Privacy in Searchable Symmetric Encrypted Cloud Data using Ranked Search

Junaid Rafiq , Professor V.K. Sharma

BHAGWANT UNIVERSITY, AJMER

ABSTRACT

Data owners are encouraged to relocate their intricate data management systems from their private locations to the public domain in order to take advantage of the enormous flexibility and cost savings offered by cloud computing, as it is growing in popularity. Outsource to the cloud. Before outsourcing, it is essential to ensure that a significant amount of sensitive data is encrypted. This assurance is important for the purpose of protecting the privacy of the data. The previous approach of data utilization, which is focused on searching for keywords in plaintext, is rendered ineffective as a result of this. As a consequence of this, the availability of a cloud-based data search service that is encrypted is of the best possible importance. Efficient searchable symmetric encryption is a cryptography system that offers an efficient design by correctly using the cryptographic primitive that is already in existence. In order to get better at what you do, you have to do this. The term "order-preserving symmetric encryption" (OPSE) refers to this particular encryption method. In comparison to previous SSE methods, the methodology that has been offered offers a security guarantee that is "as strong as possible," while at the same time making it possible to achieve the goal of ranking keyword search in an accurate way.

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

INTRODUCTION

Consumers that want a large amount of storage and computation capacity generally outsource their data and services to cloud computing, according to the present state of information technology. Cloud computing is a phenomenon that has emerged in recent years. Customers have the ability to remotely store and retrieve their data through the utilization of cloud computing, which not only diminishes the cost of hardware ownership but also provides services that are both wholesome and rapid. Particularly when it comes to cloud applications, the need of utilizing search strategies that safeguard the privacy of users becomes even more apparent. In order to protect the confidentiality and integrity of cloud service users, it is crucial to guarantee the secrecy of both the query and the returned data. Reason being, big IT companies like Amazon and Google that provide public cloud services could have special access to private information and search patterns. It is of the utmost importance to conceal both the query and the data that is obtained. Another long-term goal of cloud computing is to remove legal and technological barriers to the open market exchange of IT services, similar to the way utilities are traded. Cloud computing will keep bringing this idea to fruition. Thanks to the abundance of high-quality applications and services, the broad availability of networks, and the scalability and flexibility of cloud computing—which uses a shared pool of programmable computer resources—users of cloud computing are able to remotely store their data in the cloud. New computing models offer various advantages, such as decreased responsibility for managing storage, ubiquitous data access across different geographic locations, and the elimination of capital expenses associated with hardware and personnel maintenance, among other benefits. The rising volume of sensitive information that is being pooled into the cloud is one of the primary reasons why cloud computing is becoming more and more prevalent. This encompasses a wide range of information, including but not limited to emails, personal health records, data pertaining to company finances, and documents from the government, among other things. Due to the fact that the cloud server and the owners of the data are no longer in the same trusted domain, there is a chance that the deployment of data that is not encrypted might be impacted. There is a possibility that the cloud server may be hacked, or that it could accidentally reveal

sensitive data to those who are not authorized to receive it. Before the sensitive information is outsourced, it must first be encrypted in order to safeguard the confidentiality of the data and to prevent unwanted access to the information.

Aim

To provide a productive system that allows any authorized user to search remote databases using a variety of keywords without disclosing the keywords he uses or the information in the documents he searches. Compared to earlier studies, which assumed that the database was queried only by the data owner, the suggested approach is different. Additionally, the suggested method prioritizes the results so that the user can get only the best matches and search multiple keywords in a single query.

OBJECTIVE

- It provides 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.

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.

LITERATURE REVIEW

The idea that any PIR-based system would first require exceedingly expensive cryptographic operations in order to disguise the access pattern was first presented by Chor et al., who were the ones who initially established the concept by themselves. Because this approach does not work in cloud computing environments that are on a vast scale, an alternative method that is known as privacy-preserving search is now being deployed. As opposed to the data itself, the goal of this search strategy is to conceal the information that is included inside the data that is retrieved. Ogata and Kurosawa presented a method for doing keyword searches that is based on RSA blind signatures and protects users' anonymity. This method was demonstrated at the conference. A public key action must be performed on the user's end in conjunction with each and every query. This activity must be carried out individually. Execution of this action is required for each and every item that is contained inside the database. An alternative solution for private keyword search makes use of the homomorphic encryption and oblivious polynomial evaluation methods that were given by Freedman et al. for the sake of confidentiality. Due to the need for several homomorphic encryption operations for each search word in a query, both the server-side technique and the user-side method incur significant computation and communication overhead. This is due to the fact that each search word in a query necessitates a certain number of homomorphic encryption operations. A follow-up search approach proposed by Wang et al. makes use of inner product similarity to provide ranked search over an encrypted database. This search was carried out in order to get the most relevant results. This endeavor is restricted to searches that involve a single phrase, and its reach is confined to such searches. The method that Cao et al. provide is one of the approaches that is one of the ways that is the most similar to our solution. In addition to this, it proposes a system that makes it possible to execute a multi-keyword ranked search across a database that is encrypted. For the purpose of properly implementing this strategy, the data owner is required to disseminate a symmetric key to all of the users who are permitted to access the system. This key is employed in the manufacturing of trapdoors. Furthermore, in

order to do this task, the index has to include keyword fields. In light of this, it is essential for the user to be familiar with a list of all the keywords that are permitted to be used, as well as the locations of those keywords as obligatory information that leads to a query. However, there are a variety of circumstances in which it is feasible that this presumption might not apply. Matrix multiplication operations on square matrices become inefficient as the number of rows beyond a few thousand. These operations are considered to be ineffective. According to Wang et al.'s proposal, a trapdoor less private keyword search approach is being developed. It is essential to take into consideration that their strategy calls for the participation of a trustworthy third party, whom they have designated as the Group Manager. We tweak their indexing technique so that it is compatible with our system, but we use a completely new encryption mechanism in order to make the scheme as safe and as efficient as it can possibly be.

METHODOLOGY

The data owner uses secret keys to produce index files that are used to search for documents in the database. Because he does not know the secret keys used to build the index, the user needs the data owner's matching trapdoor in order to incorporate a search phrase in his query. A method is required to conceal the trapdoor that the user has requested from the data owner as asking for it openly would breach the user's privacy against the owner of the data. The system uses a number of modules to provide multi-keyword ranked search.

Encrypt Module

This module facilitates server-side encryption of documents using the symmetric encryption technique. It then transforms the encrypted documents into a Zip file, which is accompanied by an activation code. The activation code is subsequently sent to the user for download.

Client Module

This module allows the user to search the file with several key words and receive an accurate list of results based on their query. The user must select the required file, register, and get an email from the "customerservice404" email address with an activation code before inputting the code. The user will be able to extract the file after downloading the Zip file.

Multi-keyword Module

This module is designed to assist the user in obtaining precise results by using several keyword ideas. Users have the ability to enter queries consisting of several words. The server will consolidate these words into a single word query once it has searched for each word in our database. Lastly, show the user the database's matched word list so they may select the file from it.

Admin Module

The server may examine details and upload files with security thanks to this module. The administrator uses the log key to document the precise moment of login. Modify the log key prior to the administrator's logout. After logging in, the administrator possesses the capability to view the user's download history and the specifics of each file request shown on a flowchart, as well as modify the password. Once the Zip file format has been converted, the administrator can upload the file.

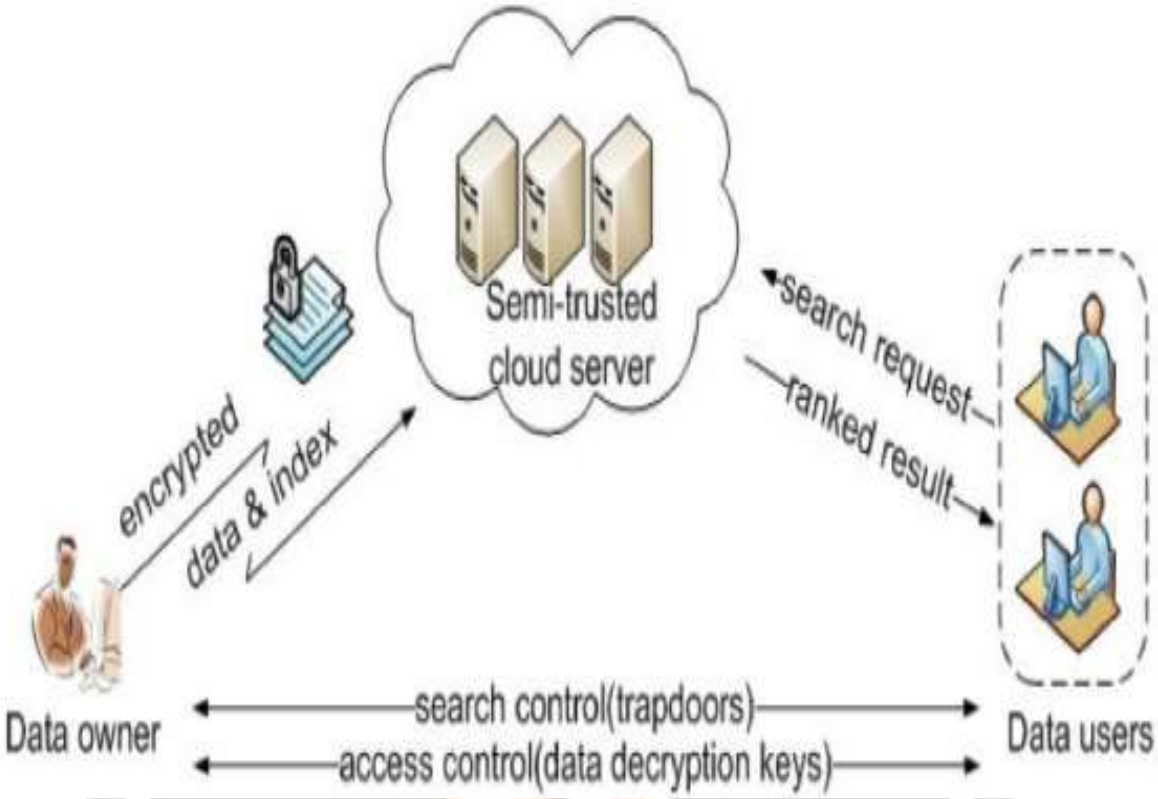

Figure 1: Architecture Diagram

PRIVACY REQUIREMENT

In the literature on the subject, searchable encryption is used as an example of a promise of privacy. It says the server should not learn anything other than search results. Keeping this broad idea of privacy in mind, we include stringent privacy standards into the MRSE architecture. The data owner can prevent unauthorized access to sensitive information while outsourcing by encrypting the data using a standard symmetric key method before sending it to the cloud. By examining the index for correlations between keywords and encrypted documents, the cloud server might potentially discern the subject matter or even the substance of a short phrase, jeopardizing the security of the index. The existing research emphasizes the significance of default measures to protect data and index privacy; yet, the overwhelming number of requests to maintain search privacy throughout the query process is difficult to handle. It is crucial to construct the searchable index in a way that prevents this type of association attack from being executed by the cloud server. Key Word Secrecy Users frequently like to keep their searches secret and concealed from third parties, such as the cloud server, thus the primary objective is to hide the keywords that they are searching for, as shown by the appropriate trapdoor. To acquire an estimate, the cloud server may do statistical analysis on the search result, even if a cryptographic trapdoor is used to secure the query terms. Document frequency, which refers to the number of documents that include a specific phrase, may be used to find keywords that are very likely to occur. It is feasible to reverse-engineer a word if the cloud server has prior knowledge of the dataset. This information is particularly relevant to the keyword. Ability to unlink trapdoors: The trapdoor generating function must be random rather than deterministic. The cloud server cannot establish a connection between two trapdoors, for instance by ascertaining that they were produced by the identical search query. Without deterministic trapdoor generation, the cloud server might gather search query frequencies linked to different keywords, thus breaking the requirement for keyword privacy. Therefore, adding enough non-determinacy to the trapdoor manufacturing process is essential to limiting the ability to unlink trapdoors. Accessing Data Method The "access pattern" defines the format of the results in a ranked search, where each result is a collection of documents arranged in a certain sequence. A search for the query set fW, specifically for page number 10, yielded a list of page IDs sorted by relevance to fW. We call this group of names FfW. Over time, the access pattern—which resembles

(FfW1, FfW2,...)—displays successive search results. When contrasted with prior searchable encryption strategies that utilize the private information retrieval (PIR) methodology to conceal the access pattern, our suggested solutions fail miserably. This is because, in a large-scale cloud system, any PIR-based technique necessitates "touching" the full server-outsourced dataset, which is wasteful.

PROPOSED SYSTEM

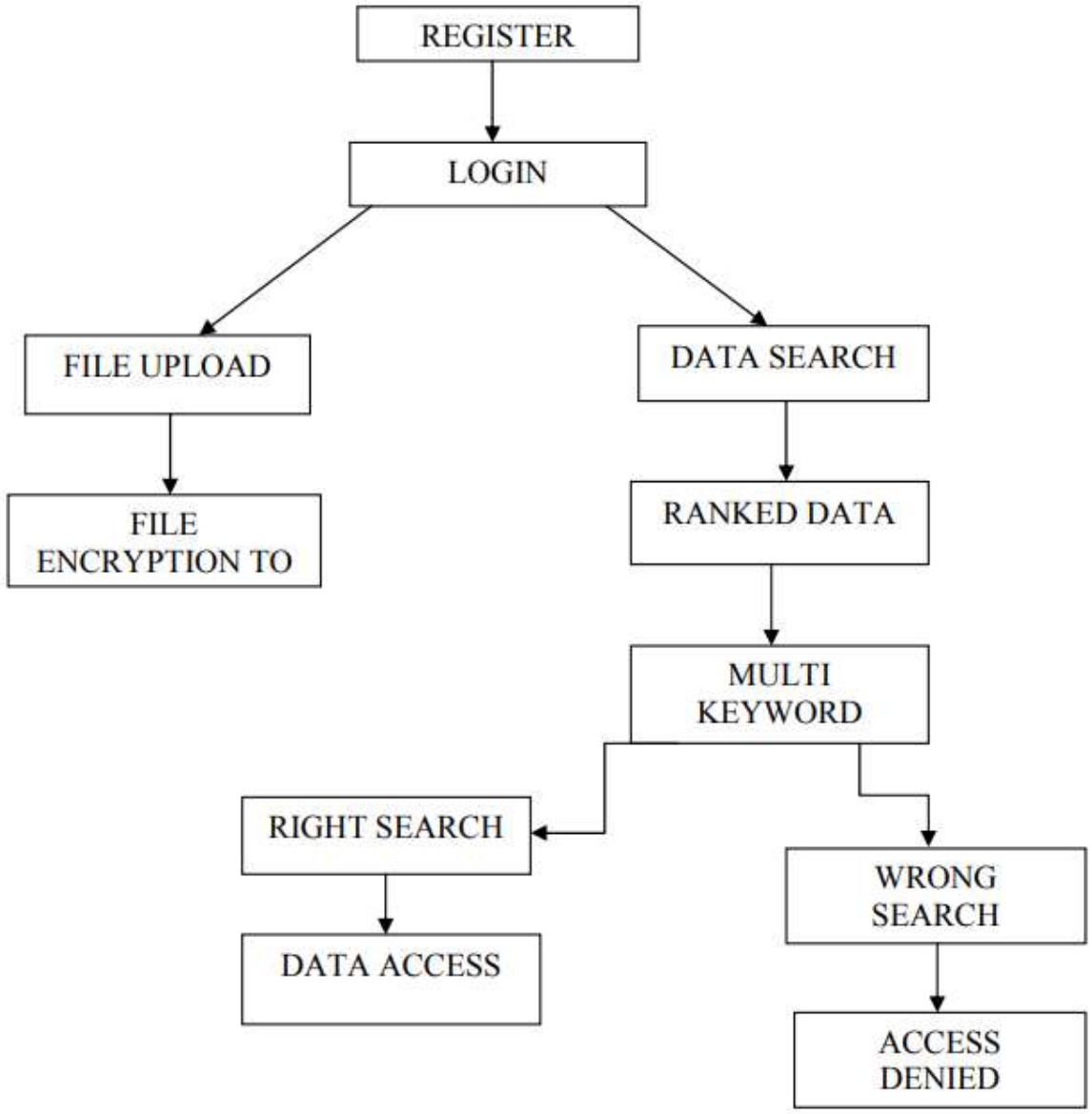

Figure 2: Flow Diagram of proposed System

The three parts that comprise a cloud data hosting service are shown in Figure 1: Users, data owners, and servers are the three main players in a cloud computing situation. The data owner will use the I-from-F technique to build a safe, searchable index before outsourcing. Once the data owner has completed step one, they can send the encrypted document collection C and index I to the server in the cloud. Here, the user can enter a set of keywords to search the document collection. Through search control measures, an authorized user will then be able to access a matching trapdoor T. The encrypted collection of data documents F will be sent to the cloud server, enabling effective data use by cross-referencing inside the encrypted form C. The cloud server's role is to retrieve index I and deliver the

accompanying encrypted documents when it receives T from a data user. To improve the accuracy of document retrieval, the cloud server should prioritize search results based on certain ranking criteria, such as coordinate matching, which will be explained shortly. Furthermore, the data consumer has the choice to provide an additional number k, which can effectively reduce transmission expenses, alongside the trapdoor T. By using the cloud server, this method reduces the search query to the top k most relevant sites. Lastly, the users' ability to decrypt will be managed via an access control mechanism.

SYSTEM TESTING

There are several types of tests. Each test category corresponds to a certain testing need.

System Test

The integrated software system satisfies all the criteria, according to the system testing. Consistent results provide a problem due to the arrangement. System testing comes in many forms, one of which is configuration-oriented system integration testing. This tool detects pre-existing linkages and integration points by analyzing process descriptions and flows.

White Box Testing:

As a software tester, white box testing entails learning the ins and outs of the program's structure, language, and mechanics—or at least its intended use. Areas inaccessible from a black box perspective can be evaluated using it.

Black Box Testing

Assessing the operational capabilities of software The phrase "black box" pertains to the practice of testing a module without any prior knowledge of its underlying structure, programming language, or functionality. Similar to other forms of testing, black box tests also need the creation of an authoritative source document, such as a specification sheet or requirements document. In this form of testing, the software under examination is considered a "black box" that is not transparent to its internal workings. The test generates outputs and reacts to them without taking into account the functioning of the application.

CONCLUSION

Searching encrypted cloud data using several keywords is the main topic of this work. In addition to developing a range of privacy constraints, this study is also working to build privacy constraints. Moreover, a variety of privacy limits are developed as part of the study. When it comes to capturing similarity between query terms and documents that have been outsourced in an efficient manner, the notion of "coordinate matching" is utilized as an efficient method. As an additional point of interest, the concept of "inner product similarity" is utilized in order to statistically codify such a theory with the intention of evaluating similarity. Performing this action is done in order to guarantee that the comparison is correct. It is essential to first give a fundamental MRSE technique that makes use of safe inner product computation before attempting to discover a solution to the challenge of providing multi-keyword semantics without sacrificing privacy. This is because the problem is difficult to solve without compromising privacy. Therefore, in order to fulfill the requirements for privacy that are imposed by two different levels of threat models, it is necessary to make considerable modifications to this approach which is now being utilized.

References

- 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.
- Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" IEEE TRANSACTION ON PARELLEL AND DISTRIBUTED SYSTEMS, VOL.25, NO 1, JANUARY 2014.
- Secure Access of Encrypted Cloud Data Based on Top-K Multikeywords with User Side Ranking IJERTCONV3IS19214, Volume & Issue: ICESMART-2015 (Volume 3 - Issue 19)

- 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. 178-188, 2015.
- 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. 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007
- 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
- R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.