www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 6 June 2015, Page No. 12676-12682

Privacy Preserving Multi-Keyword Ranked Search Over Encrypted Cloud Data

Remya P.S Puliparambil, Selvi U
Department of Computer Science And Engineering
Professional Trust's Group Of Institutions
Palladam,Tamilnadu,India
remyaps74@gmail.com
Assistant Professor,Department of Computer Science And Engineering
Professional Trust's Group Of Institutions
Palladam,Tamilnadu,India
slvunnikrishnan@gmail.com

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. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements .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

Keywords : Cloud computing,Multi-keyword,Privacy preserving, inner product similarity, coordinate matching,MRSE

1.INTRODUCTION

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 hard-ware, software, and personnel maintenances, etc.,

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized

into the cloud, such as e-mails, 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 have to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses.

However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through

keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following advance of both crypto and IR community to design the ranked searchable symmetric encryption (RSSE) scheme, in the spirit of "as-strong-as-possible" security guarantee. Specifically, explore the statistical measure approach from IR and text mining to embed weight information (i.e., relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection. As directly outsourcing relevance two main drawbacks. On the one hand, for each search request, users without preknowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead; On the other hand, invariably sending back all files solely based on presence/ absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. In short, lacking of effective mechanisms to ensure the file retrieval accuracy is a significant drawback of existing searchable encryption schemes in the context of Cloud Computing. Nonetheless, the state of the art in information retrieval (IR) community has already been utilizing various scoring mechanisms to quantify and rank order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be addressed in the context of encrypted data search.

Therefore, how to enable a searchable encryption system with support of secure ranked search is the problem tackled in this paper. 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 toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve design goals on both system security and usability, we propose to bring together the

scores will leak lots of sensitive frequency information against the keyword privacy, then integrate a recent crypto primitive order-preserving symmetric encryption (OPSE) and properly modify it to develop a one-to-many order-preserving mapping technique for our purpose to protect those sensitive weight information

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. Among various multi-keyword semantics, choose the efficient principle of "coordinate matching", its many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. 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.

2.RELATED WORKS

It is an important research problem to enable the cloud service provider to efficiently search for the keyword in encrypted files and provide user with efficient search result maintaining data privacy at the same time. We have researched on the following papers.

2.1 Practical Technique for Search over Encrypted Cloud Data

This paper discusses on sequential scanning search technique [1] that searches over encrypted data stored in cloud without losing data confidentiality. The technique is provably secure and isolates the query result whereby the server doesn't know anything other the search result. It also supports functionalities such as controlled searching by server, hidden query support for user which searches for a word without revealing it to the server. With searchable symmetric encryption [7] and pseudorandom sequence generating mechanisms that are secure, encrypted data can be effectively scanned and searched without losing data privacy. The scheme that is proposed is flexible that it can be further extended to support search queries that are combined with Boolean operators, proximity queries, queries that contain regular expression, checking for keyword presence and so on. But, in case of large documents and scenarios that demand huge volumes of storage, the technique has high time complexity.

2.2 Public Key Encryption with Keyword Search

Dan Boneh proposed a solution for searching over the cloud data that is encrypted using the Public key Crypto System [2]. The idea is to securely attach or tag the related keywords along with the each file. This will avoid the need to completely decrypt the file and save the time of scanning entire file to check if the keyword exists. The file is encrypted using a public key encryption algorithm [2] and containing keyword W, send only the Trapdoor (W) to server. He proposed two methods for construction of this scheme, one using the bilinear maps and other using Jacobi symbols. The problem with this scheme is that every tag of all the files has to be processed for finding the match.

2.3 Boolean Symmetric Searchable Encryption

Most of the techniques discussed so far focused only on single keyword matching but in real-time scenarios users may enter more than one word. Tarik Moataz came up with a solution to tackle such challenges of searching multiple keywords over the encrypted cloud data. The construction of Boolean Symmetric Searchable Encryption (BSSE) [11] is mainly based on the orthogonalization of the keyword field according to the Gram-Schmidt process. The basic Boolean operations are: the disjunction, the conjunction and the negation

2.4 Fuzzy Keyword Search

The traditional searching techniques retrieve files based on exact keyword match only but Fuzzy keyword search technique extends this feature by supporting common typos and format inconsistencies that occurs when the user types the keywords. The data privacy that is maintained during exact keyword search is ensured when this method is used. Wild card based technique [4] is

used to create efficient fuzzy keyword sets that are used for matching relevant documents. The keyword sets are created using Edit Distance algorithm that quantifies word similarity. These keyword sets reduce storage and representation overhead by eliminating the need to generate all fuzzy keywords, rather generating on similarity basis. The search result that is provided is based on a fuzzy keyword data set that is generated whenever the exact match search fails.

3.PROPOSED SYSTEM

We define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in cloud computing paradigm.For system, choose the principle of coordinate matching, to identify the similarity between search query and data documents.Specially,we use inner data correspondence, i.e., the number of query keywords appearing in a document, to evaluate the similarity of that document to the search query in coordinate matching principle. Each document is linked 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 inner product of query vector with data vector. However, directly outsourcing data vector or query vector will violate index privacy or search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, propose a basic SMS scheme using secure inner product computation, which is adapted from a secure k-nearest neighbour (kNN) technique, and then improve it step by step to achieve various privacy requirements in two levels of threat models.

1) Showing the problem of Secured Multi-keyword search over encrypted cloud data

2) Propose two schemes following the principle of coordinate matching and inner product similarity.

Advantages of using the system areMulti key word ranking for secure the cloud data,Searching on the encrypted data will give an expected data, proposed schemes indeed introduce low overhead on computation on communication and uses ranked search mechanism to support more search semantics and dynamic data operations.

3.1 System Architecture

Considering a cloud data hosting service involving three different entities, as illustrated in Fig. 1: 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 correspond- ing trapdoor T through search control mechanisms, for example, 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 (e.g., coordinate match- ing, as will be introduced shortly). 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 and the data collection can be updated in terms of inserting new documents, updating existing documents, and deleting existing documents.

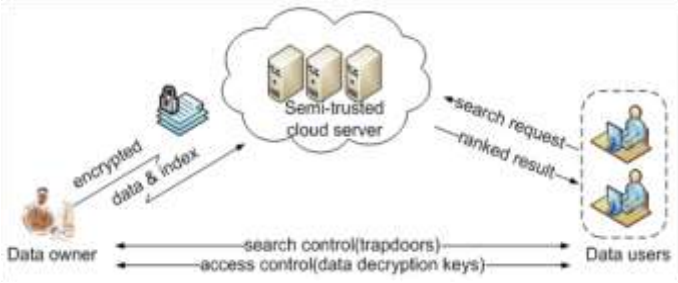

Fig:5.1 Architecture of search over encrypted data

3.2 Design Goals

To enable ranked search for effective utilization of out- sourced cloud data under the aforementioned model, 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 the cloud server from learning additional information from the data set and the index, and to meet privacy requirements specified.

- *Efficiency*. Above goals on functionality and privacy should be achieved with low communication and computation overhead.

3.3 Notations

- **F**—the plaintext document collection, denoted as a set of m data documents F=(F_1,F_2,...F_m)

- **C**—the encrypted document collection stored in the cloud server, denoted as C=(C_1,C_2,...,C_m).

- **W**—the dictionary, i.e., the keyword set consisting of n keyword, denoted as W=(W_1,W_2,...,W_n).

- **I**—the searchable index associated with C, denoted as (I_1,I_2,...,I_m) where each subindex I_i is built for F_i.

- **F_W**—the subset of W, representing the keywords in a search request, denoted as F_W=(W_{j1},W_{j2},...,W_{jt}).

- **T w** —the trapdoor for the search request F w.

- **Fe w** —the ranked id list of all documents according to their relevance to F w.

3.4 Encryption Algorithm

As we are not going to perform any operation on the outsourced files to search of the keywords, we can use any of the existing light weight symmetric key Encryption algorithms and unload the data files to the cloud. We use DES to encrypt the file and then outsource it.

3.5 Indexing

Index is created as a list of mappings [10] which correspond to each keyword. The list for a particular keyword contains details such as:

1. File ids of the files which has the particular keyword
2. Term frequency for each file which denotes the number of times the keyword has occurred in the file. This measures the importance of the keyword in that file.
3. Length of each file
4. Relevance score for each file
5. Number of files that has the particular keyword

Data structures such as B+ trees can be used to store this data. Term frequency, length of the file, number of files for the keyword are used to calculate the relevance score for each file by scoring mechanisms which is discussed later in the Ranking modules.

Whenever a data file is stored, it is preprocessed to generate a index containing the aforesaid details using the keywords extracted (using multiple string matching algorithm)from the data file.

The index creation scheme is as follows:

1. For each wi, that belongs to the keyword set W, generate F(wi) which denotes the file ids that contain wi .
2. For each wi € W For 1 <= j <= |F(wi)|

2.1. Calculate score of the file Fij (with the help of scoring mechanisms discussed later) and store as Sij .

2.2. Store it with file id id(Fij), length of the file |Fij| as (id(Fij) || |Fij| || Sij) in I(wi) which is the index list for the particular word wi

2.3. Update the total number of files that contain the keyword with the index list as (I(wi)||N).

The position of the word in the file is also considered for ranking the file. Hence the file that has the keyword in its title is considered to be more relevant than the files that has the keyword in their content. The relevance score is then stored in the index so that whenever the user requests for a word w, the top 'k' relevant files can be retrieved with this score.

3.6 Ranking

Once the documents are stored and indexed, the next important function is to rank them using details available such that the user retrieves the top „k" most relevant documents. To do so, we need to calculate a numeric score for each file. In the IR community, the most widely used ranking functions are based on the TF X IDF rule, where TF stands for Term frequency which represents the number of times a keyword is present in a file and IDF stands for Inverse Document Frequency which is defined as the ratio of number of file containing the word to the total number of files present in the server. The Ranking Function [5] used: Score (W,Fi) = Σ 1/|Fi| . (1 + ln fi,t) . (1 + N/ft) W: Keyword whose score to be calculated fi,t : Frequency of term in file Fi |Fi|: Length of the file N : Total number of files in the collection.

4.MODULE DESCRIPTION

- **Data User Module:** This module includes the user registration login details.
- **Data Owner Module**:This module helps the owner to register them details and also include login details.
- **File Upload Module:**This module helps the owner to upload his file with encryption using DES algorithm.This ensure the files to be protected from unauthorized users.
- **Rank Search Module:**This module ensure the user to search the file that are searched frequently using rank search.
- **File Download Module**:This module allows the user to download the file using his secret key to decrypt the downloaded data.
- **View Uploaded and Downloaded File:**This module allows the owner to view the uploaded files and downloaded files.

5.EXPECTED RESULTS

5.1. Data Encryption and decryption Result

When DES algorithm is applied on the data then we get encrypted data. and that encrypted data is store on the cloud. User can access the data after downloading and decrypting file. For encryption and decryption keys are provided.

5.2 Ranking Result

When any User request for the data then Ranking is done on requested data using k-nearest neighbor algorithm. For Ranking co-ordinate matching‖ principle is used. After ranking user gets the expected results of the query.

5.3Alert System Results

If any unauthorized User tries to access or updating the data on cloud, then alert will be generated in the form of mail and messages. The alert intimates the authorized user.

6. CONCLUSIONS

In this paper, we solve the problem of post processing overhead and unnecessary network traffic created when Boolean search techniques are used, by introducing the ranked keyword search scheme. The scheme generates indexes that help the user to search for his documents in a secure environment. The files matching the keyword search are further ranked based on the relevant score calculated with term frequency, file length etc. Solve the problem of multi-keyword ranked search over encrypted cloud data, and begin a variety of privacy requirements. Among different multi-keyword semantics, we choose the efficient principle of "coordinate matching", 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, the propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve 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 data set show proposed schemes introduce low overhead on both computation and communication.

This system is currently work on single cloud, In future is will extended up to sky computing & Provide better security in multi-user systems.

7.REFERENCES

1. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2014.

2. Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data Ning Cao,Cong Wang, Ming Li, Member, and Wenjing Lou, IEEE Transaction Parallel AND Distributed Ssystems, VOL. 25, NO. 1, JANUARY 2014

3. N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFO- COM, pp. 693-701, 2012.

4. PRIVACY PRESERVING RANKED KEYWORD SEARCH OVER ENCRYPTED CLOUD DATA Dinesh Nepolean, I.Karthik, Mu.Preethi, Rahul Goyal and M.K. Vanethi. Volume 4, No. 11, November 2013.
5. Secured Multiple-keyword Search over Encrypted Cloud Data. Prof. C. R. Barde,Pooja Katkade, Deepali Shewale, Rohit Khatale

6. Isha Shingari, Sourabh Singh Verma , "Achieving data integrity by forming the digital signature using RSA and SHA-1 algorithm",