IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

Secure Multi-keyword Ranked Search over Encrypted Cloud Data

Vaibhavi Kulkarni[1], Prof. Priya Pise[2]

Student, Computer Engineering, Indira College of Engineering and Management, Pune, India [1]

Professor, Computer Engineering, Indira College of Engineering and Management, Pune, India [2]

Abstract: In this era, Cloud Computing is gaining more importance. Cloud computing provides different services on demand. Due to this more and more data owners are interested to store data on cloud. Cloud computing is a model for on-demand access to a shared pool of configurable computing resources. However, for storing sensitive information, Encryption should be done. So encryption is performed before outsourcing the sensitive data. The data present in the cloud is in encrypted format so searching for appropriate documents is difficult. In this paper, we present Multi-keyword ranked search scheme over encrypted cloud data. Vector space model and TF X IDF model is used for index construction and query generation. Tree based index structure is constructed and "Greedy Depth-first search" algorithm is used for efficient search results. The Secure kNN is used to encrypt index and query vector.

Keywords: Cloud Computing, Cryptography, Multi-keyword ranked search, Searchable encryption.

I. INTRODUCTION

Cloud computing is one of the main domain. Cloud computing relies on sharing ofresources. Cloud computing provides shared resources and data to computers and other devices on demand. Cloud computing is a type of Internet-based computing. Data owners are motivated to outsource their complex data management systems from local sites to public cloud. Cloud computing provides flexibility and it saves money. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search [4], [8]. 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 [4], and rarely differentiate the search results. Multi keyword ranked search achieves more and more awareness for its functional applicability [1], [2], [3], [9]. Some dynamic schemes have been proposed to aid inserting and deleting operations on report collection. These works as it's incredibly possible that the data owners need to replace their knowledge on the cloud server. However few of the dynamic schemes help effective multi keyword ranked search [1].

This paper proposes a secure tree-structured search scheme over the encrypted cloud information, which helps multi- key phrase ranked search and dynamic operation on the file assortment. Specially, the vector area model and the generally-used term frequency inverse file frequency mannequin are combined in the index construction and question generation to provide multi- keyword ranked search.[1] As a way to receive excessive search efficiency, we assemble a tree-situated index constitution and propose a grasping Depth-first Search algorithm headquartered on this index tree. Due to the unique structure of our tree established index, the proposed search scheme can achieve sub-linear search time and handle the deletion and insertion of files. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile make sure accurate relevance ranking calculation between encrypted index and query vectors. [1] To withstand extraordinary attacks in extraordinary chance models, we construct two search schemes: the elemental dynamic multi-keyword ranked search (BDMRS) scheme in the known cipher text mannequin, and the enhanced dynamic multi-key phrase ranked search (EDMRS) scheme within the known history.

II. RELATED WORK

Searchable encryption schemes enable the users to store the encrypted data to the cloud and keyword search over encryptedtext. The searchable encryption schemes can be constructed using public key based cryptography [7]. Early works are single keyword boolean search schemes, which are very simple in terms of functionality. Afterward, a lots of work have been proposed under different threat models to achieve various search functionality, such as single keyword search[4], [8] similarity search, Multi-keyword boolean search. These all schemes allow the users to input multiple query keywords to request suitable documents.All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. [1], [2], [3], [9] Ranked search can enable quick search of the most relevant data. Multi keyword search scheme is very effective in terms of

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

getting most appropriate results. In most of the systems, searchable index tree is constructed based on vector space model and adopted cosine measure together with TF IDF to provide ranking results.

Vector Space Model and Relevance Score Function: Vector space model along with TF×IDF rule is widely used in plaintext information retrieval, which efficiently supports ranked multi-keyword search [34]. Here, the term frequency (TF) is the number of times a given term (keyword) appears within a document, and the inverse document frequency (IDF) is obtained through dividing the cardinality of document collection by the number of documents containing the keyword. In the vector space model, each document is denoted by a vector, whose elements are the normalized TF values of keywords in this document. Each query is also denoted as a vector Q, whose elements are the normalized IDF values of query keywords in the document collection.

As Cloud Computing gets to be predominant, more touchy data are being unified into the cloud. Albeit customary searchable encryption plans permit a client to safely seek over encoded information through pivotal words and specifically recover documents of interest, these procedures bolster just correct catchphrase look. In this paper, interestingly the issue of viable fluffy pivotal word seeks over encoded cloud information while keeping up magic word security is formalized and taken care of. Fluffy essential word look significantly improves framework ease of use by giving back the coordinating documents when clients seeking inputs precisely coordinate the predefined watchwords or the nearest conceivable coordinating records in light of pivotal word similitude semantics, when careful match fizzles. The alter separation to measure essential words closeness and add to two propelled methods on building fluffy watchword sets, which accomplish advanced capacity and representation overheads.

III.PROPOSED SYSTEM

Data owner has a collection of documents F = {f1; f2; :::; fn} that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. In our scheme, the data owner firstly builds a secure searchable tree index I from document collection F, and then generates an encrypted document collection C for F. Afterwards, the data owner outsources the encrypted collection C and the secure index I to the cloud server, and securely distributes the key information of trapdoor generation (including keyword IDF values) and document decryption to the authorized data users. Besides, the data owner is responsible for the update operation of his documents stored in the cloud server. While updating, the data owner generates the update information locally and sends it to the server.

Data users are authorized ones to access the documents of data owner. With t query keywords, the authorized user can generate a trapdoor TD according to search control mechanisms to fetch k encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key. Cloud server stores the encrypted document collection C and the encrypted searchable tree index I for data owner. Upon receiving the trapdoor TD from the data user, the cloud server executes search over the index tree I, and finally returns the corresponding collection of top- k ranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection C according to the received information.

We firstly describe the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme, which is constructed on the basis of vector space model and KBB tree

1) Index Construction of UDMRS Scheme:

In the process of index construction, we first generate a tree node for every document in the collection. These nodes are the leaf nodes of the index tree. After that, the internal tree nodes are generated based on these leaf nodes.

2) Search Process of UDMRS Scheme:

The search process of the UDMRS scheme is a recursive procedure upon the tree, named as “Greedy Depth-first Search (GDFS)” algorithm. Here, the RScore is the relevance score of the document fFID to the query, which is calculated according to Formula (1). The RList stores the k accessed documents with the largest relevance scores to the query. The elements of the list are ranked in descending order according to the RScore, and will be updated timely during the search process.

3) BDMRS Scheme:

Based on the UDMRS scheme, we construct the basic dynamic multi-keyword ranked search (BDMRS) scheme by using the secure kNN algorithm. The BDMRS scheme is designed to achieve the goal of privacy preserving in the known cipher text model.

Security analysis: We analyze the BDMRS scheme according to the three predefined privacy requirements in the design goals:

Index Confidentiality and Query Confidentiality: The BDMRS scheme is resilient against ciphertext-only attack (COA) and the index confidentiality and the query confidentiality are well protected.

Query Unlinkability: The trapdoor of query vector is generated from a random splitting operation, which means that the same search requests will be transformed into different query trapdoors, and thus the query unlinkability is protected. However, the cloud server is able to link the same search requests according to the same visited path and the same relevance scores.

Keyword Privacy: In this scheme, the confidentiality of the index and query are well protected that the original vectors are kept from the cloud server. And the search process merely introduces inner product computing of encrypted vectors, which leaks no information about any

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

specific keyword. Thus, the keyword privacy is protected in the known cipher text model. But in the known background model, the cloud server is supposed to have more knowledge, such as the term frequency statistics of keywords.

4) EDMRS Scheme:
The security analysis above shows that the BDMRS scheme can protect the Index Confidentiality and Query Confidentiality in the known cipher text model. However, the cloud server is able to link the same search requests by tracking path of visited nodes. In addition, in the known background model, it is possible for the cloud server to identify a keyword as the normalized TF distribution of the keyword can be exactly obtained from the final calculated relevance scores.
Security analysis: The security of EDMRS scheme is also analyzed according to the three predefined privacy requirements in the design goals:
Index Confidentiality and Query Confidentiality: Inherited from BDMRS scheme, the EDMRS scheme canprotect index confidentiality and query confidentiality in the known background model. Due to the utilization of phantom terms, the confidentiality is further enhanced as the transformation matrices are harder to figure out.
Query Unlinkability: By introducing the random value, the same search requests will generate different query vectors and receive different relevance score distributions. Thus, the query unlinkability is protected better. However, since the proposed scheme is not designed to protect access pattern for efficiency issues, the motivated cloud server can analyze the similarity of search results to judge whether the retrieved results come from the same requests. In the proposed EDMRS scheme, the data user can control the level of unlinkability by adjusting the value. This is a trade-off between accuracy and privacy, which is determined by the user.
Keyword Privacy: BDMRS scheme cannot resist TF statistical attack in the known background model, as the cloud server is able to deduce/identify keywords through analyzing the TF distribution histogram.

IV. CONCLUSION

This work uses AES algorithm for encrypting data files and GDFS. AES & GDFS increases the data security and improves privacy of data by its commutative nature. Using CRSA, data in a file can be updated dynamically without affecting the overall performance of searching on B-tree. In our proposed system, if encrypted data is modified, re-encrypting for the whole data is not required. This is a desirable feature as it reduces the computation time.

REFERENCES

[1] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, EEE, and Qian Wang, Member, IEEE, "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE transaction on parallel and distributed systems, 2015.

[2] DeepaliD.Rane, Dr.V.R.Ghorpade," Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over Encrypted Cloud Data", International Conference on Pervasive Computing 2015.

[3] Neelam S. Khan, Dr. C. Rama Krishna, AnuKhurana, "Secure Ranked Fuzzy Multi-Keyword Search overOutsourced Encrypted Cloud Data", 5th International Conference on Computer and Communication Technology 2014.

[4] Wang Jie, Yu Xiao, Zhao Ming, Wang Yong, "A Novel Dynamic Ranked Fuzzy Keyword Search Over Cloud Encrypted Data", IEEE 12th International Conference on Dependable, Autonomic and Secure Computing 2014.

[5] QunqunXu, Hong Shen, Yingpeng Sang, HuiTian, "Privacy-Preserving Ranked Fuzzy Keyword Searchsover EncryptedCloud Data", International Conference on Paralleland Distributed Computing2013.

[6] MinghuiZheng, Huihua Zhou, "An Efficient Attack on A Fuzzy Keyword SearchScheme over Encrypted Data", IEEE International Conference on High Performance Computing and Communications 2013

[7] WenjunLuo, Jianming Tan, "PUBLIC KEY ENCRYPTION WITH KEYWORD SEARCHBASED ON FACTORING", IEEE CCIS2012.

[8] Ayad Ibrahim, Hai Jin, Ali A. Yassin, DeqingZou," Approximate Keyword-based Search over EncryptedCloud Data", Ninth IEEE International Conference on e-Business Engineering 2012.

[9] M. Chuah, W. Hu, "Privacy-aware BedTree Based Solution forFuzzy Multi-keyword Search over Encrypted Data", 31st International Conference on Distributed Computing Systems Workshops 2011.