ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (855-860), Month: April - June 2015, Available at: www.researchpublish.com

Secure Conjunctive Keyword Ranked Search over Encrypted Cloud Data

Shruthishree M. K, Prasanna Kumar R.S

Abstract: **Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. 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 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 securing conjunctive keyword ranked search over encrypted cloud data.**

Keywords: **Encryption, Keyword ranked search, Pallier Cryptography, Cosine similarity.**

1. INTRODUCTION

Cloud computing is recognized as an alternative to traditional information technology and has been gradually recognized as the most significant turning point in the development of information technology due to its intrinsic resource sharing and low maintenance characters. Cloud computing is an internet based model of computing, where the shared information, software and resources are provided to computers and other devices upon demand. This enables the end user to access the cloud computing resources anytime from any platform such as a cell phone, mobile computing platform or desktop. Clouds are large pools of easily usable and accessible virtualized resources. The data and the software applications required by the users are not stored on their own computers; instead they are stored on remote servers which are under the control of other users. It is a pay-per- use model in which the infrastructure provider by means of customized service level agreements(SLAs)[1]. As cloud computing becomes prevalent, more and more sensitive information's are being centralized into the cloud. Such as emails, personal health records, photo albums, tax documents, financial transactions 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 entitles or even be hacked. To provide data privacy, sensitive data has to be encrypted before outsourcing to the commercial public cloud [2]. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of band width cost in cloud scale systems.

Exploring privacy preserving and effective search over encrypted cloud date is of paramount importance considering the potentially large amount of on demand data users & huge amount of outsourced data document in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability and scalability. Data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides in the 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 search technique allows users to selectively retrieve files of interest. Need for data retrieval is the most frequently occurring task in cloud by the user to the server. Generally cloud

server performs result relevance ranking in order to make the search as faster. Such ranked search system enables data users to find the most relevant information quickly, instead of returning undifferentiated results. 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" [4], i.e., as many matches as possible, is an efficient similarity measure using cosine similarity search among such conjunctive 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.

2. BACKGROUND AND RELATED WORK

Organizations, companies store more and more valuable information is on cloud to protect their data from virus, hacking. The benefits of the new computing model include but are not limited to: relief of the trouble for storage administration, data access, and avoidance of high expenditure on hardware mechanism, software, etc. Ranked search improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency),As directly outsourcing relevance scores will drips a lot of sensitive information against the keyword privacy, We proposed asymmetric encryption with ranking result of queried data which will give only expected data.

A. Existing system:

Existing 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 go through following disadvantage.

1. Single-keyword search with or without ranking.
2. Boolean-keyword search without ranking.
3. Do not get relevant data.
4. It still not adequate to provide users with acceptable result ranking functionality.
5. It cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. ·
6. Shared data will not be secure.

3. PROBLEM FORMULATION

A. Proposed system:

For our system, we 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.[1] 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, we propose a basic MRSE scheme using secure inner product computation and annotation based query, which is adapted from a secure k-nearest neighbor (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 and establish a set of strict privacy requirements.

2) Propose schemes follow the principle of coordinate matching and inner product similarity as well as annotation based query.

To summarize, our proposed system consists of four modules at the basic level of implementation

1. **Setup Module:** Taking security parameter as input, the data owner outputs a symmetric key for encryption.
2. **Build Data and Index Module:** Based on the dataset, the data owner builds a searchable index which is encrypted by the symmetric key generated by the above module and then outsourced to the cloud server. After the index construction, the document collection will be independently encrypted and outsourced.
3. **Trapdoor Generation Module:** With few keywords of interest from the user input search query, this algorithm generates a corresponding trapdoor.
4. **Query Module:** When the cloud server receives a query request, it performs the ranked search on the index with the help of trapdoor, and finally returns the ranked id list of top-k documents sorted by their similarity.

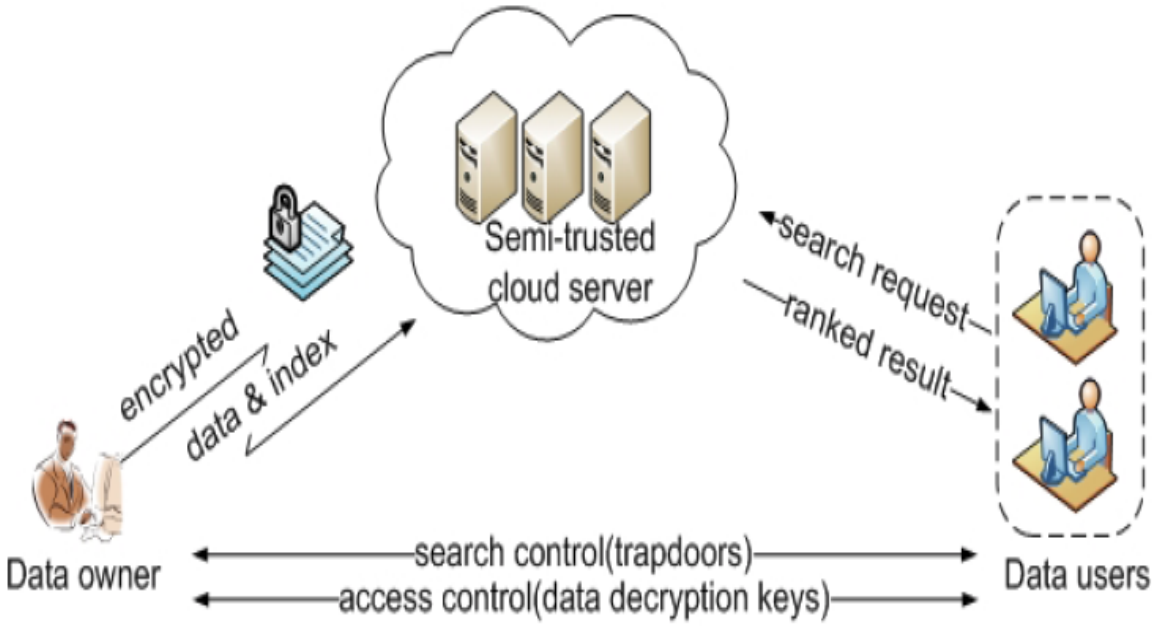

Fig1. Architecture of the search over encrypted cloud data

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 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 (e.g.,coordinate matching). 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.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (855-860), Month: April - June 2015, Available at: www.researchpublish.com

B. SYSTEM DESIGN:

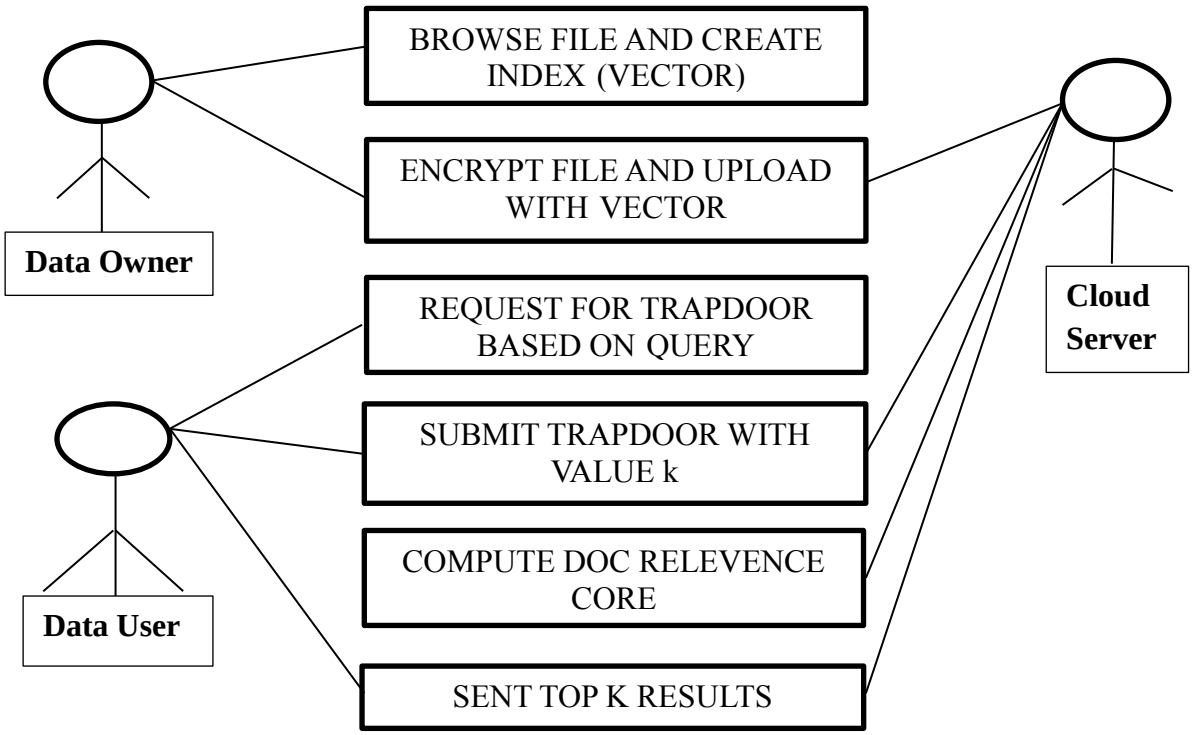

Fig 2 Usecase diagram for system module

C. System Features:

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

1. Secured Conjunctive keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.

2. Privacy: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements.

3. Effectiveness with high performance: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

ALGORITHMS USED:

A. Paillier Cryptosystem:

This algorithm is used to encrypt n decrypt file contents. It is an homomorphic algorithm. Since all the index files are in encrypted form, this is used to operate on cipher text. The Paillier cryptosytem algorithm involves steps like key generation, encryption and decryption.

Step 1: Choose two large prime numbers p and q randomly and independently of each other such that $\gcd(pq, (p-1)(q-1)) = 1$. This property is assured if both primes are of equal length.[1]

Step 2: Compute $n = pq$ and $\lambda = \operatorname{lcm}(p-1, q-1)$.

Step 3: Select random integer g where $g \in \mathbb{Z}_{n^2}^*$

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (855-860), Month: April - June 2015, Available at: www.researchpublish.com

Step 4: Ensure n divides the order of g by checking the existence of the following modular multiplicative inverse: $\mu = (L(g^{\lambda} \bmod n^2))^{-1} \bmod n$, where function L is defined as $L(u) = \frac{u-1}{n}$.

ALGORITHM FOR KEY GENERATION:

(i) For Encryption:

Step 1: Let m be a message to be encrypted where $m \in \mathbb{Z}_n$

Step 2: Select random r where $r \in \mathbb{Z}_n^*$

Step 3: Compute cipher text as: $c = g^m \cdot r^n \bmod n^2$

(ii)For Decryption:

Step 1: Let c be the cipher text to decrypt, where $c \in \mathbb{Z}_{n^2}^*$

Step 2: Compute the plaintext message as: $m = L(c^{\lambda} \bmod n^2) \cdot \mu \bmod n$

B. Rijndael Algorithm:

It is an a symmetric AES

Step 1: Secrete key is converted into byte.

Step 2: Salt data bytes are added to make it complex.

Step 3: Create a complex key byte.

Step 4: Convert the byte into text that forms a complex key

C. *Cosine Similarity Search:*

It is a nearest neighbor search identifies the top *k* relevant documents of the query. This technique is commonly used in predictive analytics to estimate or classify a point based on the consensus of its neighbors. It is used to search top k relevant document on receiving trapdoor from user. It is to calculate results between the document query. Query here refers to user query nothing but trapdoors in encrypted format.

Step 1: Select the document D and query Q.

Step 2: Calculate the Cosine co-efficient similarity between D and Q

RSV(D, Q) = Sim(D, Q).

Step 3: Calculate Euclidian dot formula

cos(d1, d2) = dot(d1, d2) / ||d1||*||d2|

where dot(d1, d2) =dot product i.e., d1[0]*d2[0] + d1[1]*d2[1] + ……

Step 4: |d| = magnitude = sqrt [d1[0]^2 + d1[1]^2+…….]

4. EXPECTED RESULTS

1. Data Encryption and decryption Result:

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

2. Ranking Result:

When any User request for the data then Ranking is done on requested data using inner similarity search. For Ranking —cosine similarity principle is used. After ranking user gets the expected results of the query.

3. Alert 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.

5. CONCLUSION

In this paper, a new framework is proposed for the problem of multi-keyword ranked search over encrypted cloud data, and constructs a variety of security requirements. From various multi-keyword concepts, we choose the efficient principle of coordinate matching as well as annotation based query (Weighted query). We first propose secure inner data computation and annotation based query. Also we achieve effective ranking result using k- nearest neighbor technique. This system is currently work on single cloud, In future is will extended up to sky computing & Provide better security in multi-user systems.

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, no. 1, pp. 50–55, 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, no. 4, pp. 35–43, 2001.

[4] I.H.Witten, A.Moffat and T.C.Bell "Managing Gigabytes: Compressing and indexing documents and images", Morgan Kaughmann Publishing, San Fransisco, 1999.

[5] 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.