ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Secured Multi-Keyword Search Over Encrypted Data in Cloud

Moturi Veena Pani
PG Scholar,
Department of CSE,
EVM College of Engineering & Technology,
Narasaraopet, AP, India.

Yalavarthi Leela Krishna
Assistant Professor,
Department of CSE,
EVM College of Engineering & Technology,
Narasaraopet, AP, India.

Abstract:

with the advent of cloud computing, 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. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE).

We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we 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. Wefurther use "inner product similarity" to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

1 INTRODUCTION:

CLOUD computing is the long dreamed vision ofcomputing as a utility, where cloud customers canremotely store their data into the cloud so as to enjoy theon-demand high-quality applications and services from ashared pool of configurable computing resources [2], [3]. Itsgreat flexibility and economic savings are motivating bothindividuals and enterprises to outsource their local complexdata management system into the cloud. To protectdata privacy and combat unsolicited accesses in the cloudand beyond, sensitive data, for example, e-mails, personalhealth records, photo albums, tax documents, financialtransactions, and so on, may have to be encrypted by dataowners before outsourcing to the commercial public cloud[4]; this, however, obsoletes the traditional data utilizationservice based on plaintext keyword search. The trivialsolution of downloading all the data and decrypting locallyis clearly impractical, due to the huge amount ofbandwidth cost in cloud scale systems. Moreover, asidefrom eliminating the local storage management, storingdata into the cloud serves no purpose unless they can beeasily searched and utilized.

Thus, exploring privacypreservingand effective search service over encryptedcloud data is of paramount importance. Considering thepotentially large number of on-demand data users andhuge amount of outsourced data documents in the cloud,this problem is particularly challenging as it is extremelydifficult to meet also the requirements of performance, system usability, and scalability. On the one hand, to meet the effective data retrievalneed, the large amount of documents demand the cloudserver to perform result relevance ranking, instead ofreturning undifferentiated results. Such ranked searchsystem enables data users to find the most relevant information quickly, rather than burdensomely sortingthrough every match in the content collection [5]. Rankedsearch can also elegantly eliminate unnecessary networktraffic by

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

sending back only the most relevant data, whichis 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 aswell as to enhance the user searching experience, it is alsonecessary for such ranking system to support multiplekeywords search, as single keyword search often yields far too coarse results.

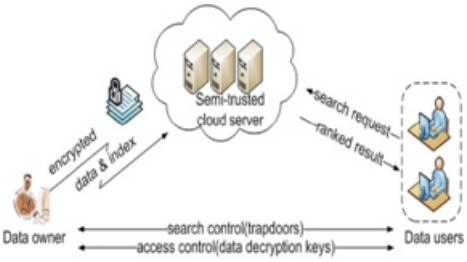

Fig. 1. Architecture of the search over encrypted cloud data

As a common practice indicated bytoday's web search engines (e.g., Google search), data usersIn this paper, for the first time, we define and solve theproblem of multi-keyword ranked search over encryptedcloud data (MRSE) while preserving strict systemwiseprivacy in the cloud computing paradigm. Among variousmulti-keyword semantics, we choose the efficient similaritymeasure of "coordinate matching," i.e., as many matches aspossible, to capture the relevance of data documents to thesearch query. Specifically, we use "inner product similarity"[6], i.e., the number of query keywords appearing in adocument, to quantitatively evaluate such similarity measureof that document to the search query. During the indexconstruction, each document is associated with a binaryvector as a subindex where each bit represents whethercorresponding keyword is contained in the document. Thesearch query is also described as a binary vector where eachbit means whether corresponding keyword appears in thissearch request, so the similarity could be exactly measuredby the inner product of the query vector with the datavector. However, directly outsourcing the data vector or thequery vector will violate the index privacy or the searchprivacy. To meet the challenge of supporting such multikeywordsemantic without privacy breaches, we proposea basic idea for the MRSE using secure inner productcomputation, which is adapted from a secure k-nearestneighbor (kNN) technique [27], and then give two significantlyimproved MRSE schemes in a step-by-stepmanner to achieve various stringent privacy requirementsin two threat models with increased attack capabilities.

Ourcontributions are summarized as follows:

1. For the first time, we explore the problem of multikeywordranked search over encrypted cloud data,and establish a set of strict privacy requirements forsuch a secure cloud data utilization system.
2. We propose two MRSE schemes based on thesimilarity measure of "coordinate matching" whilemeeting different privacy requirements in twodifferent threat models.
3. We investigate some further enhancements of ourranked search mechanism to support more searchsemantics and dynamic data operations.
4. Thorough analysis investigating privacy and efficiencyguarantees of the proposed schemes is given,and experiments on the real-world data set furthershow the proposed schemes indeed introduce lowoverhead on computation and communication.Compared with the preliminary version [1] of this paper,this journal version proposes two new mechanisms tosupport more search semantics. This version also studiesthe support of data/index dynamics in the mechanismdesign. Moreover, we improve the experimental works byadding the analysis and evaluation of two new schemes. Inaddition to these improvements, we add more analysis onsecure inner product and the privacy part.The remainder of this paper is organized as follows: InSection 2, we introduce the system model, the threat model,our design goals, and the preliminary. Section 3 describesthe MRSE framework and privacy requirements, followed bySection 4, which describes the proposed schemes. Section 5presents simulation results. We discuss related work onboth single and Boolean keyword searchable encryption inSection 6, and conclude the paper in Section 7.

2 PROBLEM FORMULATIONS

2.1 System Model:

Considering a cloud data hosting service involving threedifferent entities, as illustrated in Fig. 1: the data owner, thedata user, and the cloud server. The data owner has acollection of data documents F to be outsourced to thecloud server in the encrypted form C. To enable thesearching capability over C for effective data utilization,the data owner, before outsourcing, will first build anencrypted searchable index I from F, and then outsourceboth the index I and the encrypted document collection Cto the cloud server. To search the document collection for tgiven keywords, an authorized user acquires a correspondingtrapdoor T through search control mechanisms, for example, broadcast encryption [10].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Upon receiving T froma data user, the cloud server is responsible to search theindex I and return the corresponding set of encrypteddocuments. To improve the document retrieval accuracy,the search result should be ranked by the cloud serveraccording to some ranking criteria (e.g., coordinate matching,as will be introduced shortly). Moreover, to reduce thecommunication cost, the data user may send an optionalnumber k along with the trapdoor T so that the cloud serveronly sends back top-k documents that are most relevant tothe search query. Finally, the access control mechanism [28]is employed to manage decryption capabilities given tousers and the data collection can be updated in terms ofinserting new documents, updating existing documents,and deleting existing documents.

2.2 Threat Model:

The cloud server is considered as "honest-but-curious" inour model, which is consistent with related works on cloudsecurity [28], [29]. Specifically, the cloud server acts in an"honest" fashion and correctly follows the designatedprotocol specification. However, it is "curious" to inferand analyze data (including index) in its storage anmessage flows received during the protocol so as to learnadditional information. Based on what information thecloud server knows, we consider two threat models withdifferent attack capabilities as follows.Known ciphertext model. In this model, the cloud server issupposed to only know encrypted data set C and searchableindex I, both of which are outsourced from the data owner.Known background model. In this stronger model, the cloudserver is supposed to possess more knowledge than what canbe accessed in the known ciphertext model. Such informationmay include the correlation relationship of given searchrequests (trapdoors), as well as the data set related statisticalinformation. As an instance of possible attacks in this case,the cloud server could use the known trapdoor informationcombined with document/keyword frequency [30] todeduce/identify certain keywords in the query.

2.3 Design Goals:

To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, oursystem design should simultaneously achieve security andperformance guarantees as follows.. Multi-keyword ranked search. To design search schemes which allow multi-keyword query andprovide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.

Privacy-preserving. To prevent the cloud server fromlearning additional information from the data setand the index, and to meet privacy requirementsspecified in Section 3.2. Efficiency. Above goals on functionality and privacy should be achieved with low communication andcomputation overhead.

2.4 Preliminary on Coordinate Matching:

As a hybrid of conjunctive search and disjunctive search,"coordinate matching" [6] is an intermediate similaritymeasure which uses the number of query keywordsappearing in the document to quantify the relevance ofthat document to the query. When users know the exactsubset of the data set to be retrieved, Boolean queriesperform well with the precise search requirement specifiedby the user. In cloud computing, however, this is not thepractical case, given the huge amount of outsourced data.Therefore, it is more flexible for users to specify a list ofkeywords indicating their interest and retrieve the mostrelevant documents with a rank order.

3.PRIVACY-PRESERVING AND EFFICIENT MRSE:

To efficiently achieve multi-keyword ranked search, wepropose to employ "inner product similarity" [6] toquantitatively evaluate the efficient similarity measure"coordinate matching." Specifically, Di is a binary datavector for document Fi where each bit Di½j_ 2 f0; 1grepresents the existence of the corresponding keywordWj in that document, and Q is a binary query vectorindicating the keywords of interest where each bit Q½j_ 2f0; 1g represents the existence of the corresponding keywordWj in the query fW. The similarity score of documentFi to query fW is therefore expressed as the inner productof their binary column vectors, i.e., Di _ Q. For the purposeof ranking, the cloud server must be given the capability tocompare the similarity of different documents to thequery. But, to preserve strict systemwise privacy, datavector Di, query vector Q and their inner product Di _ Qshould not be exposed to the cloud server. In this section,we first propose a basic idea for the MRSE using secureinner product computation, which is adapted from asecure kNN technique, and then show how to significantlyimprove it to be privacy-preserving against different threatmodels in the MRSE framework in a step-by-step manner.We further discuss supporting more search semantics anddynamic operation.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

4 CONCLUSION:

In this paper, for the first time we define and solve theproblem of multi-keyword ranked search over encrypted-cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose theefficient similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture therelevance of outsourced documents to the query keywords,and use "inner product similarity" to quantitativelyevaluate such similarity measure. For meeting the challengeof supporting multi-keyword semantic without privacybreaches, we propose a basic idea of MRSE using secureinner product computation.

Then, we give two improvedMRSE schemes to achieve various stringent privacy requirementsin two different threat models. We also investigatesome further enhancements of our ranked search mechanism,including supporting more search semantics, i.e., TF_IDF, and dynamic data operations. Thorough analysisinvestigating privacy and efficiency guarantees of proposedschemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead onboth computation and communication.In our future work, we will explore checking theintegrity of the rank order in the search result assumingthe cloud server is untrusted.

REFERENCES:

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-PreservingMulti-Keyword Ranked Search over Encrypted Cloud Data," Proc.IEEE INFOCOM, pp. 829-837, Apr, 2011.

[2] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "ABreak in the Clouds: Towards a Cloud Definition," ACMSIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.

[3] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-BasedSecure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM,pp. 693-701, 2012.

[4] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc.14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[5] A. Singhal, "Modern Information Retrieval: A Brief Overview,"IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

[6] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes:Compressing and Indexing Documents and Images. Morgan KaufmannPublishing, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical Techniques forSearches on Encrypted Data," Proc. IEEE Symp. Security andPrivacy, 2000.