ISSN:0975-9646

Scalable and Efficient Keyword-based Semantic Search over Encrypted Multi-Cloud Data

Shraddha Ingale, Prof. Dr. B. D. Phulpagar

Department of Computer Engineering, Savitribai Phule Pune University

Pune, India.

***Abstract* — Cloud computing provides several attractive benefits for users like on-demand computing, pay as per use. It brings great convenience to consumers; where shared resources, data and information are provided to computers on-demand and consumer has to pay as per use. Ideally for these services, consumers should be in a position to verify the charges billed to them. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data. But on other hand consumers are facing serious difficulties that how to search the most suitable services from cloud. It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. Encryption is the process of encoding messages or information in such a way that only authorized parties can read it. Multiple encryption techniques are invented to encrypt the data. The searchable encryption increasing the attention of many researchers and hence different search methods over encrypted cloud data have been proposed. Semantic search addresses all limitations of existing search methods. This scheme not only supports keyword-based semantic search over encrypted data, but also provides verifiable searchability with data privacy preserving. This method achieves flexibility too. But these existing searchable encryption methods based on keyword search do no longer fully satisfy the new challenge and users' increasing needs. As it assumes that only limited number of data owner and data users, this can make huge load on single cloud server with more waiting time. To achieve scalability we introduced multi-cloud concept, if number of requests coming to current single cloud crosses the set threshold value it start transferring incoming requests to another cloud server to handle. This can saves lot of time, and increases the reliability and scalability of proposed system.**

Keywords* — *Consumer-centric cloud computing; privacy preserving; semantic search; verifiable search; multi cloud.

I. INTRODUCTION

Cloud computing services made available to consumers range from providing basic computational resources such as storage and compute power (infrastructure as a service, IaaS) to sophisticate enterprise application services (software as a service SaaS). A common business model is to charge consumers on a pay-per-use basis where they periodically pay for the resources they have consumed. An accounting model is said to be weakly consumer–centric if all the data that the model requires for calculating billing charges can be queried programmatically from the provider. Further, we can say that an accounting model is strongly consumer–centric if all the data that the model requires for calculating billing charges can be collected independently by the consumer. As a promise way of data storage in cloud computing environment, data outsourcing has attracted considerable attentions recently. By outsourcing their own data in the cloud, data owners can obtain high quality data storage services, while reducing the burden of local data storage and maintenance. To securely store outsourced data on an untrusted cloud server, sensitive data should be encrypted before outsourcing. Meanwhile many technical schemes related to cloud computing service are proposed by researchers. Noh et al [1] proposed a flexible communication bus model for multimedia services in cloud environment. Shahnaza et al [2] proposed a realistic IEEE 802.11e EDCA model for QoS-aware differentiated multimedia mobile cloud services. Cabarcos et al [3] proposed a middleware architecture that allows sessions initiated from one device to be seamlessly transferred to a second one under a cloud environment. The formatter will need to create these components, incorporating the applicable criteria that follow.

However, it is a difficult to search the most suitable products or services for ordinary consumers, as there are so many services and products present in cloud. It is a good practice to encrypt sensitive information before outsourcing rather than uploading plaintext directly on cloud. After encoding the plaintext become cipher text which is not readable by humans easily. However, existing search techniques fails to perform on cipher text as with plain text , thus results a big challenge to effective data utilization. The superficial solution of downloading the whole encrypted data first and then decrypting it locally is obviously impractical, due to the huge bandwidth and computation burden. Consumers might want to retrieve only certain specific data files they are interested in rather than the whole data collection. A popular way to address this problem is searchable encryption, which can retrieve specific files through keyword-based search with data protection and keyword privacy-preserving.

In this search area, many researchers have been invented multiple searchable encryption techniques. However, these searchable encryption schemes based on keyword search do no longer fully satisfy the new challenge and users' increasing needs, specifically manifested in the following two aspects.

One is that most of the existing schemes support only exact keyword search. That means the retrieved result is completely dependent on whether user enters match pre-set keywords. For example, if someone submits a query containing the term "system" ("system" is not in preset keyword sets but "machine" is), he will just retrieve an empty result ("system" and "machine" is almost same in the computer field).

The other one is that most of existing searchable encryption schemes assumes that the cloud server is honest-but-curious. However, Chai et al [4] notice that the cloud server may be selfish to save its computation or download bandwidth. That is, the cloud server might conduct only a fraction of search operation or return a part of result honestly.

Besides, Fu et al [5] proposed a multi-keyword search scheme in encrypted cloud environment and solved the problem of synonym search. Using this method, authorized cloud customers get the results as per their input and it's synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. This is a additional improvement in the field of searchable encryption. However, this scheme has not addressed the problems of semantic search and verification of search results, which is solved by semantic verifiable search scheme, proposed by Zhangjie et al [6]. This scheme not only supports keyword-based semantic search over encrypted data, but also provides verifiable search ability with data privacy preserving. This is a major contribution in searchable encryption field. But this scheme is not fully scalable and secure. To improve the scalability and provide more security to outsourced data of owner, a smart semantic search scheme with advance encryption technique is proposed in this paper. The contributions in this paper can be summarized as follows:

1) This paper proposes a smart keyword-based semantic search scheme over encrypted data. By building a semantic tree in real time according to original query terms, the scheme can find out some related words (including synonyms and various morphological forms, etc.) semantically similar to original query terms and then carry out query expansion. The expanded query can find more related result, thereby improving the flexibility of system.
2) When multiple user access the data on cloud, the request traffic is diverted to other subordinate clouds to achieve same performance and reduce the time required.
3) By combining the keyword-based semantic search scheme with verifiable symmetric searchable encryption, an efficient search scheme supporting verification of completeness and correctness of search result is proposed.
4) To make the scheme more secure and privacy preserving advanced encryption algorithm is used.

II. Related Work

A. Multi-clouds Computing Security

The use of cloud computing has increased rapidly in many organizations. Cloud computing provides many benefits in terms of low cost and accessibility of data. Ensuring the security of cloud computing is a major factor in the cloud computing environment, as users often store sensitive information with cloud storage providers but these providers may be untrusted. Dealing with "single cloud" providers is predicted to become less popular with customers due to risks of service availability failure and the possibility of malicious insiders in the single cloud. A movement towards "multi-clouds", or in other words, "inter-clouds" or "cloud-of-clouds" has emerged recently. Eric et al[7] given a recent research related to single and multi-cloud security and addresses possible solutions. It is found that the research into the use of multi-cloud providers to maintain security has received less attention from the research community than has the use of single clouds.

Multi-cloud in general is a collection of several cloud infrastructures. In single cloud infrastructure, if the environment is subjected to any type of attack then the data confidentiality is lost and may lead to loss of data. An example of Multi-cloud architecture is DepSky architecture. It is a combination of several different storage clouds. Therefore to secure the data in multi-cloud, Shamir proposed to store the data in more than one cloud and encrypt the same in the cloud before it transferred and saved. The goal of the algorithm is to divide the data **DATA** into **n** pieces (DATA1, DATA2, DATA3, DATA4 …..DATAn)

B. Searchable Encryption in Cloud

To apply the searchable encryption to cloud computing, some researchers have been studying further on how to search over encrypted cloud data efficiently. Li et al [9] firstly proposed a fuzzy keyword search scheme over encrypted cloud data. Wang et al [10] proposed a secure ranked search scheme. But this scheme supports only single keyword search. Then Cao et al [11] proposed a privacy-preserving ranked scheme supporting multi-keyword, which is lack of flexibility. And the fuzzy keyword search scheme proposed by Li [9] just tacks the problems of minor typos and format inconsistence, but does not meet the needs of users retrieving as more relevant data files as possible.

Chai et al [4] propose a verifiable search scheme, which can prove the correctness and completeness of result efficiently. However, there are some security problems that are not addressed properly in the paper. Based on VSSE and fuzzy keyword search, Wang et al [12] propose a scheme supporting both verification and fuzzy search, but the scheme ignores result ranking. Fu et al [5] proposed a multi-keyword search scheme in encrypted cloud environment which solved the problem of synonym query. This is a good step forward in the field of searchable encryption. However, this scheme has not addressed the problems of semantic search and verification of search results. This paper proposes a semantic search scheme over encrypted data by building a semantic tree in real time according to original query terms. The scheme is smarter than previous scheme, and can improve user experience.

III. SYSTEM FLOW

A. System Architecture

The system model considered in the paper involves three different entities: the data owner, the data user and the cloud server, as illustrated in Fig.1.

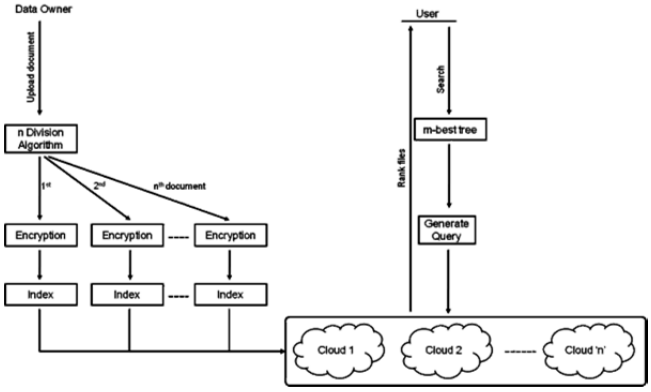

Fig: System Architecture of Semantic Search

B. Overview

1) When data owner uploads a file, system divides it as per the count of clouds, constructs a metadata for each file from collection of text files and outsources the encrypted metadata set to the private cloud servers. The text files are encrypted by using advanced encryption algorithm and uploaded to the public cloud server.

2) Private cloud server constructs the inverted index and semantic relationship library using metadata set provided by data user. Then the Inverted index is outsourced to the public cloud server for retrieval.

3) The authorized data users provide the search trapdoor to the private cloud server. Here, the authorization between the data owner and users is appropriately done.

4) Upon receiving the request, the private cloud server extends the query keyword upon semantic relationship library (SRL) and uploads the extended query keyword set to the public clouds.

5) Upon receiving the search request, the public cloud retrieves the index, and returns the matching files to the user in order.

6) The files are merged from the public clod severs and returned to user.

7) Finally, the access control mechanism is employed to manage the capability of the user to decrypt the received files.

C. Algorithm

- **Setup:** In this algorithm the data owner initiates the scheme to generate the random key and a secret key.
- **GenIndex:** To improve the search efficiency, a symbol-based tree to store elements in a finite symbol set is built.

  Pre-process:

  1) The data owner scans the plaintext document collection D and extracts the distinct keywords of D, denoted as W;

  2) The data owner computes the score of all distinct keywords on basis of presence in number of documents from collection.

  3) The data owner divides the file in to number of clouds and uploads it in encrypted form.
- **GenQuery:** When the user inputs the query terms Q, first builds term similarity tree TST (Q,v,m) and executes keyword semantic extension, getting the extended query.
- **Search:** Upon receiving the search request, the cloud server performs the search operation over the index G. The search is principally to find a path in G according to the search request, from the root node to the leaf node. The existence of a path indicates that the queried word includes in the targeted data files. The relevant files merged from different sever and returned to user.
- **Verify & Rank:** When the user receives the ranked outcome from the cloud server, he can verify the correctness and completeness of search result.

IV. CONCLUSION

In this paper, the semantic search scheme with more secured encryption technique over multi cloud is proposed. Comparing to most of the existing searchable encryption schemes, the proposed scheme is more practical and flexible, better suiting users' different search intentions. Moreover, the proposed scheme protects data privacy and supports verifiable search ability, in the presence of the semi-honest server in the cloud computing environment. Using multi-cloud the processing speed is decreased which returns the result to consumer in less time.

REFERENCES

[1] W. Noh and T. Kim, "Flexible communication-bus architecture for distributed multimedia service in cloud computing platform," *IEEE Trans. Consumer Electron.*, 2013.

[2] T. Shahnaza and Y. Kim, "Realistic IEEE 802.11 e EDCA model for QoSaware mobile cloud service provisioning," *IEEE Trans. Consumer Electron*, 2012.

[3] P. A. Cabarcos, F. A. Mendoza, R. S. Guerrero, A. M. Lopez, and D. Diaz-Sanchez, "SuSSo: seamless and ubiquitous single sign-on for cloud service continuity across devices," *IEEE Trans. Consumer Electron.*, 2012.

[4] Q. Chai and G. Gong, "Verifiable Symmetric Searchable Encryption for SemiHonest-but-Curious Cloud Servers," *Proceedings of IEEE International Conference on Communications (ICC'12)*, 2012.

[5] Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query," *IEEE Trans. Consumer Electron.*, 2014.

[6] Zhangjie Fu, Member, IEEE, Jiangang Shu, Xingming Sun, and Nigel Linge, "Smart Cloud Search Services: Verifiable Keyword-based Semantic Search over Encrypted Cloud Data," *IEEE Transactions on Consumer Electronics*, November 2014.

[7] Mohammed A. AlZain, Eric Pardede, Ben Soh, James A. ThomCloud, "Computing Security: From Single to Multi-Clouds," *Hawaii International Conference on System Sciences.*, 2012.

[8] Aakash Goplani, Jyoti Vaswani, Sneha Kukreja, Prof. Anjali Yeole, "A Review on Techniques for Searching and Indexing over Encrypted Cloud Data," *International Journal of Emerging Technology and Advanced Engineering.*, January 2015.

[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. J. Lou, "Fuzzy keyword search over encrypted data in cloud computing," *Proceedings of IEEE INFOCOM 2010*, San Diego, CA, USA, pp. 1-5, 2010.

[10] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure ranked keyword search over encrypted cloud data," *Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, pp. 253-262, 2010.

[11] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *Proceedings of IEEE INFOCOM 2011*, pp. 829-837, 2011.

[12] M. Muhil, U. Hemanth Kishra, R. Kishore Kumar, E. A. Mary, "Securing Multi-Cloud using Secret Sharing Algorithm," *2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)*, 2015.