INTERNATIONAL JOURNAL · ENGINEERING RESEARCH

International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 4th Aug 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-08

DOI: 10.48047/IJIEMR/V10/I08/08

Title **PRIVACY PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA**

Volume 10, Issue 08, Pages: 46-51

Paper Authors

Mr. V. RAHAMATHULLA, Mr. U. KUMARA SWAMY

PRIVACY PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

Mr. V. RAHAMATHULLA,Assistant Professor in Dept Of MCA, SVIM, India.
Mr. U. KUMARA SWAMY, IVth Semester, Dept Of MCA, SVIM, India, email id: kumar.uddandam@gmail.com

ABSTRACT

With the fast growth of cloud computing services, an increasing number of people and businesses are opting to outsource their data and processing to the cloud. Data should be encrypted before outsourcing to protect data privacy, because performing searches over encrypted data is difficult. In this work, we present the MRSE-HC, a privacy-preserving multi-keyword ranked search strategy for encrypted data in hybrid clouds. A bisecting k-means clustering based keyword partition technique divides the document's keyword lexicon into balanced parts. Keyword partition based bit vectors are used for documents and queries that are used as the index of searches, according to the partitions. The public cloud utilises the trapdoor to determine the outcome in the candidates after the private cloud filters out the candidate documents using keyword partition based bit vectors. An improvement scheme EMRSE-HC is presented based on the MRSE-HC scheme, which includes a complete binary pruning tree to further increase search efficiency. MRSE-HC and EMRSE-HC are privacy-preserving multi-keyword ranked search schemes for hybrid clouds, according to the security analysis and performance assessment, and outperform the current system FMRS in terms of search efficiency.

***Keywords:* Hybrid Cloud, Multi-keyword Ranked Search, Privacy-preserving, Searchable Encryption.**

I INTRODUCTION

Cloud computing is a long-awaited concept of computing as a utility, in which cloud users may store their data remotely in the cloud and access high-quality apps and services on demand from a common pool of programmable computer resources [1]. Individuals and businesses alike are being enticed to outsource their local sophisticated data management system to the cloud because of its excellent flexibility and cost savings. Sensitive data, such as emails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud [2] to protect data privacy and combat unsolicited accesses in the cloud and beyond; however, this renders the traditional data utilisation service based on plaintext keyword search obsolete. Due to the enormous amount of bandwidth expense in cloud scale systems, the simple option of downloading all the data and decrypting locally is plainly unfeasible. Furthermore, aside from removing local storage management, putting data on the cloud is useless unless it can be found and used quickly. As a result, it's critical to investigate privacy-preserving and effective search services for encrypted cloud data. Given the possibility for a high number of on-demand data users and a significant number of outsourced data documents in the cloud, this challenge is particularly tough to solve, since it is exceedingly difficult to fulfil performance, system usability, and scalability criteria. On the one hand, the huge number of documents necessitates the cloud server performing result relevance rating rather than delivering undifferentiated results in order to fulfil the effective data retrieval need. Instead of sifting through every match in the content collection, data consumers may rapidly locate the most relevant information using a ranked search algorithm [3].

In the “pay-as-you-use” cloud model, ranked search may also neatly minimise needless network traffic by returning just the most relevant data, which is very desirable. Such a ranking process, however, should not disclose any keyword-related information for privacy reasons. However, in order to increase the accuracy of search results as well as the user searching experience, such a ranking system must also enable multiple keyword searches, as single term searches typically produce much too coarse results. Data consumers may choose to give a list of keywords rather than just one as a signal of their search interest to obtain the most relevant data, as suggested by today's web search engines (e.g., Google search). And each keyword in the search request has the potential to further limit the search results. In the plaintext information retrieval (IR) community, “coordinate matching” [4], i.e., as many matches as feasible, is an effective similarity measure among such multi-keyword semantics to improve the result relevance.\

II RELATTED WORK

Traditional single keyword searchable encryption techniques typically provide an encrypted searchable index whose information is concealed from the server unless proper trapdoors produced through secret key(s) are provided [2]. Song et al. investigated it initially in the symmetric key setting, while Goh, Chang et al., and Curtmola et al. [8] provide refinements and improved security definitions. Our early work addresses secure ranked keyword search, which ranks results based on term frequency rather than delivering undifferentiated results. However, it just allows you to search for a single term. Boneh et al. describe the first searchable encryption architecture in the public key setting, in which anybody with the public key may write to the data stored on the server, but only authorised users with the private key can search. However, public key solutions are typically computationally costly. Furthermore, in a public key scenario, keyword privacy may not be guaranteed since the server might encrypt any keyword with the public key and then assess the ciphertext using the received trapdoor.

These systems have significant overhead due to their core primitives, such as the cost of computation with a bilinear map, for example [16], or the cost of communication with secret sharing, for example [15]. Predicate encryption methods, which enable both conjunctive and disjunctive search, have lately been proposed as a more generic search strategy. Conjunctive keyword search returns "all-or-nothing" results, meaning it only returns documents that contain all of the keywords specified in the search query; disjunctive keyword search returns "undifferentiated" results, meaning it returns any document that contains a subset of the specific keywords, even if it only contains one keyword of interest. In brief, none of the current Boolean keyword searchable encryption systems enable multiple keywords ranked search over encrypted cloud data while maintaining privacy, which is what we propose to investigate in this article. Inner product searches in predicate encryption only determine if two vectors are orthogonal, i.e., the inner product value is hidden unless it equals zero. Predicate encryption isn't eligible for ranked search since it doesn't allow you to compare hidden inner goods. Furthermore, the majority of these methods are based on the time-consuming assessment of elliptic curve pairing procedures. When implemented in the cloud, this inefficiency disadvantage also restricts their actual performance. On a side note, database community research on top-k retrieval [27] is also somewhat related to our challenge.

III PROPOSED SYSTEM

We propose a system in which any authorised user may conduct a search on encrypted data using multiple keywords without disclosing the keywords he is looking for or the data of the documents that fit the query. Authorized users can conduct cloud searches using specific keywords to obtain relevant documents. Our proposed method makes it possible for a group of users to query the database if they have so-called trapdoors for search words that allow them to incorporate them in their searches. Our suggested system can search for many keywords in a single query and rank the results so that the user may see just the most relevant matches in a logical order. We also develop a set of stringent privacy guidelines. We choose the effective principle of "coordinate matching" from among various multi-keyword semantics.

IV SYSTEM DESIGN

The creation of a basic structural framework for a system is what system architecture is all about. It establishes the project's broad framework, which briefly outlines the structure's operation, and the goal of the project phase is to devise a solution to the problem mentioned in the requirement file. The structure's outline is seen in Figure 1 below. In our system design, we examine three components: the Data Owner, the Data User, and the Cloud Server.

- The Data Owner is in charge of the database's construction; Data Users are members of a group who have access to the database's files.
- The Cloud Server provides information services to authorised users. It is important for the server to be unaware of the contents of the database it maintains.

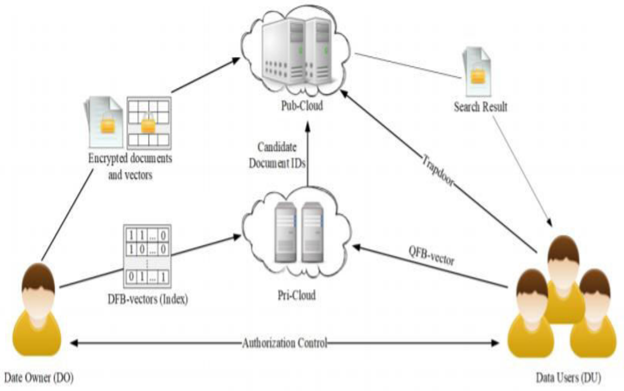

Fig 1:Architecture Diagram

The data owner has a large number of data records that he wants to store in encrypted form on a cloud server. Before outsourcing, the data owner will create a safe searchable index from a list of different keywords extracted from the file collection, and then store both the index and the encrypted file on the cloud server. We handle the approval process between the data owner and the users. A certified user makes and transmits a search request to the cloud server in a secret forma trapdoor of the keyword to search the file collection for a specified keyword. The server is responsible for searching the index and returning the matched set of files to the user after receiving the search request. We investigate the problem of secure ranked keyword search as follows: To improve file retrieval correctness for users, the search result must be returned according to certain ranking relevance criteria. Cloud servers, on the other hand, must learn little or nothing about the main concepts themselves, as they expose highly sensitive information in violation of keyword privacy. To save bandwidth, the user can provide the trapdoor with a potential value of k, and the cloud server will only give back the top-k most relevant files

V METHODOLOGY

System Model

Consider a cloud data hosting service with three separate entities: the data owner, the data user, and the cloud server, as shown in Fig. 1. The data owner has a set of data documents F that will be encrypted and sent to the cloud server. Before outsourcing, the data owner will first construct 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 provide searching capabilities over C for effective data usage. An authorised user obtains a matching trapdoor T using search control methods, such as broadcast encryption [8,] to search the document collection for t specified keywords. The cloud server is responsible for searching the index I and returning the matching collection of encrypted documents after receiving T from a data user. To increase document retrieval accuracy, the cloud server should rank the search results based on specific ranking criteria (e.g., coordinate matching, as will be introduced shortly). Furthermore, the data user may provide an optional number k along with the trapdoor T to have the cloud server only give back the top-k pages that are most relevant to the search query, reducing transmission costs.

Threat Model

In our approach, the cloud server is seen as “honest-but-curious,” which is consistent with prior cloud security research. The cloud server, in particular, behaves in a “honest” manner and accurately respects the protocol definition. However, inferring and analysing data (including index) in its storage and message flows received via the protocol in order to gain more information is "curious." We explore two threat models with distinct attack possibilities based on the information the cloud server has.

Security

When the same queries are used in the search stage, the Pub-Cloud creates various trapdoors, but the candidate documents and computed relevance scores are the same. These conversion channels may be used by Pub-Cloud to link the same search queries and derive the most popular terms found often in documents. Extending dimension by inserting certain phantom words into vectors to break such conversion channels is a feasible and successful countermeasure to address this problem. This approach is also introduced to improve the security of our scheme by protecting document confidentiality, index and trapdoor privacy, and trapdoor unlikability.

Security analysis

Confidentiality of documents The documents and vectors are encrypted in the scheme before being outsourced to Pub-Cloud. The EncData algorithm specifies the encryption processes. Documents are encrypted using the secret key g in SK using a symmetric encryption method (such as AES), while the associated vectors are encrypted using the secure inner product operation in SK using one random bit vector and two random invertible matrix. Pub-Cloud is unaware of the secret keys in SK since SK created by DO is only shared with authorised DU. As a result, Pub-Cloud is unable to extract plaintext information from encrypted documents and vectors, and document confidentiality is ensured.

Privacy, Index, and Trapdoor The index in this system is the CBP-Tree that is stored in Pri-Cloud, which is separate from PubCloud. Trapdoors are created by executing a secure inner product on the query vectors using the secret key SK, which is only shared by DO and DU. Several random elements are also added to the trapdoor creation process, increasing the unpredictability of vector values.

VI CONCLUSION

First, we explain and overcome the challenges of multi-keyword ranked search over encrypted cloud data, as well as develop a range of privacy criteria in this paper. We choose the effective similarity measure of "coordinate matching," i.e., as many matches as possible, from a variety of multi-keyword semantics to effectively convey the importance of outsourced documents to query communication. In the future, we'll look at supporting various multi-keyword semantics over encrypted data and ensuring the rank order in the search result keywords is accurate. We present a fundamental notion of MRSE to address the problem of supporting multi-keyword semantic without privacy breaches. Then, in two different threat models, we provide two improved MRSE outlines to realise numerous strict privacy criteria. A detailed analysis of the proposed schemes' privacy and efficiency guarantees is provided, and experiments on a real-world data set indicate that our future systems have little overhead on computation and transmission.

VII REFERENCES

[1] Chi Chen, Member, IEEE, Xiaojie Zhu, Student Member,IEEE, Peisong Shen, Student Member, IEEE,Jiankun Hu, Member, IEEE, Song Guo, Senior Member, IEEE, Zahir Tari, Senior Member, IEEE, and Albert Y. Zomaya, Fellow, IEEE, " An Efficient PrivacyPreserving Ranked Keyword Search Method" , IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 4, APRIL 2016

[2] Ning Cao†, Cong Wang‡, Ming Li†, Kui Ren‡, and Wenjing Lou††Department of ECE, Worcester Polytechnic Institute,

Email: {ncao, mingli, wjlou}@ece.wpi.edu,‡Department of ECE, Illinois Institute of Technology, “Privacy Preserving Multi Keyword Ranked Search Over Encrypted Cloud Data.

[3] Keiko Hashizume, David G Rosado2, Eduardo FernándezMedina and Edu ardo B Fernandez, “An analysis of security issues for cloud computing”, Hashizume et al. Journal of Internet Services and Applications 2013, 4:5

[4] Cong Wang†, Ning Cao‡, Jin Li†, Kui Ren†, and Wenjing Lou‡†Department of ECE, Illinois Institute of Technology, Chicago, IL 60616 Email: {cong, jli, kren}@ece.iit.edu ‡Department of ECE, Worcester Polytechnic Institute, Worcester, MA 01609 Email: {ncao, wjlou}@ece.wpi.edu,

[5] “Secured ranked Search over Encrypted Data” 2010 International Conference on Distributed Computing Systems

[6] W. K. Wong,The University of Hong Kong wkwong2@cs.hku.hk ,David W. Cheung The University of Hong Kong dcheung@cs.hku.hk Ben Kao The University of Hong Kong kao@cs.hku.hk Nikos Mamoulis The University of Hong Kong nikos@cs.hku.hk, “Secure kNN Computation on Encrypted Databases”

[7] Cong Wang, Student Member, IEEE, Qian Wang, Student Member, IEEE,Kui Ren, Senior Member, IEEE, Ning Cao, and Wenjing Lou, Senior Member, IEEE, “Toward Secure and Dependable Storage Services in Cloud Computing”, 220 IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 5, NO. 2, APRIL-JUNE 2012

[8] Keiko Hashizume1*, David G Rosado2, Eduardo Fernández-Medina2 and Eduardo B Fernandez1,” An analysis of security issues for cloud computing, Hashizume et al. Journal of Internet Services and Applications 2013, 4:5

[9] T. Jothi Neela1* and N. Saravanan2,” Privacy Preserving Approaches in Cloud”, Vol 6 (5) | May 2013

[10] Yong-Il Kim1, Yoo-Kang Ji2 and Sun Park3*,” Big Text Data Clustering using Class Labels and Semantic Feature Based on Hadoop of Cloud Computing”, International Journal of Software Engineering and Its Applications Vol.8, No.4 (2014)

[11] Muhammad Yasir Shabir, Asif Iqbal, Zahid Mahmood_, and AtaUllah Ghafoor, ” Analysis of Classical Encryption Techniques in Cloud Computing”, TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll09/10llpp102- 113 Volume 21, Number 1, February 2016o the user's concerned keyword.