Applied Science & Engineering Journal for Advanced Research
ISSN (Online): 2583-2468
Volume-1 Issue-2 || March 2022 || PP. 13-16 DOI: 10.54741/asejar.1.2.3

Encrypted Data and Indexing Search Keyword on Multi Cloud

Swetha.V
M.Tech Scholar, Department of Computer Science and Engineering, Veerammal College of Engineering, Dindigul, India

Corresponding Author: swetha.v2009@gmail.com

Received: 27-02-2022 Revised: 10-03-2022 Accepted: 22-03-2022

ABSTRACT
Many data owners are embracing cloud computing to outsource their complex data management systems because of its elasticity and also the cost savings. Prior to outsourcing, all sensitive data must be encrypted to ensure its privacy. The search service must be able to handle multi keyword queries and also provide resemblance ranking in order to meet the statistics repossession needs of cloud users. Privacy-preserving multi-keyword ranking search over encoded cloud data is defined and solved in this paper for the first time, and we establish a set of strict privacy requirements for such a secure cloud data utilisation system when it comes to multi-keyword semantics, "coordinate matching" is the most efficient method. We begin by proposing a basic MRSE scheme based on secure inner product computation, and experiments on the real-world dataset show that the schemes we've proposed do indeed have low computation and communication overhead.. Pseudo (fake) documents that can accurately reflect user information requirements are the focus of a new optimization method we've developed. We also propose a new criterion for evaluating the performance of the restructured web search results, called average precision (CAP). By further extending these two schemes, we can better serve our data search customers. Real-world data experiments show that the proposed schemes do indeed have low computational and communication overhead.

Keywords*: multi keyword search, privacy search, cloud computing*

I. INTRODUCTION

On-demand of the high quality presentations and their services are made available to cloud customers by using a shared pool of configurable computing resources provided by CLOUD computing, a vision that has been a long time in the making. For the sake of privacy and security, sensitive data, such as emails, photo albums ,medical accounts, tax documents, and financial transactions should need to be encrypted before being transferred to a public cloud service provider. Traditional data utilisation services that use plaintext keyword search are now rendered obsolete. This problem is particularly challenging because it is extremely difficult to meet the requirements of performance, system usability, and scalability, given the large number of on-demand data users and the enormous amount of outsourced data documents in the cloud. To indicate their search interest by providing a list of keywords rather than just one, so that the most relevant data can be retrieved. There are many ways to narrow down the results, and each keyword in the search request can help. As many matches as possible, also known as "coordinate matching," is an effective similarity measure among the multi keyword semantics to refine the relevance of the search results.

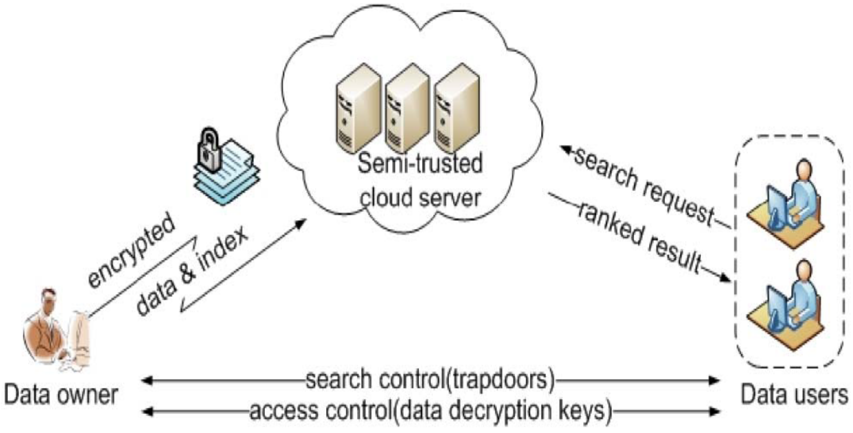

Figure 1: Encrypted cloud data search architecture

II. PROBLEM OF THE FORMULATION

2.1 Organization Model

As presented in Figure, there are three parties involved in a cloud data hosting service: the data user, the data owner, and the cloud server itself. The data owner has a collection of encrypted data documents F that they want to transfer to the cloud server C. Before outsourcing, the data owner will create an encrypted searchable index I from F and then send the index I and the encrypted document collection C to the cloud server. This will allow effective data utilisation by enabling searching over C.

Search control mechanisms, such as broadcast encryption, allow an authorised user to acquire a corresponding trapdoor T for t given keywords. The cloud server is responsible for searching the index I and returning the corresponding set of encrypted documents when it receives T from a data user. According to some ranking criteria, the cloud server would rank the examine results to improve the accuracy of document retrieval (e.g., coordinate matching, as will be introduced shortly). Additionally, the data user may send an optional number k along with the trapdoor T so that the cloud server only returns the top-k documents that are most relevant to the search query. It is possible to insert new documents, update existing documents, and delete existing ones using the access regulator instrument used to accomplish decryption competencies granted to managers and the data assortment.

2.2 Threat Model

According to the model, a cloud attendant is "honest but curious," which is in line with other research on cloud security. As a result, the cloud server acts in a "honest" manner and adheres to the specified protocol specification. The inference and analysis of data (including the index) in its storage and message flows received during the protocol in order to learn additional information is, however, "curious". The following are two threat models with varying attack capabilities based on the information that the cloud server has access to.

This cipher text model is well known. Data set C and searchable index I, both of which are hosted by a third party, are all that the cloud server is supposed to know in this model.

Unknown foreground fashion model It is assumed that the cloud server in this more secure model has access to information that is not available to it in the more traditional ciphertext model. Trapdoors (search requests) and data set statistics are examples of correlations that may be included in this type of information.

2.3 Design Goals

The system design should be instantaneously achieve the following security and enactment guarantees in order to enable categorized search for effective utilisation of outsourced cloud data under the aforementioned model. A search that ranks results based on a variety of different keywords. Instead of displaying a slew of similar results, search schemes should be developed that allow users to enter multiple keywords and rank the results based on how closely they match one another. Privacy-preserving. The data set and index have been encrypted to prevent the cloud server from learning additional information, and to meet the specified privacy requirements Efficiency Low communication and computation overhead should be used to achieve the above stated functional and privacy objectives.***2.4 Notations***

- F—the plaintext document collection, denoted as a set of m data documents
- F =(F1,F2,….Fm)
- C—the encrypted document collection stored in the cloud server, denoted as
- W—dictionary, i.e., the keyword set consisting of n keyword, denoted as
- C=(C1, C2,… upto..Cm)

2.4 Preliminary on Coordinate Matching

"Coordinate matching" is a hybrid of conjunctive search and disjunctive search that uses the number of query keywords in a document to measure the relevance of that document to the query. Boolean queries perform well when the user knows exactly what data they want to search for and what they want to find.

III. FRAMEWORK AND PRIVACY

"Coordinate matching" is a conjunctive/disjunctive search hybrid that measures the consequence of a document to a query by counting number of query keywords that document contains. When a user knows exactly what data they're looking for and how they want to find it, boolean queries work well.

3.1 MRSE Framework

The framework does not include operations on the data documents because the data titleholder could easily engagement the outmoded symmetric key cryptography to translate and then transfer facts. The MRSE system consists of four algorithms, which are outlined as follows:

- Setup(1^l). Taking a security parameter l as input, the data owner outputs a symmetric key as SK.
- BuildIndex (SK, F). Based on the data set F, the data owner builds a searchable index. After the index construction, the document collection can be independently encrypted and outsourced. I which is encrypted by the symmetric key SK and then outsourced to the cloud server.
- Trapdoor ($w^{\tilde{}}$) With t keywords of interest in W as input,this algorithm generates a corresponding trapdoor TW

3.2 MRSE Privacy Requirements

Searchable encryption is an example of a privacy guarantee where the server is guaranteed to learn nothing but the search results. The MRSE framework's strict privacy requirements are explored and established in light of this general privacy description. Traditional symmetric key cryptography can be used to encrypt outsourced data before it's sent to the cloud, keeping it safe from prying eyes on the cloud server.

Privacy for keywords. Since most users prefer not to have their search results made public, such as by the cloud server, the most pressing concern is to hide the keywords that the trapdoor indicates the user is searching for. Disconnection of trapdoors. Instead of using a deterministic trapdoor generation function, use a random one. Cloud servers should be unable to deduce the relationship between any given trapdoors, such as determining if the two trapdoors are formed by the same search request. Pattern for gaining access. The ranked search's access pattern is the sequence of search results in which each search result is a set of documents ranked according to their importance. FeW denotes the id list of all documents ranked by their relevance to the query keyword set fW in the search results for the query keyword set fW. As a next step, we examine the patterns of access that are generated by sequential searches.

IV. EFFICIENT MRSE & PRIVACY-PRESERVING

We propose to use "inner product similarity" to quantitatively evaluate the efficient similarity measure "coordinate matching" in order to achieve efficient multi-keyword ranked search. Document Fi contains the keyword Wj, and Q is a binary query vector that identifies which bits in document Fi correspond to the relevant keyword Wj.

Before showing how it can be significantly improved to be privacy-preserving against various threat models within the framework of the MRSE, we first propose an initial idea for the use of secure inner product computation, which is adapted from a secure kNN technique. Additional search semantics and dynamic operation are also discussed.

4.1 MRSE_I Scheme

As opposed to the initial plan, which was to simply remove the extended dimension from the query vector, in our more advanced design, we keep the dimension extending operation while assigning a new random number t to the extended dimension in each query vector. The cloud server is expected to have a more difficult time figuring out the relationship between the trapdoors it receives. Randomness should also be carefully calibrated in the search results to obfuscate document frequency and reduce the chances of reidentification of keywords, as stated in the keyword privacy requirement.

4.2 Analysis

We examine this MRSE I scheme in terms of the three design goals outlined in Section. The ability to do something with ease and effectiveness. Assume the number of keywords in a document. is expected to be smaller in order to achieve high precision, which indicates the good purity of the retrieved documents, from a consideration of effectiveness.

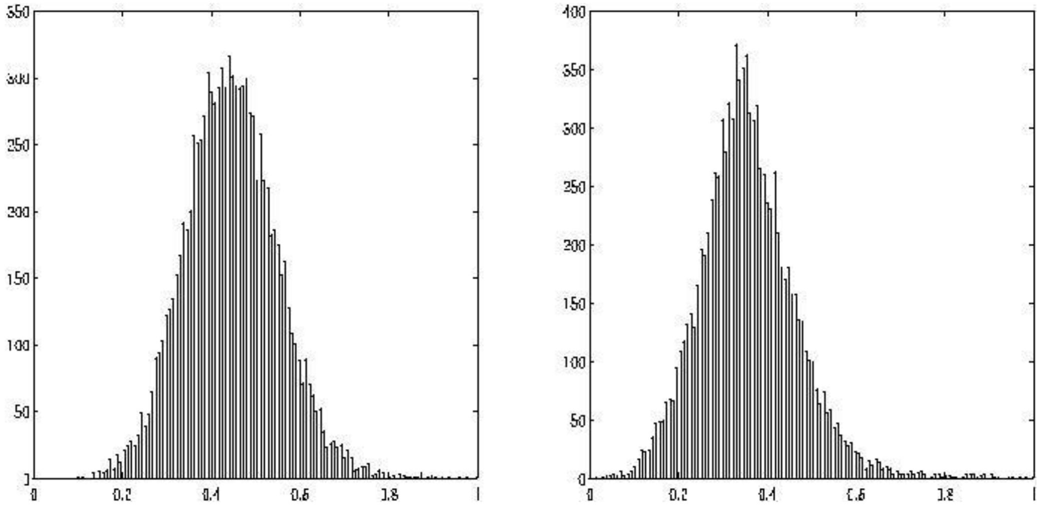

Figure 2: Distribution of final similarity score with different standard deviations, 10k documents, 10 query keywords.

4.3 MRSE

The manifestation of a keyword in the document / query is represented in the data vector or query direction according to the ranking principle of "coordinate matching.". There are, in fact, a number of other factors that could affect the usability of a search engine. When a keyword appears in the majority of the documents in a data set, its importance in the query is reduced relative to keywords that appear in a smaller number of documents.

V. PERFORMANCE ANALYSIS

The Enron Email Data Set is used in this section to demonstrate a thorough experimental evaluation of the proposed technique. E-mails are selected at random to build the data set. On a Linux Server with an Intel Xeon Processor 2.93 GHz, the entire experiment system is written in C. Numerical Recipes' public utility routines are used to compute the inverse of a matrix. The efficiency of four proposed MRSE schemes, as well as the tradeoff between search precision and privacy, is evaluated by our technique.

VI. CONCLUSION

The Multi-keyword categorized search over translated haze data is defined and solved in this paper for the first time, with a variety of privacy requirements established. For the purpose of capturing the relevance of outsourced documents to query keywords, we select the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, and use "inner product similarity" to quantitatively evaluate such a similarity measure. We propose a basic MRSE idea based on secure inner product computation to meet the challenge of supporting multi-keyword semantic without compromising privacy. Two improved MRSE schemes are then provided to meet variously strict seclusion necessities in 2 distinctive threat prototypes. We're also looking into ways to improve our categorized search appliance, such as allowing more search semantics to be supported.

REFERENCES

1. L.M. Vaquero, L. Rodero-Merino, J. Caceres, & M. Lindner. 92009). A break in the clouds: Towards a cloud definition. *ACM SIGCOMM Comput. Commun. Rev., 39*(1), 50-55.
2. A. Singhal. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull., 24*(4), 35-43.1
3. I.H. Witten, A. Moffat, & T.C. Bell. (1999). *Managing gigabytes: Compressing and indexing documents and images*. Morgan Kaufmann Publishing.
4. D. Song, D. Wagner, & A. Perrig. 92000). Practical techniques for searches on encrypted data. In: *IEEE Symp. Security and Privacy*.