ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING

National Conference on Advanced Innovation in Engineering and Technology (NCAIET-2015)

Alva's Institute of Engineering and Technology, Moodbidri

Vol. 3, Special Issue 1, April 2015

Apriori Based Muti-Keyword Search Over Encrypted Cloud Data

Sujatha Manni[1], Parikshith Nayak[2]

M.Tech Scholar, Department of CS&E, Alva's Institute Of Engineering and Technology, India[1]

Assistant professor, Department of CS&E, Alva's Institute Of Engineering and Technology, India[2]

Abstract: With the data services provided by cloud, many data owners are motivated to outsource their data into public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing. Thus, data search service is very important for this Encrypted cloud data. 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. In this paper, Multi-keyword ranked search is proposed over encrypted cloud data. In which user can search for data with multiple keywords of interest in search query and get ranked result of related documents. Coordinate Matching, i.e., as many matches as possible, to capture the relevance of data documents to the search request. Apriori Algorithm is used for the efficient search in the proposed system for faster access of the related documents to the User.

Keywords: Keyword search, Ranked Search, Apriori, Disjunctive search.

I. 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 [1]. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud [9]; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus privacy preserving and effective search service over encrypted cloud data is of paramount importance. Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability.

On the one hand, to meet 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 [9]. 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" [6], i.e., as many matches as possible, is an efficient similarity measure among such multi-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 other.

II. LITERATURE SURVEY

In the literature Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing [3] Signature is attached to each block in data for integrity. Modified block is computed with new signature to achieve integrity. For security reasons when user leaves the group or

DOI 10.17148/IJIREEICE

misbehaves then they must be revoked from the group. It supports sharing of data within the group but user revocation overhead and if TPA is compromised then user identities will be revealed.

In Compressing and Indexing Documents and Images [4] Inverted indexes are used to evaluate queries which finds the frequency of keywords. Compression of these indexes using byte-wise integer compression gives faster query evaluation.

In this technique Certificate less Public Auditing for Data Integrity in the Cloud [11] public verifier does not need to manage certificate to choose right public key for auditing instead use name and email address of data owner which can ensure the right public key is used but it reveals user details.

In LT Codes-Based Secure and Reliable Cloud Storage Service [2] by allowing a third party to perform the public integrity verification. Data owners are significantly released from periodically checking data integrity, data owners need not be online. In which communication cost increases.

And in Secure and Reliable Data Outsourcing in Cloud Computing [10] Encrypt and upload data between users of the group, to improve data reliability and scalability increase the number of group managers dynamically. Backup manager will allow requests when group manager not available but user revocation persists.

A. Existing System

In Single Keyword Searchable Encryption, Traditional single keyword searchable encryption scheme encrypt the data files and provide searchable index. It supports single keyword search. Public key is used to write the data stored on server but only authorized users with private key can search for the required file, but Multi-keyword search is not possible and does not support Conjunctive search.

And in Boolean Keyword Searchable Encryption it Support both conjunctive and disjunctive search, Conjunctive keyword search returns all-or-nothing, Disjunctive keyword search returns undifferentiated results, which means even only one keyword of interest. It does not qualified for performing ranked search and Multi-keyword is not possible.

III. PROPOSED SYSTEM

In the proposed system Multi-Keyword Ranked Search is proposed for the Encrypted data. And for the efficient searching here coordinate matching is proposed in which as many matches as possible to give search result.

The proposed System display the ranked result for better user experience and Apriori algorithm is applied for efficient retrieval of the data.

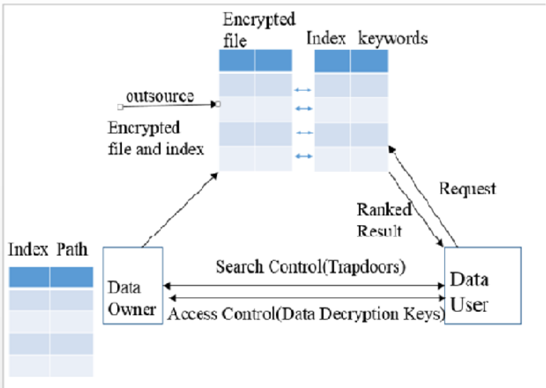

Fig 1: Proposed Architecture of the search over encrypted cloud data.

A. Problem Statement

In the proposed system Multi-Keyword Ranked Search is proposed for the Encrypted data. And for the efficient searching here coordinate matching is proposed in which as many matches as possible to give search result. The proposed System display the ranked result for better user experience and Apriori algorithm is applied for efficient retrieval of the data.

B. Objectives

The main objective is to provide Multi-keyword Ranked Search and replacing the Existing System with AES and Apriori algorithms. And to achieve efficient search.

C. Methodology

In the proposed system the data is encrypted using symmetric key cryptographic algorithms for security and stored in cloud. User can request for the files with multiple keywords, and get ranked result. Use Apriori Algorithm for faster retrieval of the requested data.

IV. ADVANTAGES AND CHALLENGES

A. Advantages

1. Efficient search
2. Less computational overhead
3. Security of sensitive data due to Encryption
4. Faster and relevant search

B. Challenges

Getting the better Response Time is one of the main challenges in the proposed system and efficient search experience for the user.

V. IMPLEMENTATION

So far in the project single keyword is implemented. The files are searched based on the first keyword entered by the user.

This search is also called as normal search where only one keyword is considered and ranked results are returned to the user. The result of this keyword search is not very accurate. To overcome this issue multi keyword search needs be implemented.

DOI 10.17148/IJIREEICE

VI. CONCLUSION AND FUTURE WORK

In the proposed model the search is implemented in two flavors, one is Single keyword Ranked Search which gives user to search the files using a single keyword and in response user gets ranked results where file having highest rank displayed first. Later as an Advanced Search user can search for files with Multiple Keywords of interest and get ranked results which has to be implemented yet. These proposed methods using Apriori Algorithm will access the required files with less response time, so the proposed method is better than the Existing Single keyword searching Methods.

REFERENCES

[1]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 1, JANUARY 2014

[2]. 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.

[3]. C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, 2010.

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

[5]. D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[6]. D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[7]. Y. Hwang and P. Lee, "Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System," Pairing, vol. 4575, pp. 2-22, 2007.

[8]. C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

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

[10]. "Secure and Reliable Data Outsourcing in Cloud Computing" by Ning Cao in July 2012

[11]. "Certificateless Public Auditing for Data Integrity in the Cloud" by Boyang Wang, Baochun Li, Hui Li and Fenghua Li,2012.