IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 5, May 2021

DOI: 10.17148/IARJSET.2021.8561

Conservation Of Privacy In Searchable Symmetric Encryption Cloud Data Using Ranked Search

Kalpesh Nilkanth Badhe[1], Dr. Dinesh Patil[2]

*1 M.Tech (Student), Computer Science & Engineering, Shri Sant Gadge Baba COET, Bhusawal, Maharashtra, India.

*2 Head of Dept. & Associate Professor Computer Science & Engineering, Shri Sant Gadge Baba COET, Bhusawal, Maharashtra, India.

Abstract: In this paper, for the first time we define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing.
We first give a straight forward yet ideal construction of ranked keyword search under the state-of-the-art searchable symmetric encryption (SSE) security definition, and demonstrate its inefficiency. To achieve more practical performance, we then propose a definition for ranked searchable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE). Thorough analysis shows that our proposed solution enjoys "as-strong-as possible" security guarantee compared to previous SSE schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

Keywords: Ranked Search, Encrypted Cloud, Privacy-Preserving Data, Order-Preserving Symmetric Encryption, Cryptographic Primitive Accesses.

I. INTRODUCTION

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.
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. The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc. As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as emails, personal health records, company finance data, and government documents, etc. The fact that data owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted data at risk. The cloud server may leak data information to unauthorized entities or even be hacked. It follows that sensitive data has to be Encrypted prior to outsourcing for data privacy and combating unsolicited accesses.

II. PROBLEM STATEMENT

Considering a cloud data hosting service involving three different entities of the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build 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 search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms, e.g., broadcast encryption.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 5, May 2021

DOI: 10.17148/IARJSET.2021.8561

Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. To improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria.
Moreover, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. Finally, the access control mechanism is employed to manage decryption capabilities given to users.
In our model, there is a set of users, a server, and a set of documents. The server stores encrypted documents. Each user has access to a subset of the documents. A user can create a document and then give access to other users to the document by giving them the decryption key of the document.

III. 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. 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.

Disadvantage:

a) It still not adequate to provide users with acceptable result ranking functionality

b) It cannot accommodate such high service-level requirement like system usability, user

c) Searching experience and easy information discovery.

IV. PROBLEM STATEMENT

Considering a cloud data hosting service involving three different entities of the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build 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 search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms, e.g., broadcast encryption. Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. To improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria. Moreover, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. Finally, the access control mechanism is employed to manage decryption capabilities given to users. In our model, there is a set of users, a server, and a set of documents. The server stores encrypted documents. Each user has access to a subset of the documents. A user can create a document and then give access to other users to the document by giving them the decryption key of the document. At a high level, the following security guarantees are desirable. If some user was not given access to a document, the user should not be able to read the contents of that document or search over that document, even if the user colludes with the server. The setting is entirely distributed. Each user generates his key and there is no trusted party for choosing keys, and no globally trusted user. Moreover, there is no trusted party to create document keys or to help with providing access to documents.

V. PROPOSED SYSTEM

We first purpose a basic idea for the MRSE based on secure inner product computation and then give two significantly improved MRSE scheme to achieve various stringent privacy requirements in two different threat models. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching". To improve search experience of data search services, we further extend these two schemes to support more search sementics. Through analysis investigating privacy and efficiency guarantees of proposed scheme indeed introduce low overhead on computation and communication. The proposed scheme introduces nearly constant overhead while increasing the

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 5, May 2021

DOI: 10.17148/IARJSET.2021.8561

number of query keywords. Therefore our scheme cannot compromise by timing-based side channel attacks that try to differentiate certain queries based on their query time. 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

Advantages:-

a) It proposed schemes indeed introduced low overhead on computation and communication.
b) To reduce the communication cost.
c) It uses rank search mechanism to support more search semantic and dynamic data
d) Operation.
e) It is more secure and efficient.

VI. PRIVACY REQUIREMENT

The representative privacy guarantee in the related literature, such as searchable encryption, is that the server should learn nothing but search results. With this general privacy description, we explore and establish a set of strict privacy requirements specifically for the MRSE framework. As for the data privacy, the data owner can resort to the traditional symmetric key cryptography to encrypt the data before outsourcing, and successfully prevent the cloud server from prying into the outsourced data. With respect to the index privacy, if the cloud server deduces any association between keywords and encrypted documents from index, it may learn the major subject of a document, even the content of a short document. Therefore, the searchable index should be constructed to prevent the cloud server from performing such kind of association attack. While data and index privacy guarantees are demanded by default in the related literature, various search privacy requirements involved in the query procedure are more complex and difficult to tackle as follows. Keyword Privacy As users usually prefer to keep their search from being exposed to others like the cloud server, the most important concern is to hide what they are searching, i.e., the keywords indicated by the corresponding trapdoor. Although the trapdoor can be generated in a cryptographic way to protect the query keywords, the cloud server could do some statistical analysis over the search result to make an estimate. As a kind of statistical information, document frequency (i.e., the number of documents containing the keyword) is sufficient to identify the keyword with high probability .When the cloud server knows some background information of the dataset, this keyword specific information may be utilized to reverse-engineer the keyword. Trapdoor Unlink ability the trapdoor generation function should be a randomized one instead of being deterministic. In particular, the cloud server should not be able to deduce the relationship of any given trapdoors, e.g., to determine whether the two trapdoors are formed by the same search request. Otherwise, the deterministic trapdoor generation would give the cloud server advantage to accumulate frequencies of different search requests regarding different keyword(s), which may further violate the aforementioned keyword privacy requirement. So the fundamental protection for trapdoor unlink ability is to introduce sufficient non determinacy into the trapdoor generation procedure. Access Pattern Within the ranked search, the access pattern is the sequence of search results where every search result is a set of documents with rank order. Specifically, the search result for 10 the query keyword set fW is denoted as FfW, consisting of the id list of all documents ranked by their relevance to fW. Then the access pattern is denoted as (FfW1, FfW2, . . .) which are the results of sequential searches. Although a few searchable encryption works, has been proposed to utilize private information retrieval (PIR) technique , to hide the access pattern, our proposed schemes are not designed to protect the access pattern for the efficiency concerns. This is because any PIR based technique must “touch” the whole dataset outsourced on the server which is inefficient in the large scale cloud system.

VII. 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. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we first propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve privacy requirements in two levels of threat models.

REFERENCES

1) Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data” IEEE TRANSACTION ON PARELLEL AND DISTRIBUTED SYSTEMS,VOL.25,NO 1,JANUARY2014.
2) S. Kamara and K. Lauter, “Cryptographic cloud storage,” in RLCPS, January 2010, LNCS. Springer, Heidelberg.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 5, May 2021

DOI: 10.17148/IARJSET.2021.8561

3) 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.
4) M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007
5) R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.
6) Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.
7) D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.
8) E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http://eprint.iacr.org/2003/216.
9) A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, 2001.
10) H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.