International Journal of Science, Technology and Management (IJSTM) ISSN (online): 2321-774X
Volume 9, Issue 1, 2022

Protection of Data Stored in Cloud Computing Using ESSMO and OSNQSC

Nisha, Satvender Kumari*

Department of Computer Science and Engineering

BRCM College of Engineering & Technology, Bahal, Bhiwani, INDIA.

**Corresponding Author email: satvender@brcm.edu.in*

Abstract - Retrieval of precise information is challenging over encrypted data stored in cloud. There is a large volume of data set which spans across multiple independent clouds. It includes the process of information aggregation from source generators and modelling of such information is done along with categorizing the data. In this article two schemes ESSMO (An Efficient Search Scheme on Encrypted Cloud Data using Modular OPE) and OSNQSC (Optimized Selection of Nodes for Enhanced QoS in Cloud Environment) are presented. An efficient search scheme using ESSMO reduces the risk of message leakage. Modular OPE doesn't reveal the physical location of the file stored and reduces the risk of message leakage. OSNQSC achieves data integrity by assuring the user about not corrupting the data. Privacy of the data is ensured by giving the access to upload the data only to authorized users.

Index Terms – Cloud computing, Data storage, ESSMO, OSNQSC

I. Introduction

Cloud is based on service model, where data is maintained offloading the contribution of the device resources. Mobile Cloud Storage (MCS) [1] provides good amount of on-line services and provide a file storage system based on devices. The users will be allowedto store files and retrieve files from the cloud with the help of wireless communication. The data is encrypted, stored in cloud as mobile devices and PCs have security threats. The mobile device with limited battery capacity faces the challenges of the computing power while encrypting the data [2]. Across the globe, enormous amount of outsourced information are stored and retrieved. In the process, various nuisances regarding data security arise while providing retrieval and searching procedures. Bearing the various security concerns into consideration, the solution for protection is to upload data into the cloud server after encryption. Before uploading the file to the server, it has to be encrypted and then decryption mechanism can be run only after downloading the files. For the end users, the devices have limited battery capacity and computing power that leads to high energy consumption [3]. So the responsibility is transferred to the cloud in the proposed method. It performs the search of the articles and ranks them in an optimize fashion so that the accuracy is increased. All traditional Searchable Symmetric Encryption (SSE) [4] schemes allow the user to search on cipher-text and securely retrieve the cipher-text over encrypted cloud data through keywords without decrypting the files. This supports only Boolean-keyword search without considering any relevance ofthe document. Boolean-keyword search has a main drawback whenever a huge number of documents are involved.

II. ESSMO (An Efficient Search Scheme on Encrypted Cloud Data using Modular OPE)

Figure 1 shows the model that contains the owner, the user and the cloud as entities. The owner has the authority to upload the encrypted articles after completing the processes such as stemming, tokenization, frequency computation, TF-IDF calculation.

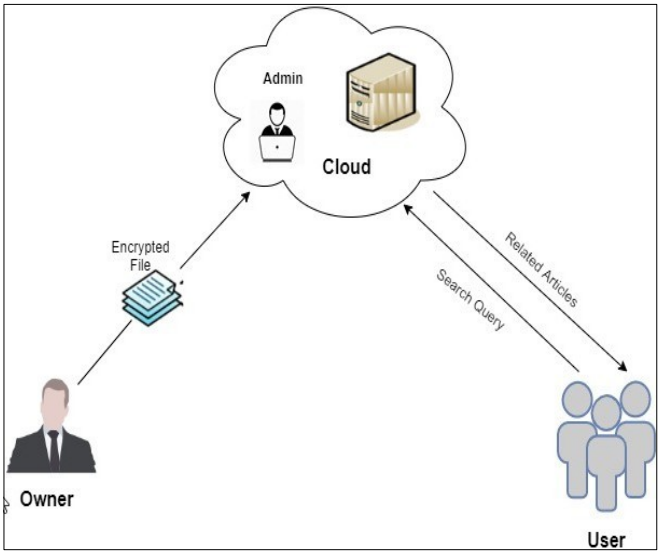

Fig.1: Proposed Method

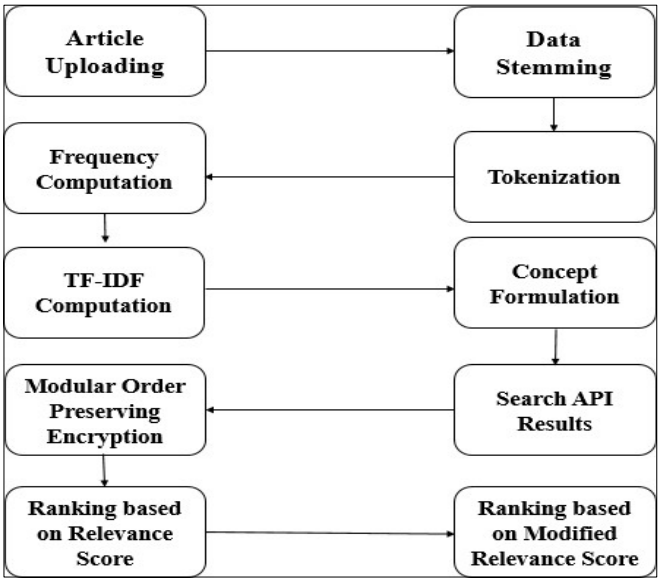

Fig.2: Steps of ESSMO

Once the articles are uploaded, the admin who resides inside the cloud has the access to view the user details, encrypted articles. The user who is in need of the article sends the search query in the encrypted form. The cloud calculates the relevance score of the keyword with the help of Google API. Then the articles with higher relevance score value are sent to the user along with few suggestions. The work flow of the proposed method with 10 modules is as shown in Fig. 2. Data cleaning is a pre-processing step which is responsible for removal of special symbols and stop words from the article. Stop words are list of words like is, the, about, etc. In the implementation of the proposed method, more than 1000 standard stop words with unlimited custom stop words are used to clean an article data.

The sequence of words are stored in a separate queue in orderto obtain clean article details. The data cleaning and stemming algorithm is summarized in Algorithm 1. Tokenization is responsible for converting the clean data into a set of tokens across all thepages of the articles. The tokenization algorithm is provided in Algorithm 2.

Algorithm 1 Stopword Cleaning Algorithm

Input: Article A_i

Output: Cleaned article pages A_iPj

begin

Step1: Divide the article into number of pages P_1, P_2,., P_n

Step2: for each page P_i, remove the unwanted symbols and form a Queue Q_i of the words in a page

Step 3: Measure the count of number of words W_i in Q_i.*for* each W_i in Q_i

if $W_i \in SW_1$, SW_2, , SW_n

move on to next element of Q_i

else

add W to C_A and move to next element of Q_i

end ifend for

Step 4: After the above steps are done a set of clean article pages are obtained

A_1P_1, A_1P_2, , A_1P_n

end

Algorithm 2 Tokenization Algorithm

Input: Set of cleaned article pages A_1P_1, A_1P_2,….., A_1P_n.

Output: Tokenized Words W_1,W_2,. ,W_n.

begin

Step1: Measure the count of FIFO Queue CA_{count}.*for* each word in CA_{count} assign ID and move to next word

end for

Step 2: Measure the number of elements of FIFO queue N_{FIFO}.

Step 3: The set of words $W_1, W_2, \ldots, Wn$ is the tokenized set.

end

An efficient search scheme using ESSMO reduces the risk of message leakage. Modular OPE doesn't reveal the physical location of the file stored and reduces the risk of message leakage. Additional suggestions are also considered from the user prospective to improve the searching. Network traffic is reduced by communicating the selected index. The proposed method reduces the file retrieval time and reduces the workload on mobile devices.

III. OSNQSC: Optimized Selection of Nodes for Enhanced QoS in CloudEnvironment

OSNQSC architecture has two entities as shown in Fig. 3 i.e., the cloud and the user.

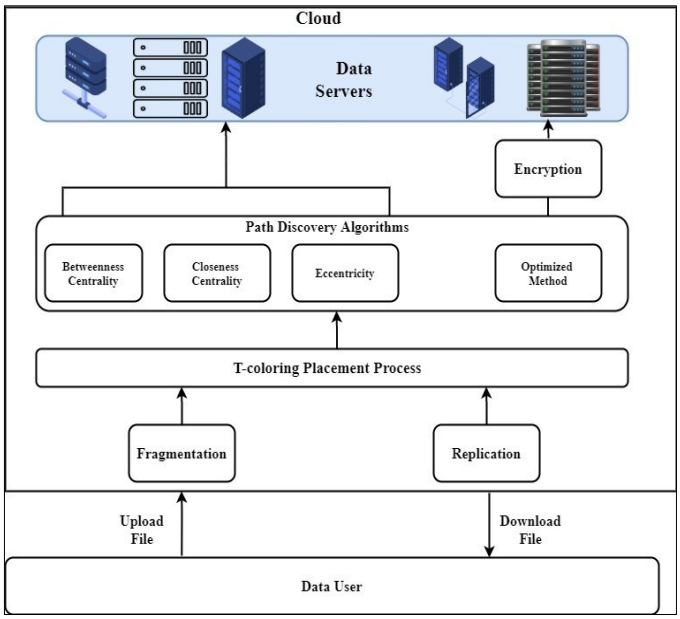

Fig.3: OSNQSC Architecture

The user who wants to upload the data safely in cloud. The cloud is the one who provides the storage services to the user. The interaction between the user and the cloudis as follows:

1. The user has to register themselves to get authenticated.
2. The user sends the file to be stored in the cloud.
3. In proposed optimized method, the file fragments are encrypted before placing anduploaded to cloud.
4. The file is divided into number of fragments using SPLIT method. Along with this,ISPs are deployed in the cloud using T-coloring Placement process.
5. The file fragments are placed on the selected ISPs using Path discovery algorithms.

Data fragmentation means, the data is broken into multiple pieces stored in memory that are not close together. The main objective of fragmentation is to efficiently use the storage space [5-6]. The importance of fragmentation depends on the specific storage allocation system.

The file that is uploaded by the user should be in pdf format. The file is divided into multiple pages. Each page is considered as the independent data fragment. These fragments will be stored in the ISP which is present in the nodes selected by the path discovery algorithms. Along with fragments, the replicated fragments [7] are also placed in other nodes of the network.

T-coloring is responsible for placing the ISP in a region of bounded limits on cloud. Each ISP will have its own unique location. The location of two ISPs cannot be same as well. The ISP location can change for each iteration and each iteration position are presented in differentcolors for all 25 ISPs within boundary range of 25*25 $m^2$and same is depicted.

Algorithm 3 shows the node position by T-color Placement Process in which ISP are distributed within cloud by taking number of nodes N_{ISP} required to deploy in cloud, minimum x position, maximum x position, minimum y position and maximum y posi- tion. Position of node is generated randomly between x_{min} to x_{max} and y_{min} to y_{max} and information is stored in (i, x_{pos}, y_{pos}) until all the nodes are placed in the network.

Algorithm 3 T-color Placement Process

Input: N_{ISP}, ha_{start}, ha_{end}, va_{start}, va_{end}.

Output: T-coloring Placement Matrix $TCPM$.

begin

for $l = 1$ N_{ISP}.→

do

Step 1 : Find the first dimension for the ISP in the area under coverage

ha_{start}, ha_{end}. The dimension must satisfy the condition.

$ha_i = ha_v$ for any ha_v

$ha_{start} \leq ha_v \leq ha_{end}$ and ha_v

ha_h

Step 2 : Find the second dimension for the ISP in the area under coverage

va_{start}, va_{end}. The dimension must satisfy the condition.

$va_i = va_v$ for any v_{av} which satisfies

$va_{start} \leq va_v \leq va_{stop}$ and $va_v \neq va_h$.

Step 3 : The complete set is formed using $(l, (ha_l, va_l))$.

Step 4 : Save the l'h data of Matrix

Step 5 : $l = l + 1$

End for

end

OSNQSC is proposed for outsourcing data that considers both performance and security. The method fragments and replicates the file over cloud nodes. The nodes are selected by considering quality of the nodes in terms of energy and the distance between the nodes. Dividing the file into fragments provides the confidentiality to the user's data. No meaningful information is revealed to an attacker on any successful attack on the nodes. The fragments are encrypted to ensure the double security on the fragments of file and the controlled replication of the file fragments, where each fragments is replicated only once for the purpose of improved security. OSNQSC achieves data integrity by assuring the user about not corrupting the data. Privacy of the data is ensured by giving the access to upload the data only to authorized users.

IV. Conclusion

In this work, a multi-keyword search scheme on encrypted data is proposed that uses the Modular OPE algorithm to encrypt the keywords in roder to provide the double security to the stored data. In ESSMO, TF table is used as index and the cloud server calculates the relevance score using the encrypted TF value. Modified TF-IDF contains the TF-IDF values along with score of search results obtained for top keywords of Google API for eacharticle. The proposed method is performs better as compared to TEES with respect to time taken to search an article and throughput.

From OSNQSC, a cloud storage system in this work enhances the QoS using optimized method. This method selects only those nodes to store which have high energy capacity and also have shorter distance. The files are fragmented and then stored in the selected ISPs of optimized method. An encryption method is introduced before uploading to provide the double security to the data. Thus, the proposed method helps in reducing the time toupload, download file that intern reduces the routing overhead with higher throughput.

References

[1] Xiaojun Yu and Qiaoyan Wen, "Design of Security Solution to Mobile Cloud Storage", Springer Book Title Knowledge Discovery and Data Mining, vol. 135, pp. 255-263, 2012.

[2] Oleksiy Mazhelis, Gabriella Fazekas, and Pasi Tyrvinen, "Impact of Storage Acquisition Intervals on the Cost-Efficiency of the Private vs. Public Storage", IEEE Fifth International Conference on Cloud Computing, pp. 646-653, June 2012.

[3] Ian Witten, Alistair Moffat, and Timothy Bell, "Managing Gigabytes: Compressing and Indexing Documents and Images", Morgan Kaufmann publication, 1999.

[4] Qi Chai and Guang Gong, "Verifiable Symmetric Searchable Encryption for Semi Honest-but-Curious Cloud Servers", IEEE International Conference on

Communications (ICC), pp. 917-922, June 2012.

[5] Asma H. Al-Sanhani, Amira Hamdan, Ali B. Al-Thaher, and Ali Al-Dahoud, "A Comparative Analysis of Data Fragmentation in Distributed Database", 8th International Conference on Information Technology (ICIT), pp. 724-729, May 2017.

[6] Suganya and R. Kalaiselvi, "Efficient Fragmentation and Allocation in Distributed Databases", International Journal of Engineering Research and Technology, vol. 2, issue. 2, pp. 1-7, January 2013.

[7] Anca-Georgiana Fodor and Ion Lungu, "Implementation of Fragmentation and Replication Methods in Distributed Systems", Journal of Information Systems and Operations Management, pp. 373-383, 2016. Francis Bloch, Matthew O. Jackson, and Pietro Tebaldi, "Centrality Measures in Networks", Elsevier SSNR Publishing, June 2019.

[8] Mohammad Samadi Gharajeh, "A Dynamic Replication Mechanism in Data Grid based on a Weighted Priority-based Scheme", i-manager's Journal on Cloud Computing,

[9] vol. 6, no. 1, pp. 9-13, January - June 2019. Jeevitha B K, Sindhura D, Thriveni J, and Venugopal K R, "DSCESM: Data Security for Cloud Environment with Scheduled Key Managers", International Conference on Advances in Electronics, Electrical and Computational Intelligence, May 2019. Conor McBay, Genard Parr and Sally McClean, "Energy Saving in Data Center Servers using Optimal Scheduling to Ensure QoS", Seventh International Conference on Cloud Computing, Grid and Virtualization, pp. 57-60. 2016.

[10] Dzmitry Kliazovich, Pascal Bouvry, and Samee Utiab Khan, "DENS: Data Center Energy-Efficient Network-Aware Scheduling", Cluster Computing, pp. 65-75, 2011.

[11] Shruthi Dadhich and Vibhakar Pathak, "An Approach to Optimal Strategy for Energy Efficiency in Cloud System", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 5, issue. 6, pp. 1097-1101, June 2017.

[12] Stefan Rass and Peter Schartner, "Towards using Homomorphic Encryption for Cryptographic Access Control in Outsourced Data Processing", Seventh International Conference on Cloud Computing, Grid and Virtualization, pp. 7-13. 2016.

[13] K. Subramanian and F. Leo John, "Dynamic Data slicing in Multi Cloud Storage using Cryptographic Techniques", World Congress on Computing and Communication Technologies (WCCCT), pp. 159-161, 2017.

[14] http://www.cloudbus.org/cloudsim/

[15] Rodrigo N. Calheiros, Rajiv Ranjan, Cesar A. F. De Rose and Rajkumar Buyya, "CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructure and Services", Distributed, Parallel, and Cluster Computing, April 2009.

[16] https://towardsdatascience.com/graph-analytics-introduction-and-concepts-of centrality -8f5543b55de3.

[17][17]

https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to centrality-measures.