IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# PRIVACY-PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

BADHE KALPESH NILKANTH.
STUDENT OF M.TECH LAST YEAR
DR. DINESH D. PATIL
HEAD OF DEPT. & ASSOCIATE PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SHRI SANT GADGE BABA
COLLEGE OF ENGINEERING AND TECHNOLOGY, BHUSAWAL-425203
DR.BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY, LONERE, INDIA
2020-21

*Abstract:* The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, 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.

Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

Index Terms - cloud computing, multi-keyword, searchable encryption.

## I.INTRODUCTION

We are living in a highly networked environment, where huge amounts of data are stored in remote, but not necessarily trusted servers. There are several privacy issues regarding to accessing data on such servers; two of them can easily be identified: sensitivity of i) keywords sent in queries and ii) the data retrieved; both need to be hidden. A related protocol, Private Information Retrieval (PIR) enables the user to access public or private databases without revealing which data he is extracting. Since privacy is of a great concern, PIR protocols have been extensively studied in the past.

In today's information technology landscape, customers that need high storage and computation power tend to 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 robust and fast services. The importance and necessity of privacy preserving search techniques are even more pronounced in the cloud applications. Due to the fact that 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.

To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, our system design should simultaneously achieve security and performance guarantees as follows.

- Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.
- Privacy-Preserving: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements specified in section III-B.
- Efficiency: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

## II. LITERATURE REVIEW

### 1. HISTORY

The problem of Private Information Retrieval was first introduced by Chor et al. Recently Growth et al propose a multi query PIR method with constant communication rate. However, any PIR-based technique requires highly costly cryptographic operations in order to hide the access pattern. This is inefficient 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. The scheme requires a public key operation per item in the database for every query and this operation must be performed on the user side.

Freedmanetal, proposed an alternative implementation for private keyword search that uses homomorphic encryption and oblivious polynomial evaluation methods. The computation and communication costs of this method are quite large since every search term in a query requires several homomorphic encryption operations both on the server and the user side.

A recent work proposed by Wang et al allows ranked search over an encrypted database by using inner product similarity. However, this work is only limited to single keyword search queries.

One of the closest methods to our solution is proposed by Cao et al. Similar to our approach presented here, 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. Additionally, 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 to generate a query. This assumption may not be applicable in several cases. Moreover, it is not efficient due to matrix multiplication operations of square matrices where the number of rows is in the order of several thousands.

Wangetal propose 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 a totally different encryption methodology to increase the security and efficiency of the scheme.

### 2. RELATED WORK

We defined and solve the challenging problem of Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data in cloud computing(MRSE).We establish a set of strict privacy requirement for such a secure cloud data utilization system.

Among various multi-keywords sementics. We choose the efficient similarly measure of “coordinate matching” i.e. as many matches as possible, to capture the relevance of data documents to the search query. We further use “inner product similarly” to quantitatively evaluate such similarly measure.

Firstly, the user side of proposed system will be implemented on mobile devices running Android and iOS operating systems since the potential application scenario envisions that users access the data anywhere and anytime.

And secondly, the proposed method will be tested on a real dataset in order to compare the performance of our ranking method with the ranking methods used in plain datasets that do not involve any security or privacy-preserving techniques.

As a special case of modification the operation of deleting existing documents introduce less computation and communication cost since it only requires to update the document frequency of all the keywords contained by these document.

Today, there are large number of data users and documents in cloud. It is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need.

The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results.

## 3. FRAMEWORK

Preliminary investigation examine project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited recourses and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

- ➢ Technical Feasibility
- ➢ Operational Feasibility
- ➢ Economical Feasibility

Technical Feasibility:-

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipment's have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
- Can the system be upgraded if developed?
- Are there technical guarantees of accuracy, reliability, ease of access and data security?

The current system developed is technically feasible. It is a many-to-many user interface for interaction. The purpose of the web application is to mainly facilitate communication between the client and the Server. As the Server Application is exclusively run only on the Server, it provides Security and no chances of misuse. The software and hard requirements for the development of this project are not many and are already available in any organizations. The work for the project is done with the current equipment and existing software technology. Necessary bandwidth exists for providing a fast feedback to the users irrespective of the number of users using the system.

Operational Feasibility:-

Proposed projects are beneficial only if they can be turned out into information system. They will meet the organization's operating requirements. Operational feasibility aspects of the project are to be taken as an important part of the project implementation. Some of the important issues raised are to test the operational feasibility of a project includes the following:-

- Is there sufficient support for the management from the users?
- Will the system be used and work properly if it is being developed and implemented?
- Will there be any resistance from the user that will undermine the possible application benefits?

This system is targeted to be in accordance with the above-mentioned issues. Beforehand, the management issues and user requirements have been taken into consideration. So there is no question of resistance from the users that can undermine the possible application benefits.

The well-planned design would ensure the optimal utilization of the computer resources and would help in the improvement of performance status.

Economical Feasibility:-

A system can be developed technically and that will be used if installed must still be a good investment for the organization. In the economical feasibility, the development cost in creating the system is evaluated against the ultimate benefit derived from the new systems. Financial benefits must equal or exceed the costs.

The system is economically feasible. It does not require any additional hardware or software. The tolls of the web application are already available in the Organization and it is economically feasible to buy a domain name (if necessary) in order to implement the project. Since the interface for this system is developed using the existing resources and technologies available at any organization. There is nominal expenditure and economical feasibility for certain. Also as the System is designed with a purpose to curtail the time of searching, it adds up to the Economic value of the Project.

Compared to the original SSE, the new scheme embeds the encrypted relevance scores in the searchable index in addition to file ID. Thus the encrypted scores are the only additional information that the adversary can utilize against the security guarantee, i.e., keyword privacy and file confidentiality. Due to the security strength of the file encryption scheme, the file content is clearly well protected.

Thus, we only need to focus on keyword privacy. From previous discussion, we know that as long as data owner properly chooses the range size R sufficiently large, the encrypted scores in the searchable index will only be a sequence of order-preserved numeric values with very low duplicates. Though adversary may learn partial information from the duplicates (e.g., cipher text duplicates may indicate very high corresponding plaintext duplicates), the fully randomized score-to-bucket assignment (inherited from OPSE) and the highly flattened one-to-many mapping still makes it difficult for the adversary to predict the original plaintext score distribution, let alone reverse engineer the keywords. Also note that we use different order-preserving encryption keys for different posting lists, which further reduces the information.

## 4. SCOPE

As our future work, we will explore supporting other multi keyword semantics (e.g. Weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in more stronger threat model.

Following the current research, there are possible improvements and undergoing efforts that will appear in the future work.

Firstly, the user side of proposed system will be implemented on mobile devices running Android and iOS operating systems since the potential application scenario envisions that users access the data anywhere and anytime.

And secondly, the proposed method will be tested on a real dataset in order to compare the performance of our ranking method with the ranking methods used in plain datasets that do not involve any security or privacy-preserving techniques**.**

## 5. CONCLUSION

In this paper, for 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 principle of "coordinate matching", i.e., as many matches 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.

## REFERENCES

[1] 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.

[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.

[3] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, 2001.

[4] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

[5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.

[6] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http:// eprint.iacr.org/2003/216.

[7] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.

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

[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.

[10] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007

## ACKNOWLEDGMENTS

It gives us great pleasure in presenting the project report on **'Privacy-Preserving Multi-Keyword Ranked Search Over Encrypted Cloud Data'.**

We would like to take this opportunity to thank our internal guide and Head of Computer Engineering Department, Shri Sant Gadge Baba College Of Engineering AndTechnology, Bhusawal Prof. D. D. Patil for giving us all the help and guidance weneeded. We are really grateful for his kind support.