013

[10] Yong-Il Kim1, Yoo-Kang Ji2 and Sun Park3*," Big Text Data Clustering using Class Labels and Semantic Feature Based on Hadoop of Cloud Computing", International Journal of Software Engineering and Its Applications Vol.8, No.4 (2014)

[11] Muhammad YasirShabir, Asif Iqbal, Zahid Mahmood_, and AtaUllahGhafoor, " Analysis of Classical Encryption Techniques in Cloud Computing", TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll09/10llpp102-113 Volume 21, Number 1, February 2016