Secured Multi-keyword Ranked Search over Encrypted Cloud Data

Vinita Aher [1], Archana Algat [2], Reshma Bhagure [3], Priyanka Malode [4], Priti Lahane [5]

[1] *Student, Information Technology, MET's IOE Nasik, Maharashtra, India*
[2] *Student, Information Technology, MET's IOE Nasik, Maharashtra, India*
[3] *Student, Information Technology, MET's IOE Nasik, Maharashtra, India*
[4] *Student, Information Technology, MET's IOE Nasik, Maharashtra, India*
[5] *Assistant Prof., Information Technology, MET's IOE Nasik, Maharashtra, India*

ABSTRACT

To design schemes which allow to return effective data retrieval when provide multi-keyword query, instead of returning unwanted result. To enable ranked search for effective utilization of out-sourced cloud data under the system and threat model, system design should simultaneously achieve security and performance guarantees of Multi-keyword Ranked Search, Privacy-preserving, Efficiency."Co-ordinate matching" is an intermediate similarity measure which uses the number of query keywords appearing in the document to quantify the relevance of that document to the query.

This technique treats encrypted data as documents and allows a user to securely Search through a single keyword and retrieve documents of interest. Although some recent designs have been proposed to support multi-keyword ranked search over encrypted cloud data (MRSE) search attempt to enrich the search flexibility, they are still not sufficient to provide users with acceptable result ranking functionality. To prevent the cloud server from learning additional information from the data set and the index, and to meet privacy requirement . Such as data privacy, index privacy, keyword privacy, trapdoor unlinkability, access pattern.

Keyword: - *Co-ordinate matching, Privacy-preserving, MRSE, Ranking.*

1. INTRODUCTION

Cloud is used to stored large amount of data which is accessible by multiple users. Cloud computing is an on-demand and Internet-based computing that provides shared processing resources, data and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources. Cloud computing [2][3] is the long dreamed vision, where customers can remotely store their data. So as to retrieve the on-demand high-quality applications and services from cloud by using multi-keyword ranked search technique. To meet the effective data retrieval [5] on large amount of documents, the cloud server performs result relevance ranking[3], instead of returning undifferentiated results. Such multi keyword ranked search system enables users to find the most relevant data quickly, rather than burden to sorting through every match in the content collection. On the other hand, to improve the search result accuracy and for user searching experience enhancing.

To meet the challenge of supporting such multi-keyword semantic with privacy, To propose a basic idea for the MRSE using secure inner product computation, which is taken from a secure k-nearest neighbor (kNN)

technique and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various privacy requirements in two threat models [1]. The Proposed scheme is summarized as follows:

1. For the first time, to explore the problem of multi-keyword ranked search over encrypted cloud data, and establish a strict privacy requirements for such a secure cloud data utilization system.

2. Propose two MRSE schemes based on the similarity measure of "co-ordinate matching"[6] while meeting different privacy requirements in two different threat models.

3. To investigate some further expansion of ranked search mechanism for supporting more search semantics and dynamic data operations.

4. Through analysis investigating privacy and efficiency guarantees of the proposed schemes is given, and experiments on the real-world data set show the proposed schemes indeed introduce low overhead on computation and communication.

2. LITERATURE SURVEY

Organizations, companies store more valuable information is on cloud to protect their credential data from hacking, virus . 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 cost on hardware mechanism, software, etc. Ranked search improves system usability by normal matching files in a ranked order regarding to certain relevance criteria , 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 [1]. Existing searchable encryption[4] schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search[7], without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, having following disadvantage:

- Single-keyword search without ranking.
- Boolean- keyword search without ranking.
- Single-keyword search with ranking.
- Do not get relevant data.

Keyword-based retrieval [5] is a common data service which is widely applied in plaintext scenario. Single-keyword search without ranking system provides searching data with single keyword. Boolean- keyword search [7][8] without ranking provides true or false scenario but without ranking of data. Single keyword search with ranking provides ranking of data with single keyword.

Disadvantages:

- Undesirable security and privacy risks of data.
- Only allowed for Single keyword Search.
- Downloading all the data and Decrypting is impractical.

3. PROPOSED SYSTEM

Searching keyword on relational databases is useful for many users without any technical background. Recently, aggregate keyword search on relational databases was proposed and has attracted interest, so that search multiple keyword on encrypted cloud data is big challenging and this will provide efficient searching, and proposed scheme provide this approach of searching.

In this System solve the problem of searching over encrypted cloud data (MRSE) while preserving strict system-wise security in cloud computing paradigm using multiple queries. Among various multi-keyword semantics, to choose the efficient principle of "co-ordinate matching", it many matches as possible, to capture the similarity between search query and data documents. Specifically, by using "inner product similarity", the number of keywords of query appearing in a document, and evaluate the similarity of that document to the search query result by using "co-ordinate matching" principle.

The set of privacy requirements are improved in threat models of two level. The first time explore the problem of multi keyword ranked search over encrypted cloud data, and provide a set of strict privacy requirements for secure cloud data utilization system. This proposed system introduce low overhead on computation and communication. The working of proposed system is as following:

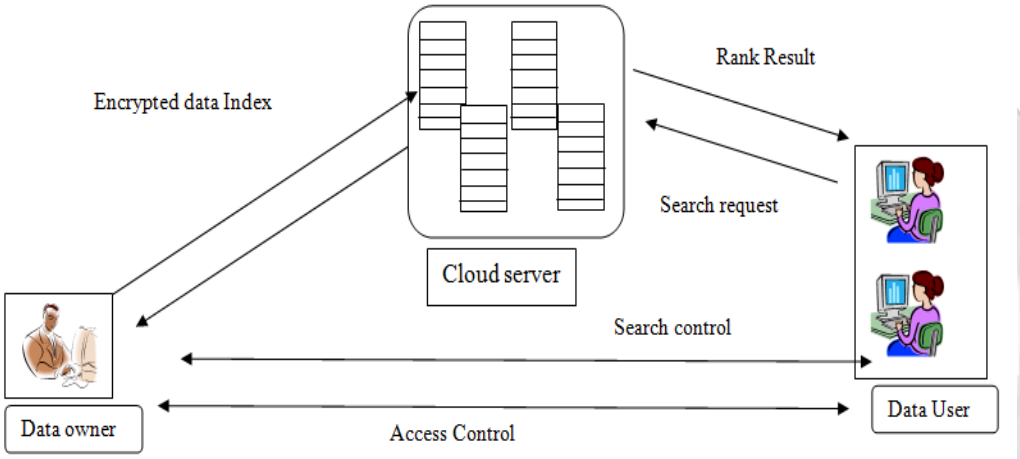

Fig -1: Architecture of the search over encrypted cloud

The owner creates the file that he has to be uploading on the cloud. He selects the important keyword from that file for index generation. After that index is to be generated according to the keyword as well as file is to be encrypt by using some encryption algorithm, finally upload that particular file on the cloud. Suppose user wants to search some files on a cloud then by using keyword he or she search that particular file, after by using particular keyword, cloud server could do some statistical analysis over the search result, And using statistical information analyze 'document frequency' (i.e., the number of documents containing the keyword) And 'Keyword frequency' (i.e., the number of keyword containing by particular documents) is sufficient to identify the keyword with high probability. On the basis of analysis documents gets ranked. After ranking select efficient document and get retrieved.

Advantage:-

- Multi key word ranking for secures the cloud data.
- Searching on the encrypted data will give an expected data.

4. ALGORITHM USED

4.1 AES ALGORITHM FOR ENCRYPTING FILE

AES is a variant of Rijndael which has a fixed block size of 128 bits and key size of 128, 192 or 256 bits by contrast; the Rijndael specification per se is specified with a block and key sizes that may be any multiple of 32 bits, both with a maximum of 128 and a maximum of 256 bits. This algorithm specified here is referred as " AES algorithm". The algorithm is used with the three different key lengths indicated above, and therefore these different "flavors" is referred as "AES-128", "AES-192", and "AES-256".

AES operates on a 4*4 column-major order matrix of bytes, termed the state, although some versions of Rijndael have a larger block size and have additional columns in the state. Most AES calculations are done in special finite field.

The key size used for an AES cipher specifies the number of operations of transformation rounds that convert the input, called the plaintext, into the final output, called the cipher text. The numbers of cycles of repetition are as follows:

- 10 cycles of repetition for 128 bit keys.
- 12 cycles of repetition for 192 bit keys.
- 14 cycles of repetition for 256 bit keys.

Each round consists of several processing steps, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key.

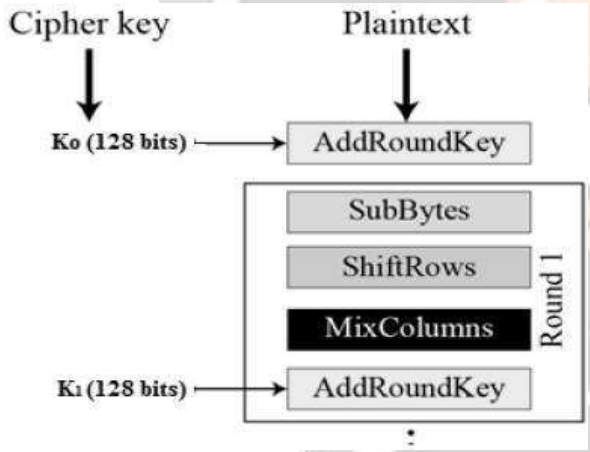

Fig -2: High-level description of the AES algorithm

1. Key Expansion:

Round keys are derived from the cipher key using Rijndael's key schedule.

2. Initial Round:

Add Round Key- Each byte of the state is combined with the round keys using bitwise XOR

3. Rounds

SubBytes - a non-linear substitution step where each byte is replaced with another according to a lookup table.

ShiftRows - a transposition step where each row of the state is shifted cyclically a certain number of steps.

MixColumns- a mixing operation which operates on the columns of the state, combining the four bytes in each column.

4. AddRoundKey:

Final Round (No MixColumns)

SubBytes

ShiftRows

AddRoundKey

4.2 COSINE SIMILARITY

Cosine similarity is a measure of similarity between two non zero vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0,1]. The name derives from the term "direction cosine": in this case, note that unit vectors are maximally "similar" if they're parallel and maximally "dissimilar" if they're orthogonal (= perpendicular). It should not escape the alert reader's attention that this is analogous to cosine, which is unity (maximum value) when the segments subtend a zero angle and zero (uncorrelated) when the segments are perpendicular.The cosine of two non zero vectors can be derived by using Euclidean dot product formula:

$$a .b = \|a\| \|b\| \cos(\theta)$$

Given two vectors of attributes, *A* and *B*, the cosine similarity, $\cos(\theta)$, is represented using a dot product and magnitude as

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

where and are components of vector and respectively.

5. CONCLUSIONS

The solution of multi-keyword ranked search over encrypted cloud data enhances the user to receive the relevant data in the search and establish a variety of privacy requirements. And solve the problem of secure multi keyword top-k retrieval over encrypted cloud data. Among various multi-keyword semantics ,to 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. Specifically use “inner product similarity” i.e., the number of query keywords appearing in a document, then evaluate such similarity measure of that document by using search query.

6. 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, 2011.

[2] 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.

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

[4] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and privacy, 2000

[5]S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology (EDBT '09), pp. 439-449, 2009.

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

[7] D. Boneh and B. Waters, "Conjunctive, Subset, and Range Queries on Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC), pp. 535-554, 2007.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS '10), 2010.