International Journal of Research
Available at https://www.internationaljournalofresearch.org/
ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

Privacy-Preserving Multi keyword Similarity Search over Outsourced Cloud Data

[1]POLISHETTY PRIYANKA, [2]Dr VIJAY REDDY MADIREDDY

M.Tech, Department of Computer Science and Engineering,

Associate Professor, Department of CSE,

Swami Ramananda Tirtha Institute of Science & Technology, Nalgonda.

Abstract - Cloud computing provides individuals and enterprises massive computing power and scalable storage capacities to support a variety of big data applications in domains like health care and scientific research, therefore more and more data owners are involved to outsource their data on cloud servers for great convenience in data management and mining. However, data sets like health records in electronic documents usually contain sensitive information, which brings about privacy concerns if the documents are released or shared to partially untrusted third-parties in cloud. A practical and widely used technique for data privacy preservation is to encrypt data before outsourcing to the cloud servers, which however reduces the data utility and makes many traditional data analytic operators like keyword-based top-*k* document retrieval obsolete. In this paper, we investigate the multi-keyword top-*k* search problem for big data encryption against privacy breaches, and attempt to identify an efficient and secure solution to this problem. Specifically, for the privacy concern of query data, we construct a special tree-based index structure and design a random traversal algorithm, which makes even the same query to produce different visiting paths on the index, and can also maintain the accuracy of queries unchanged under stronger privacy. For improving the query efficiency, we propose a group multi-keyword top-*k* search scheme based on the idea of partition, where a group of tree-based indexes are constructed for all documents. Finally, we combine these methods together into an efficient and secure approach to address our proposed top-*k* similarity search. Extensive experimental results on real-life data sets demonstrate that our proposed approach can significantly improve the capability of defending the privacy breaches, the scalability and the time efficiency of query processing over the state-of-the-art methods.

Index Terms—mobile cloud computing, data encryption, access control, user revocation.

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

1. INTRODUCTION

Cloud computing has emerged as a disruptive trend in both IT industries and research communities recently its salient characteristics like high scalability and pay-asyou-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management [1], [2]. Nowadays, more and more companies and individuals from a large number of big data applications have outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks.

But when the companies and individuals enjoy these advantages in cloud computing, they also need to take the privacy concern of the outsourced data into account. Because data sets in many applications often contain sensitive information like e-mails, electronic health records and financial transaction records, when the data owner outsourcing such sensitive data to the cloud servers which are considered to be partially trusted, the data can be easily accessed and analyzed by cloud service providers illegally. Since the analysis of these data sets may provide profound insights into a number of key areas in society (such as e-research, healthcare, medical and government services), thus data owners need effective, scalable and privacy-preserving services before releasing their data to the cloud. Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plaintext into cyphertext, which is a non-readable form to unauthorized parties. A variety of data encryption models have been proposed [3], [4], [5] and they are used to encrypt the data before outsourcing to the cloud servers. However,

applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plaintext data no longer work well over encrypted data. The keyword-based search is such one widely used data operator in many database and information retrieval applications, and its traditional processing methods cannot be directly applied to encrypted data. Therefore, how to process such queries over encrypted data and at the same time guarantee data privacy becomes a hot research topic. Fortunately, many methodologies based on searchable.

2. Literature Survey

Ensuring Security and Privacy Preservation for Cloud Data Services

We first present security threats and requirements of an outsourcing data service to a cloud, and follow that with a high-level overview of the corresponding security

technologies. We then dwell on existing protection solutions to achieve secure, dependable, and privacy-assured cloud data services including data search, data computation, data sharing, data storage, and data access. Finally, we propose open challenges and potential research directions in each category of solutions.

Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions

In this paper we begin by reviewing existing notions of security and propose new and stronger security definitions. We then present two constructions that we show secure under our new definitions. Interestingly, in addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions.

Further, prior work on SSE only considered the setting where only the owner of the data is capable of submitting search queries. We consider the natural extension where an arbitrary group of parties other than the owner can submit search queries. We formally define SSE in this multi-user setting, and present an efficient construction

Public Key Encryption with keyword Search

We define and construct a mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word "urgent" is a keyword in the email without learning anything else about the email. We refer to this mechanism as Public Key Encryption with keyword Search. As another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using our mechanism Alice can send the mail server a key that will enable the server to identify all messages containing some specific keyword, but learn nothing else. We define the concept of public key encryption with keyword search and give several constructions.

Practical Techniques for Searches on Encrypted Data

In this paper, we offer solutions for this problem under well-defined security requirements. Our schemes are efficient in the sense that no public-key cryptosystem is involved. Indeed, our approach is independent of the encryption method chosen for the remote files. They are also incremental, in that UU can submit new files which are secure against previous queries but still searchable against future queries.

3. OVERVIEW OF THE SYSTEM

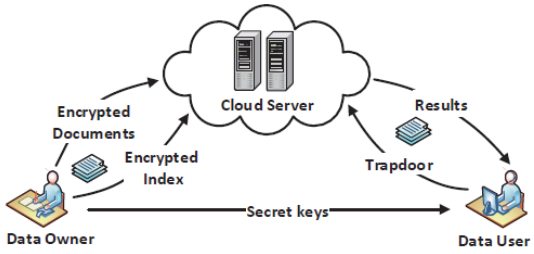

Fig 3.1 System Architecture

3.1 EXISTING SYSTEM:

- The keyword-based search is such one widely used data operator in many database and information retrieval applications, and its traditional processing methods cannot be directly applied to encrypted data. Therefore, how to process such queries over encrypted data and at the same time guarantee data privacy becomes a hot research topic.

- Song et al. first defined the problem of searching on encrypted data and proposed a symmetric searchable encryption scheme with linear complexity.

3.2 DISADVANTAGES OF EXISTING SYSTEM:

Most of these methods cannot meet the high search efficiency and the strong data security simultaneously, especially when applying them to big data encryption poses great scalability and efficiency challenges

3.3 PROPOSED SYSTEM:

- In this paper, we focus on a special type of multi-keyword ranked search, namely the multikeyword top-*k* search, which has been a very popular database operator in many important applications, and only needs to return the *k* documents with the highest relevance scores.
- We propose a group multi-keyword top-*k* search scheme (GMTS), which is based on partition and supports top-*k* similarity search over encrypted data.
- We propose a random traversal algorithm (RTRA) to strengthen the data security, where the data owner builds a binary tree as searchable index and assigns a random switch to each node, so the data user can assign a random key to each query.

3.4 ADVANTAGES OF PROPOSED SYSTEM:

- Improving the efficiency and the security of multi-keyword top-*k* similarity search over encrypted data.
- Data user receives different results but with the same high level of query accuracies in the mean time.
- Experimental results show that our methods are more efficient and more secure than the state-of-the-art methods.

3.5 IMPLEMENTATION

MODULES:

Data owner:

In the first module, we develop the Data Owner Module. Owner Will Signup and Wait for the approval from cloud server. After authentication is successful Owner can login, and upload files with encrypted index

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

using hash code along with encrypted keyword and send file to cloud service provider.

In this module, data owner can check uploaded file details and downloaded file details. Owner can check requests received from users who are requesting for file permissions to download file. Owner can give permissions by sending decryption key to requested user mail id.

Data User:

In this module, we develop the User Module. User Will registries and Wait for the approval from cloud server and login on the user's page. We develop the module, such that, User will search with keyword and keyword in encrypted and send to cloud server to get all related files which are matched with same keyword and get tapdoor (hashcode). User will enter hashcode to access all top k similar files. User will send download file request to data owner and receive decryption key to user mail id.

Cloud Server:

CSP stores the data for data owner. It faithfully executes the operations requested by data owner and data user by activating and deactivating users and sending files which are requested by users, when data user requests for hashcode for requested files based on keyword CSP will send hashcode to user mail id.

4. SYSTEM DESIGN

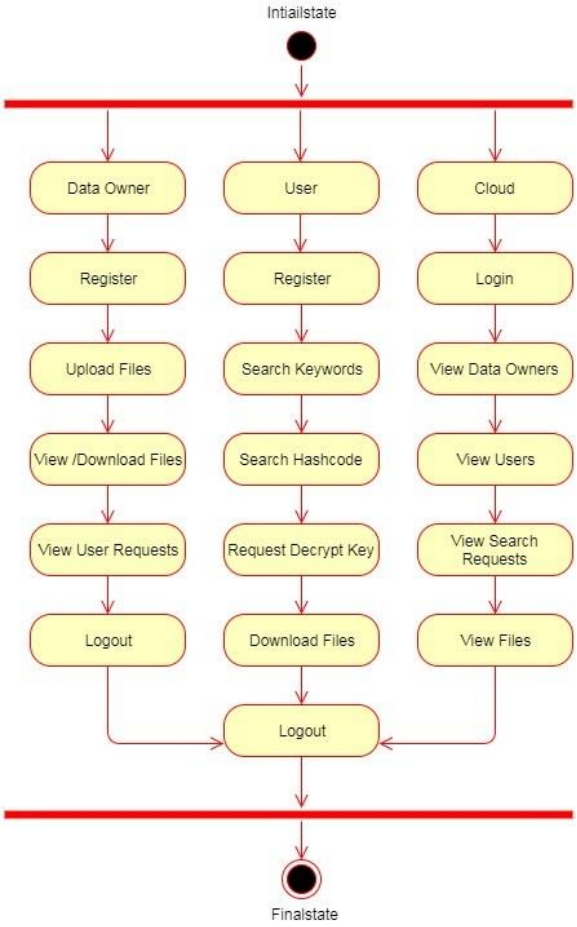

Fig 4.1: Activity Diagram

5. OUTPUT SCREEN SHOTS

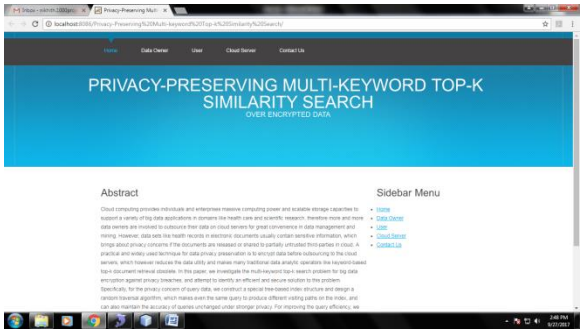

Fig 5.1: Home Page

International Journal of Research
Available at
https://www.internationaljournalofresearch.org/
ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

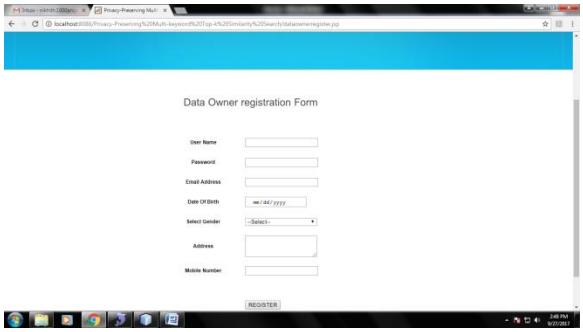

Fig 5.2: Data Owner registration Page

Fig 5.3: User Login Page

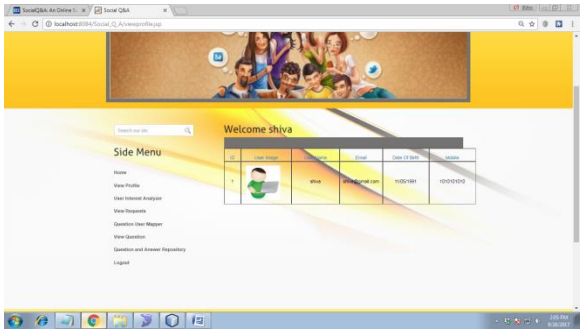

Fig 5.4: View Profile Page

Fig 5.5: User Interest Analyzer Page

6. CONCLUSION AND FUTURE SCOPE

In this paper, we focus on improving the efficiency and the security of multi-keyword top-*k* similarity search over encrypted data. At first, we propose the random traversal algorithm which can achieve that for two identical queries with different keys, the cloud server traverses different paths on the index, and the data user receives different results but with the same high level of query accuracies in the mean time. Then, in order to improve the search efficiency, we design the group multi-keyword top-*k* search scheme, which divides the dictionary into multiple groups and only needs to store the top-*ck* documents of each word group when building index. Next, to protect the query unlinkability, we apply the random traversal algorithm to get the RGMTS, which can increase the difficulty of cloud servers to conduct linkage attacks on two identical queries, and we can also tune the value of *E* to make the level of query unlinkability flexible for data owners. Finally, the experimental results show that our methods

are more efficient and more secure than the state-of-the-art methods.

7. REFERENCES

[1] J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, "Ensuringecurity and privacy preservation for cloud data services," *ACMComputing Surveys*, 2016.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski,G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia,"A view of cloud computing," *Communications of the ACM*, vol. 53,no. 4, pp. 50–58, 2010.

[3] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchablesymmetric encryption: Improved definitions and efficient constructions,"in *Proceedings of the 13th ACM Conference on Computerand Communications Security*. ACM, 2006, pp. 79–88.

[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Publickey encryption with keyword search," in *Advances in Cryptology-Eurocrypt 2004*. Springer, 2004, pp. 506–522.

[5] Z. Ying, H. Li, J. Ma, J. Zhang, and J. Cui, "Adaptively secureciphertext-policy attribute-based encryption with dynamic policyupdating," *Sci China InfSci*, vol. 59, no. 4, pp. 042 701:1–16, 2016.

[6] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques forsearches on encrypted data," in *Security and Privacy, 2000. SP 2000.Proceedings. 2000 IEEE Symposium on*, 2000, pp. 44–55.

[7] E.-J. Goh*et al.*, "Secure indexes." *IACR Cryptology ePrint Archive*,vol. 2003, p. 216, 2003.