TOWARDS SECURE MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

Mrs. B. Jeeva, Dr. S. Rajalakshmi,
M.E., (CSE), Jay Shriram Group of Institutions, Tirupur.
Associate Professor/CSE, Jay Shriram Group of Institutions, Tirupur.
jeeva2mail@gmail.com mrjaislm@gmail.com

Abstract- The project defines and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wise privacy in the cloud computing paradigm. Data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have 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. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. Among various multi-keyword semantics, choosing the efficient similarity measure of “coordinate matching,” i.e., as many matches as possible, to capture the relevance of data documents to the search query. Specifically, here I use the “inner product similarity” i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query. During the index construction, each document is associated with a binary vector as a sub index where each bit represents whether corresponding keyword is contained in the document. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy. The vector space model helps to provide sufficient search accuracy, and the DES encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on cipher text. As a result, information leakage can be eliminated and data security is ensured. Thorough security and performance analysis show that the proposed scheme guarantees high security and practical efficiency.

Index Terms—MRSE, OTP, Cloud, Product similarity

1. INTRODUCTION

Cloud computing is a colloquial expression used to describe a variety of different types of computing concepts that involve large number of computers that are connected through a real-time communication network (typically the Internet). Cloud computing is a jargon term without a commonly accepted non-ambiguous scientific or technical definition. In science, cloud computing is a synonym for distributed computing over a network and means the ability to run a program on many connected computers at the same time. The popularity of the term can be attributed to its use in marketing to sell hosted services in the sense of application service provisioning that run client server software on a remote location.

Cloud computing relies on sharing of resources to achieve coherence and economies of scale similar to a utility (like the electricity grid) over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services.

The cloud also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but as well as dynamically re-allocated as per demand. This can work for allocating resources to users in different time zones. For example, a cloud computer facility which serves European users during European business hours with a specific application (e.g. email) while the same resources are getting reallocated and serve North American users during North America's business hours with another application (e.g. web server). This approach should maximize the use of computing powers thus reducing...
environmental damage as well, since less power, air conditioning, Rackspace, and so on, is required for the same functions.

The term "moving to cloud" also refers to an organization moving away from a traditional CAPEX model (buy the dedicated hardware and depreciate it over a period of time) to the OPEX model (use a shared cloud infrastructure and pay as you use it).

Proponents claim that cloud computing allows companies to avoid upfront infrastructure costs, and focus on projects that differentiate their businesses instead of infrastructure. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and enables IT to more rapidly adjust resources to meet fluctuating and unpredictable business demand.

**Multi-Keyword Ranked Search over Encrypted (MRSE)**

CLOUD computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, 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; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus, exploring privacy preserving and effective search service over encrypted cloud data is of paramount importance. Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability.

On the one hand, to meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the “pay-as-you-use” cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. As a common practice indicated by today’s web search engines (e.g., Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. “Coordinate matching”, as many matches as possible, is an efficient similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information retrieval (IR) community. However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, and many others.

Encryption is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality. Our early works have been aware of this problem, and provide solutions to the secure ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem.

In the project, for the first time, define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wide privacy in the cloud computing paradigm. Among various multi-keyword semantics, choose the efficient similarity measure of “coordinate matching,” i.e., as many matches as possible, to capture the relevance of data documents to the search query. Specifically, inner product similarity the measure of “coordinate matching,” as many matches as possible, to capture the relevance of data documents to the search query. During the index construction, each document is associated with a binary vector as a sub-index where each bit represents whether corresponding keyword is contained in the document. The search query is also described as a binary vector
where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy. To meet the challenge of supporting such multi keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique, and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various stringent privacy requirements in two threat models with increased attack capabilities. Our contributions are summarized as follows:

1. For the first time, we explore the problem of multi keyword ranked search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system.

2. We propose two MRSE schemes based on the similarity measure of “coordinate matching” while meeting different privacy requirements in two different threat models.

3. We investigate some further enhancements of our ranked search mechanism to support more search semantics and dynamic data operations.

4. Thorough analysis investigating privacy and efficiency guarantees of the proposed schemes is given, and experiments on the real-world data set further show the proposed schemes indeed introduce low overhead on computation and communication.

Compared with the preliminary version of this paper, this journal version proposes two new mechanisms to support more search semantics. This version also studies the support of data/index dynamics in the mechanism design. Moreover, we improve the experimental works by adding the analysis and evaluation of two new schemes. In addition to these improvements, we add more analysis on secure inner product and the privacy part.

II.OBJECTIVE

User Ranking Guarantee why something is mentioned a lot, and that it isn't due to marketing, or self-promotion, rather than importance. Proposed cloud storage systems that provide confidentiality, integrity and verifiability of client data against a UN trusted cloud provider. This OTP used to see data in cloud and it can be used once only in a time, when you search a file and tend to see the file the OTP will send to email and you get the OTP and apply to see the file.

III. EXISTING SYSTEM

The cloud server hosts third-party data storage and retrieve services. Since data may contain sensitive information, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Any kind of information leakage that would affect data privacy is regarded as unacceptable. To meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the “pay-as-you-use” cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results.

IV.PROPOSED SYSTEM

In the Proposed work, we will explore checking the integrity of the rank order in the search result assuming the cloud server is untrusted.

To propose OTP (one Time Password) as our future work. This OTP used to see data in cloud and it can be used once only in a time, when you search a file and tend to see the file the OTP will send to email and you get the OTP and apply to see the file.

System Architecture

Fig.2 Architecture diagram of the MRSE Implementation.

Data Flow Diagram
**MODULE DESCRIPTION**

The following modules are implemented in this technique

a. Cloud Setup

b. Cryptography cloud Storage

c. Vector Model

**Cloud Setup**

In this module we have setup data owner and cloud server. So the data owner is going push the data into the cloud sever. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud will be secured.

**Cryptography cloud Storage**

In this module while the data is uploaded into the Estorage and retrieve services. Since data may contain sensitive information, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Any kind of information leakage that would affect data privacy are regarded as unacceptable.

**Vector Model**

In this model we used a series of searchable symmetric encryption schemes have been enable search on cipher text. In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy.
E. User login

F. Search Keyword

VII. CONCLUSION AND FUTURE WORK
The first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of “coordinate matching,” i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use “inner product similarity” to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we propose a basic idea of MRSE using secure inner product computation. Then, we give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. We also investigate.

Some further enhancements of our ranked search mechanism, including supporting more search semantics, i.e., TF _ IDF, and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduce low overhead on both computation and communication.

VIII. REFERENCES


AUTHORS BIOGRAPHY

B.Jeeva received her B.E degree in Idhaya engineering college for women chinnasalem, India and currently pursuing M.E degree in Jay Shriram Group of Institutions, Tiruppur, India. Her research interests include Data mining, Cloud Computing and Big Data.

Dr.S. RajaLakshmi received her B.E. degree in Periyar University, Salem, India and M.E. degree in Anna University, Chennai, India and Ph.D. in Data mining in Anna University, India. Currently she is working as an Associate Professor in Jay Shriram Group of Institutions, Tirupur, India. Her research interests include Data mining, Cloud Computing and Big Data.