IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

Privacy in Searchable Symmetric Encrypted Cloud Data using Ranked Search

Nutan Pramod Chaudhari[1], Dr. Dinesh D. Patil[2], Prof. Rahul P. Chaudhari[3]
Department of Computer Science and Engineering[1]
Head of Department, Department of Computer Science and Engineering[2]
Associate Professor, Department of Computer Science and Engineering[3]
Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal, Maharashtra, India

Abstract: *The appearance of cloud computing, data owners are motivatedfor great flexibility and economic savings to outsource their complex data management systems from local sites to commercial public cloud. For protecting data privacy, to ensure adequate sensitive data has 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. To achieve more competence, ranked searchable symmetric encryption is a cryptography scheme which gives an efficient design by properly utilizing the existing cryptographic primitive it is known as order-preserving symmetric encryption (OPSE). The proposed solution enjoys "as-strong-as possible" security guarantee compared to previous SSE schemes, while correctly realizing the goal of ranked keyword search.*

Keywords: Ranked Search, Encrypted Cloud, Privacy Preserving Data, Order-Preserving Symmetric Encryption

I. INTRODUCTION

In today's information technology, customers that need high storage and computation power usually outsource their data and services to clouds. Clouds enable customers to remotely store and access their data by lowering the cost of hardware ownership while providing wholesome and fast services. The essential requirement of privacy preserving search techniques are even more pronounced in the cloud applications. Hence large companies that operate the public clouds like Google or Amazon may access the sensitive data and search patterns, hiding the query and the retrieved data has great importance in ensuring the privacy and security of those using cloud services. The another continuing vision of cloud computing is that IT services are traded as a utilities in an open market without technological and legal barriers, where cloud customers can remotely store their data into the cloud due to on-demand high quality applications and services, broad network access, and being very elastic and scalable from a shared pool of configurable computing resources. The benefits of new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, and personnel maintenances, etc. More and more sensitive information are being centralized into the cloud, such as emails, personal health records, company finance data, and government documents, etc. for this reason Cloud Computing becomes prevalent. The fact that data owners and cloud server are no longer in the same trusted domain may put deploy unencrypted data at risk. The cloud server may leak data information to unauthorized entities or may be hacked. The sensitive data has to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses.

1.1 Aim

To achieve an efficient system where any authorized user can perform a search on a remote database with multiple keywords, without revealing neither the keywords he searches for, nor the contents of the documents he retrieves. The proposed system differs from the previous works which assume that only the data owner queries the database. Moreover, the proposed system is able to perform multiple keyword search in a single query and ranks the results so the user can retrieve only the top matches.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

1.2 OBJECTIVE

- It provide formal definitions for the safe and privacy requirements of keyword search on encrypted cloud data
- In conformity with the defined requirements.
- To propose a ranking method that proves to be efficient to implement and effective in returning documents highly relevant to submitted search terms
- And finally implement the proposed scheme and demonstrate that it is much more efficient than existing methods in literature.

1.3 Design Goals

- **Multi-keyword Ranked Search:** To design search schemes which allow multi-keyword query and provide result resemblance ranking for effective data retrieval, instead of returning indeterminate results
- **Privacy-Preserving:** To prevent the cloud server from learning additional information from the dataset and the index, and encounter privacy requirements.
- **Efficiency**: Above goals on functionality and privacy should be achieved with low communication and computation expenses.

II. LITERATURE REVIEW

Any PIR-based technique requires highly costly cryptographic operations in order to hide the access patternwas first proposed by Chor et al. This is ineffective in the large scale cloud system and as an alternative approach, privacy preserving search is employed which aims to hide the content of the retrieved data instead of which data is retrieved Ogata and Kurosawa show privacy preserving keyword search protocol based on RSA blind signatures. It requires a public key operation per item in the database for every query and this operation must be performed on the user side. An alternative implementation for private keyword search that uses homomorphic encryption and oblivious polynomial evaluation methods proposed by Freedman et al. The computation and communication costs of both the server and the user side method are quite large since every search term in a query requires several homomorphic encryption operations. A latter search proposed by Wang et al allows ranked search over an encrypted database by using inner product similarity. This work is only finite to single keyword search queries. One of the closest methods to our solution is proposed by Cao et al. Similarly, it proposes a method that allows multi-keyword ranked search over encrypted database. In this method, the data owner needs to distribute a symmetric key which is used in trapdoor generation to all authorized users. Further, this work requires keyword fields in the index. This means that the user must know a list of all valid keywords and their positions as compulsory information lead to a query. This supposition may not be applicable in several cases. Moreover, it is ineffectual due to matrix multiplication operations of square matrices where the number of rows is in the order of several thousands. Wang et al suggest a trapdoor less private keyword search scheme, where their model requires a trusted third party which they named as the Group Manager. We adapt their indexing method to our scheme, but we use completely different encryption methodology to increase the security and efficiency of the scheme.

III. METHODOLOGY

To search index files for the documents in the database are generated by the data owner using secret keys. A user who wants to include a search term in his query, needs the corresponding trapdoor from the data owner since he does not know the secret keys used in the index generation. Asking for the trapdoor openly would violate the privacy of the user against the data owner, therefore a technique is needed to hide the trapdoor asked by the user from the data owner. There are various modules used in the system for the multi keyword ranked search.

3.1 Encrypt Module

This module is used to help the server to encrypt the document using Symmetric encryption Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

3.2 Client Module

This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required file and register the user details and get activation code in mail from the "customerservice404" email before enter the activation code. After user can download the Zip file and extract that file.

3.3 Multi-keyword Module

This module is used to help the user to get the accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list.

3.4 Admin Module

This module is used to help the server to view details and upload files with the security. Admin uses the log key to the login time. Before the admin logout, change the log key. The admin can change the password after the login and view the user downloading details and the counting of file request details on flowchart. The admin can upload the file after the conversion of the Zip file format.

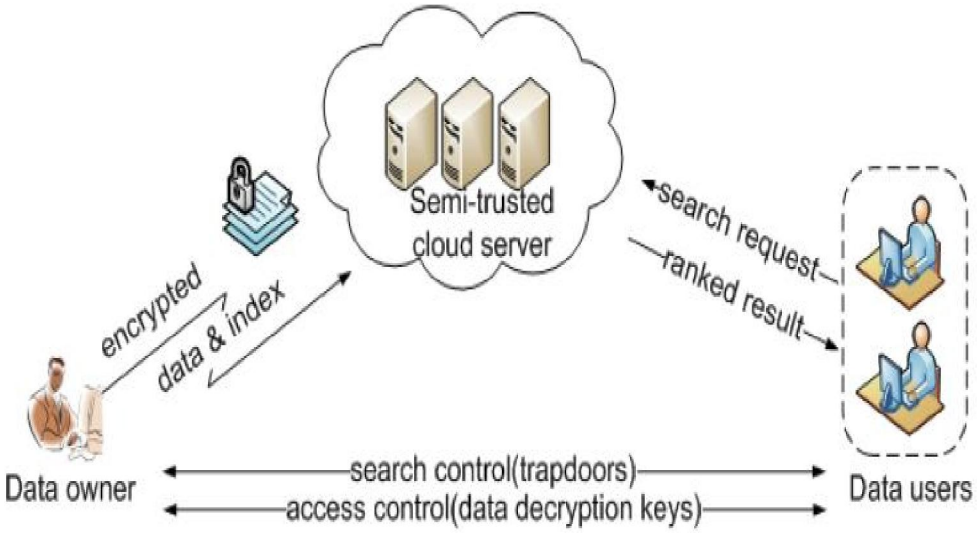

Figure 1: Architecture Diagram[5]

IV. PRIVACY REQUIREMENT

The representative privacy guarantee in the related literature, such as searchable encryption, is that the server should learn nothing but search results. With this general privacy description, we explore and establish a set of strict privacy requirements specifically for the MRSE framework. The data owner can resort to the traditional symmetric key cryptography to encrypt the data before outsourcing, and successfully prevent the cloud server from prying into the outsourced data as for the data privacy. With respect to the index privacy, if the cloud server deduces any association between keywords and encrypted documents from index, it may learn the major subject of a document, even the content of a short document. Hence, the searchable index should be constructed to prevent the cloud server from performing such kind of association attack, while data and index privacy guarantees are demanded by default in the related literature, various search privacy requirements involved in the query procedure are more complex and tough to tackle as follows. Keyword Privacy As users usually prefer to keep their search from being exposed to others like the cloud server, the primary concern is to hide what they are searching, i.e., the keywords indicated by the corresponding trapdoor. Despite the trapdoor can be generated in a cryptographic way to protect the query keywords, the cloud server could do some statistical analysis over the search result to make an estimate. As a kind of statistical information, document frequency (i.e., the number of documents containing the keyword) is sufficient to identify the keyword with high probability .When the cloud server knows some background information of the dataset, this keyword specific

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

information may be utilized to reverse-engineer the keyword. Trapdoor Unlink ability the trapdoor generation function should be a randomized one instead of being deterministic. In particular, the cloud server should not be able to deduce the relationship of any given trapdoors, e.g., to determine whether the two trapdoors are formed by the same search request. Otherwise, the deterministic trapdoor generation would give the cloud server advantage to accumulate frequencies of different search requests regarding different keyword(s), which may further violate the aforementioned keyword privacy requirement. So the fundamental protection for trapdoor unlink ability is to introduce sufficient non determinacy into the trapdoor generation procedure. Access Pattern Within the ranked search, the access pattern is the sequence of search results where every search result is a set of documents with rank order. Specifically, the search result for 10 the query keyword set fW is denoted as FfW, consisting of the id list of all documents ranked by their relevance to fW. Then the access pattern is denoted as (FfW1, FfW2 . . .) which are the results of sequential searches. Although a few searchable encryption works, has been proposed to utilize private information retrieval (PIR) technique , to hide the access pattern, our proposed schemes are not designed to protect the access pattern for the efficiency concerns. This is because any PIR based technique must "touch" the whole dataset outsourced on the server which is inefficient in the large scale cloud system.

V. PROPOSED SYSTEM

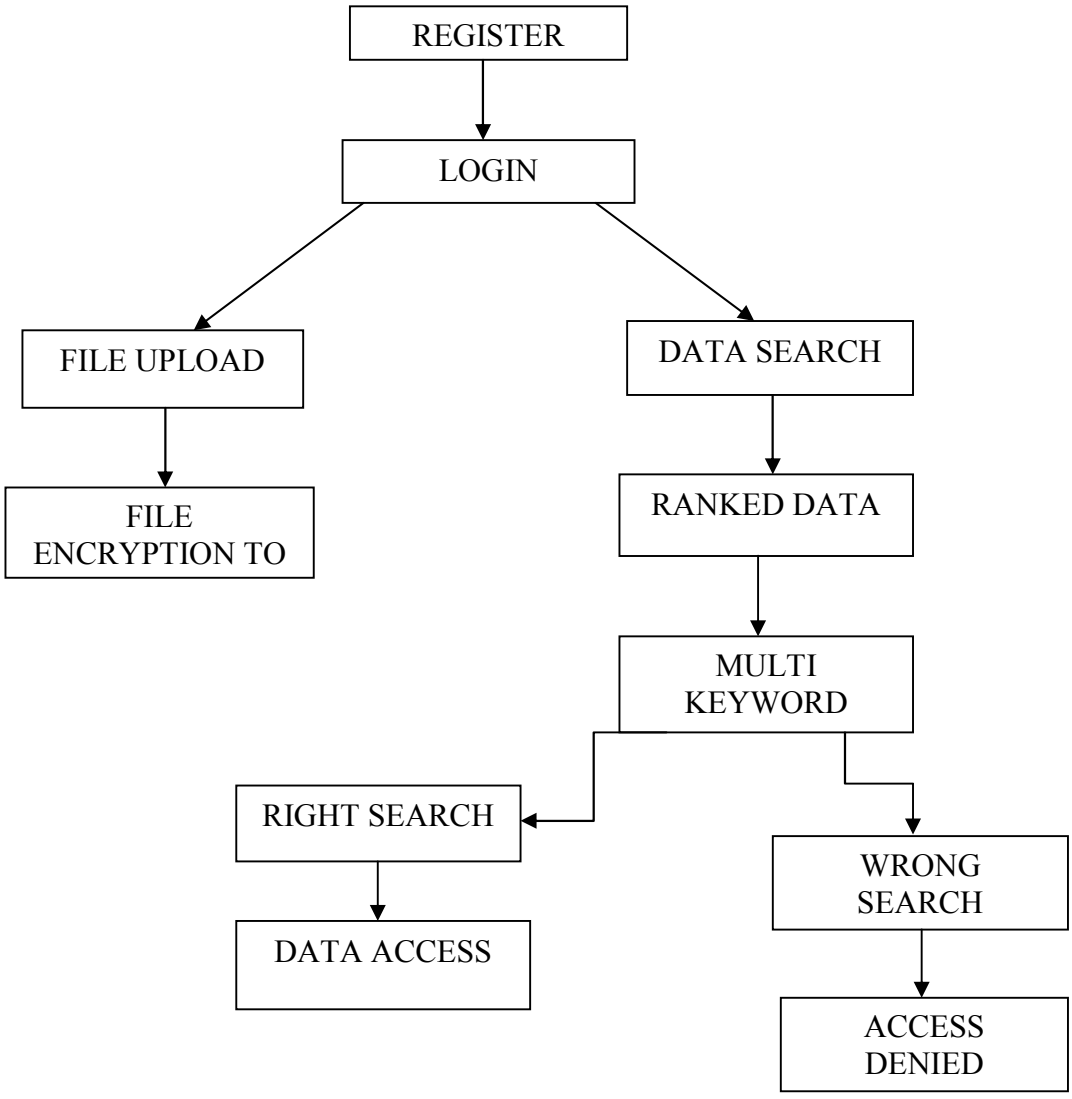

Figure 2: Flow Diagram of proposed System

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

Cloud data hosting service involving three different entities, as shown in Fig. 1: the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C which is enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index I from F, and then outsource both the index I and the encrypted document collection C to the cloud server to search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms. After receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. The search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching, as will be introduced shortly) which improve the document retrieval accuracy. Besides, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query and at lastthe access control mechanism is employed to manage decryption capabilities given to users [3].

IV. SYSTEM TESTING

There are various types of test. Each test type addresses a specific testing requirement.

4.1 System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is to configuration oriented system integration test. It is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

4.2 White Box Testing:

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or atleastitspurpose.ItisusedtotestareasthatcannotbereachedfromablackboxlevelBox level.

4.3 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

V. CONCLUSION

This study focuses on the problem of multi-keyword ranked search over encrypted cloud data established variety of privacy requirements. The efficient principle of “coordinate matching”, i.e., as much as possible, to effectively capture similarity between query keywords and outsourced documents, and use “inner product similarity” to quantitatively formalize such a principle for similarity measurement. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, firstly propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve privacy requirements in two levels of threat models.

REFERENCES

[1]. H. Dai, Y. Ji, L. Liu, G. Yang and X. Yi, "A privacy-preserving multi-keyword ranked search over encrypted data in hybrid clouds", Proc. 5th Int. Conf. Artif. Intell. Secur. (ICAIS), pp. 68-80, 2019.

[2]. Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data” IEEE TRANSACTION ON PARELLEL AND DISTRIBUTED SYSTEMS,VOL.25,NO 1,JANUARY2014.

[3]. Secure Access of Encrypted Cloud Data Based on Top-K Multikeywords with User Side RankingIJERTCONV3IS19214, Volume & Issue : ICESMART – 2015 (Volume 3 – Issue 19)

[4]. S. Kamara and K. Lauter, “Cryptographic cloud storage,” in RLCPS, January 2010, LNCS. Springer, Heidelberg Chengwei Liu, Yixiang Chan, Syed HasnainAlamKazmi, Hao Fu, “Financial Fraud Detection Model: Based on Random Forest,” International Journal of Economics and Finance, Vol. 7, Issue. 7, pp. 178-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, September 2022
Impact Factor: 6.252

188, 2015.

[5]. S. Grzonkowski, P. M. Corcoran and T. Coughlin, "Security analysis of authentication protocols for next-generation mobile and CE cloud services", Proc. IEEE Int. Conf. Consum. Electron., pp. 83-87, Sep. 2011.

[6]. N. Cao, M. Li and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", Proc. IEEE INFOCOM, pp. 829-839, Apr. 2011.

[7]. N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy–preserving multi–keyword ranked search over encrypted cloud data", IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222-233, Jan. 2014.

[8]. J. Xu, W. Zhang, C. Yang, J. Xu and N. Yu, "Two-step-ranking secure multi–keyword search over encrypted cloud data", Proc. Int. Conf. Cloud Service Comput., pp. 124-130, Nov. 2012.

[9]. C. Yang, W. Zhang, J. Xu, J. Xu and N. Yu, "A fast privacy–preserving multi-keyword search scheme on cloud data", Proc. Int. Conf. Cloud Service Comput., pp. 22-24, Nov. 2012.

[10]. H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou and X. S. Shen, "Enabling fine–grained multi–keyword search supporting classified sub–dictionaries over encrypted cloud data", IEEE Trans. Dependable Secure Comput., vol. 13, no. 3, pp. 312-325, May 2016.

[11]. Z. Xia, X. Wang, X. Sun and Q. Wang, "A secure and dynamic multi–keyword ranked search scheme over encrypted cloud data", IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 2, pp. 340-352, Feb. 2016.

[12]. Z. Xiangyang, D. Hua, Y. Xun, Y. Geng and L. Xiao, "MUSE: An efficient and accurate verifiable privacy-preserving multikeyword text search over encrypted cloud data", Secur. Commun. Netw., vol. 2017, Nov. 2017.

[13]. M. Bellare, A. Boldyreva, and A. ONeill, “Deterministic and efficiently searchable encryption,” in Proc. of CRYPTO, 2007

[14]. L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A break in the clouds: towards a cloud definition,” ACM SIGCOMM Comput. Commun. Rev., vol. 39, 2009

[15]. R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions,” in Proc. of ACM CCS, 2006.