IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

An Efficient Privacy-Preserving Ranked Keyword Search Method

Prof. Supriya Bhosale[1], Akshay Khirolkar[2], Rohit Gaikwad[3], Bhushan Mirase[4], Vinod Bakse[5]

ME (Computer), DYPCOE, Ambi, Maharashtra, India[1]

BE (Information Technology), DYPCOE, Ambi, Maharashtra, India[2,3,4,5]

Abstract: As the volume of data in data center is experiencing a tremendous growth, so owner of data prefer to outsource sensitive and important documents for the purpose of privacy conserving .The documents are stored in encrypted format so it is essential to develop efficient search cipher text technique. In the process of encryption relationship between document is concealed which leads to perform deterioration. In this paper a quality hierarchical clustering (QHC) method is proposed to support searching mechanism and to meet fast searching within cloud environment. In this paper multi keyword ranked search hierarchical clustering index (MRSE-HCI) architecture is used. For search result verification, minimum hash sub tree is used. The proposed method has several advantages over traditional method in document's retrieval and rank privacy.

Keywords: Cloud Computing, Hierarchical Clustering, Security, Cipher text search, multi keyword search, ranked search.

1. INTRODUCTION

As we step into digital era, terabyte's of data is generated worldwide per day, so any data owner who want to outsource their data need to be step up in a cloud computing. Organizations and enterprisers with large amount of data prefer to outsource data in order to reduce data management cost and storage facility. Although cloud service provider's (CSP) claim for their services regarding security, privacy measure security and privacy are major obstacles preventing for wider acceptances.

A traditional way of data encryption makes server side utilization such as searching on encrypted data becomes a challenging task. Therefore, proposing a method which can maintain and utilize this relationship to speed the search phase is desirable. Cloud computing is the open source platform, broad network access, on demand capability, rapid elasticity- are the several advantages of cloud computing but security and privacy are the major issue. In existing solutions data are stored in plain text due to which data is vulnerable for attacks.

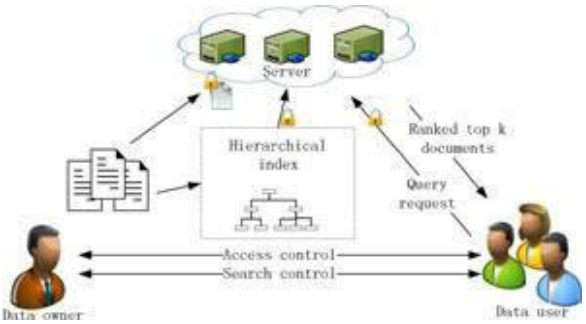

Fig. 1. Architecture of cipher text search

In this paper, a vector space model is used and every document is represented by a vector, which means every document can be seen as a point in a high dimensional space.

2. LITERATURE SURVEY

With the advantage of storage as a service many enterprises are moving their valuable data to the cloud, since it costs less, easily scalable and can be accessed from anywhere any time [1]. The trust between cloud user and provider is paramount. Here security as a parameter is used to establish trust. Cryptography is one way of establishing trust. Searchable encryption is a cryptographic method to provide security. In literature many researchers have been working on developing efficient searchable encryption schemes. This paper explores some effective cryptographic techniques based on data structures like CRSA and B-Tree to increase the level of security. It tried to implement the search on encrypted data using Azure cloud platform [1].

Cloud computing is generating lot of interest to provide solution for data outsourcing and high quality data services. More and more institution, organizations and corporations are exploring the possibility of having their applications, data and their IT assets in cloud [2]. As the data and there cloud's size increases searching of the relevant data is expected to be a challenge. To overcome this challenge, search index is created to aid in faster search. However, search Index creation and computation has been complex and time consuming, leading to cloud-down time there by hindering the swiftness in reacting to data request for mission critical requirements. Focus of this paper is to explain how reusability of search index is

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

helping to reduce the complexity of search index com put action. Search index is proposed to be created using parameters like similarity relevance, user ranking and scheme robustness. User ranking helps to guarantee a keyword is used frequently in the uploaded data [2]. The proposed system defined that the reusability of search index concept reduces cloud consuming time while maintaining the security using searchable symmetric encryption (SSE).The file requested from user is fetched from the cloud, using Two- round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval [2].

Nowadays, more and more people are motivated to outsource their local data to public cloud servers for great convenience and reduced costs in data management. But in consideration of privacy issues, sensitive data should be encrypted before outsourcing, which obsoletes traditional data utilization like keyword based document retrieval. This paper present a secure and efficient multi keyword ranked search scheme over encrypted data, which additionally supports dynamic update operations like deletion and insertion of documents [3]. Specifically, we construct an index tree based on vector space model to provide multi keyword search, which meanwhile supports flexible update operations. Besides, cosine similarity measure is utilized to support accurate ranking for search result. To improve search efficiency, we further propose a search algorithm based on "Greedy Depth first Traverse Strategy". Moreover protect the search privacy, propose a secure scheme to meet various privacy requirements in the known cipher text threat model. Experiments on the real world dataset show the effectiveness and efficiency of proposed scheme [3].

3. OUR CONTRIBUTION

The problem of maintaining the close relationship between different plain documents over an encrypted domain has been investigate and propose a clustering method to solve this problem. We design a search strategy to enhance the rank privacy. This search strategy adopts the backtracking algorithm upon the above clustering method, by using backtracking algorithm we solve any query where it occurs. By applying the Merkle hash tree and cryptographic signature to authenticated tree structure, we provide a verification mechanism to assure the correctness and completeness of search results.

4. DEFINITIONS AND BACKGROUND

4.1 Threat Model

The adverse ability is concluded in two threat models.

Known cipher text model: In this model, Cloud server can get encrypted document collection, encrypted query keywords and encrypted data index.

Known background model: In this model, cloud server knows more information than that in known cipher text model. Statistical background information of dataset, such as the document frequency and term frequency information of a specific keyword, can be used by the cloud server to launch a statistical attack to infer or identify specific keyword in the query which further reveals the plain-text content of documents.

4.2 Design Goals

Search efficiency. The time complexity of search time of the MRSE-HCI is less where scheme needs to be logarithmic against the size of data collection in order to deal with the explosive growth of document size in big data scenario.

Retrieval accuracy. Retrieval precision is related to two factors: the relevance between the query and the documents in result set, and the relevance of documents in the result set.

Integrity of the search result the correctness, completeness and freshness of the document should be maintain.

5. SYSTEM ARCHITECTURE AND ALGORITHM

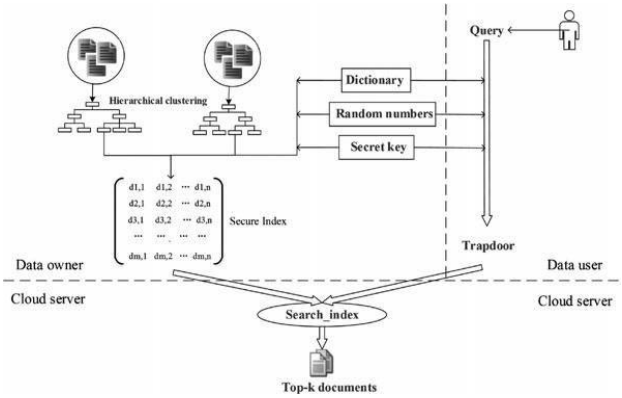

Fig. 2. MRSE-HCI architecture

The MRSE-HCI architecture is depicted by Fig.2.where the data owner is responsible for collecting documents, building document index and outsourcing them in an encrypted format to the cloud server. The cloud server provides a huge storage space, and the computation resources needed by cipher text search. Upon receiving a legal request from the data user, the cloud server searches the encrypted index, and sends back top-k documents that are most likely to match users query. The number k is properly chosen by the data user. Our system aims at protecting data from leaking information to the cloud server while improving the efficiency of cipher text search. In this model, both the data owner and the data user are trusted, while the cloud server is semi-trusted, which is consistent with the architecture in In other words, the cloud server will strictly follow the predicated order and try to get more information about the data and the index. The data user needs to get the authorization from the data owner before accessing to the data.

Until now many hierarchical clustering methods have been proposed but all this method are not comparable to

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

the partition clustering method , K-mean and K-Medoid are popular clustering algorithms but the size of k is fixed here. So we proposed a quality hierarchical clustering algorithm based on dynamic k-means.
Algorithmic Dynamic k-means 1, input the initial set of k cluster C 2, set the threshold TH_{min} 3, while k s not stable 4, generate a new set of cluster center C_0 by k-means 5, for every cluster center C_0, i 6, get the minimum relevance score: min(S_i) 7, if the min(S_i)<TH_{min} 8, add new cluster center: k=k+1 9, go to while 10, Until k is steady

Every cluster is checked whether its size exceeds TH or not. If the size exceeds the cluster will split into child cluster which is formed by dynamic k-means this procedure will be repeated until all the cluster meet the requirement of maximum cluster size.

Algorithm Quality Hierarchical Clustering (QHC)
1, input document and set the size threshold TH
2, build cluster ser C_0 in first level by dynamic k-means
3, while there new cluster set C_i
4, for every cluster $C_{i,j}$
5, if the size $C_{i,j}$ is bigger than TH
6, split this cluster into sub-cluster C_{i+1}
7, until all clusters match the size constraint

The retrieved document has possibility to be wrong because of unstable network and the data may damage due to hardware or software failure, so verifying the authenticity of search result is critical issue in cloud environment. Therefore, the minimum hash sub tree is designed to verify the correctness and freshness of search result.

Algorithm building minimum hash sub-tree(MHST)
1, build hast tree based on hierarchical clustering result 2, for every leaf node I,
3, calculate its hash value: 4, while not tree root
5, for every non leaf node j, 6, calculate its hash value 7, construct node (idj)
8, goto the upper level
9, calculate tree root's hash value:
10, calculate the signature of hash value

6. CONCLUSION

The problem of maintaining the semantic relationship between different plain documents has been explored and given the design method to enhance the performance of the semantic search. For the adaptation to the requirements of data explosion, online information retrieval and semantic search, MRSE-HCI architecture has been proposed. Accordingly, a verifiable mechanism is proposed for the guarantee of correctness and completeness of search results. In addition, the search efficiency and security under two popular threat models has been analyzed. An experimental platform is built to evaluate the search efficiency, accuracy, and rank security. The proposed architecture not only properly solves the multi- keyword ranked search problem, but also brings an improvement in search efficiency, rank security, and the relevance between retrieved documents.

ACKNOWLEDGEMENTS

We would like to thank **Prof. Supriya Bhosale** for her valuable advice and suggestion; we would also like to thank **Prof. Sunita Patil** for her valuable support for timely completion of project. We also like to thank **Prof. Shubhangi Sonone** for her valuable comments on the paper.

REFERENCES

[1] Dynamic Multi-Keyword Ranked Searchable Security Algorithm Using CRSA and B-Tree Prasanna B T#1, C B Akki*2 #Department of ISE, EPCET Associate Professor, Bengaluru, INDIA-560049 *Department of ISE, SJBIT Professor, Bengaluru, INDIA-560060
[2] Reusability of Search Index over Encrypted Cloud Data on Dynamic update Kavitha R1, R J Poovaraghan2 Student, M.Tech, SRM University, Chennai, India1 Assistant Professor (OG), Department of Computer Science, SRM University, Chennai, India2
[3] Dynamic Multi-keyword Top-k Ranked Search over Encrypted Cloud Data Xingming Sun, Xinhui Wang, Zhihua Xia, Zhangjie Fu and Tao Li Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, 210044, China sunnudt@163.com, wxh_nuist@163.com, xia_zhihua@163.com, wwwfzj@126.com