INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 5 ISSUE 3 (2019) PAGES 4401 - 4406
RECEIVED : 12.04.2019 PUBLISHED : 18.05.2019
May 18.2019

DYNAMIC INDEX STRUCTURE BASED ON MULTI-KEYWORD RANKED SEARCH IN CLOUD ENVIRONMENT

[1]S.Kayalvizhi, [2]T.Aravind,
[1]PG Scholar, Department of Computer Science Engineering, Muthayammal Engineering College,
[2]Assitant professor, Department of Computer Science Engineering,
Muthayammal Engineering College.

ABSTRACT

The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. 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 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, and rarely differentiate the search results. We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", its many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

Index Terms: Multi-keyword ranked search over encrypted cloud data, OTP, Product resemblance, Cloud, Data owners

1. INTRODUCTION

Cloud computing has been considered as a new version of company IT infrastructure, that may prepare large useful resource of computing, storage and applications, and allow users to experience ubiquitous, convenient and on-demand community get admission to a shared pool of configurable computing sources with incredible performance and minimum financial overhead . Attracted by these attractive features, each individuals and firms are inspired to outsource their data to the cloud, instead of buying software and hardware to manage the data themselves. Regardless of the various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health statistics, business enterprise finance records, authorities files , and many others) to remote servers brings privacy concerns.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 5 ISSUE 3 (2019) PAGES 4401 - 4406
RECEIVED : 12.04.2019 PUBLISHED : 18.05.2019
May 18.2019

The cloud service providers (CSPs) that hold the information for users may additionally get admission to users' sensitive data with out authorization. A fashionable method to shield the records confidentiality is to encrypt the data earlier than outsourcing but, this could cause amassive fee in terms of information usability. For an example, the existing techniques on keyword-based data retrieval, which are broadly used on the plaintext records, can not be directly implemented on the encrypted data. Downloading all the data from the cloud and decrypt regionally is impractical. On the contrary, more practical special purpose solutions, such as searchable encryption (SE)schemes have made specific contributions in terms of performance, capability and safety. Searchable encryption schemes permit the client to store the encrypted data to the cloud and execute keyword search over ciphertext domain. To this point, abundant works had been proposed under different threat models to obtain various search capability, which include single keyword search, similarity search, multi-keyword boolean search, ranked search, multi-keyword ranked search, and so on. Among them, multi-keyword ranked search achieves increasingly more attention for its realistic applicability.

2. LITERATURE SURVEY

1. Secure and privacy preserving keyword search:

Qin Liu [3] proposed in this paper the pursuit that gives catchphrase protection, information protection and semantic secure by open key encryption. CSP is included in fractional decipherment by lessening the correspondence and computational flying in decoding process for end clients. The client presents the watchword trapdoor encoded by user's private key to CS (Cloud Server) safely and recovers the scrambled records.

Limitations: The correspondence and computational expense for encryption and decoding is more

2. Secure and Efficient Ranked Keyword Search:

Cong Wang [4] proposed seek which unravels preparing overhead, information and catchphrase security, least correspondence and calculation aeronautical. The information proprietor assemble list alongside the catchphrase recurrence based importance scores for documents. Client ask for „w— to cloud server with discretionary k— as Tw utilizing the private key. The cloud server looks the list with scores and sends scrambled document in light of positioned grouping.

Limitations: It doesn't play out numerous catch phrase seeks. Minimal overhead in record building

3. Single Keyword Search over Encrypted data on cloud:

Reachable searchable encryption plan agree to a client to solidly search for over scrambled information through watchwords without first applying decoding on it, the proposed systems bolster just traditional Boolean catchphrase look, without catching any relevance of the records in the query item. At the point when straightforwardly connected in substantial joint information outsourcing cloud environment, they experience next deficiency Limitations: Single-watchword hunt without positioning. Boolean-catchphrase look without

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 5 ISSUE 3 (2019) PAGES 4401 - 4406
RECEIVED : 12.04.2019 PUBLISHED : 18.05.2019
May 18.2019

3. PROPOSED SYSTEM:

Privacy Preserving Multi-Keyword Ranked Search (MRSE):

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. This system defines and solves the challenging problem of privacy-preserving multi keyword ranked search over encrypted data in cloud computing (MRSE). A set of strict privacy requirements for such a secure cloud data utilization system is established. As a hybrid of conjunctive search and disjunctive search, "coordinate matching" [12] 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. When users know the exact subset of the data set to be retrieved, Boolean queries perform well with the precise search requirement specified by the user. In cloud computing, however, this is not the practical case, given the huge amount of outsourced data. Therefore, it is more flexible for users to specify a list of keywords indicating their interest and retrieve the most relevant documents with a rank order.

Fig 1.Rank search over encrypted cloud data

K-NEAREST NEIGHBORS ALGORITHM

In pattern recognition, the *k*-nearest neighbors algorithm (*k*-NN) is a non-parametric method used for classification and regression.[1]In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression:

- In *k-NN classification*, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If *k* = 1, then the object is simply assigned to the class of that single nearest neighbor.

- In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms.

Both for classification and regression, a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/*d*, where *d* is the distance to the neighbor.[2]

The neighbors are taken from a set of objects for which the class (for *k*-NN classification) or the object property value (for *k*-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.

A peculiarity of the *k*-NN algorithm is that it is sensitive to the local structure of the data.

4. SYSTEM MODEL

A. Data Owner Module

This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user.

O-Module that runs in the side of data owner is a documentation which is used by the owner to carry out the owner function in the system and file training phase. Furthermore, this documentation is used by the owner at some stage in the dynamic operations on the cloud data.

B. Data User Module

This module includes the user registration login details. This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required file and register the user details and get activation code in mail email before enter the activation code. After user can download the Zip file and extract that file.

C. Encryption Module:

This module is used to help the server to encrypt the document using RSA Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

An encryption module (also known as a decryption module) was a palm-sized encryption device used to encode messages sent over or through subspace.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 5 ISSUE 3 (2019) PAGES 4401 - 4406
RECEIVED : 12.04.2019 PUBLISHED : 18.05.2019
May 18.2019

D. Rank Search Module

These modules ensure the user to search the files that are searched frequently using rank search. This module allows the user to download the file using his secret key to decrypt the downloaded data. This module allows the Owner to view the uploaded files and downloaded files

E. Cloud server stores

The encrypted document accumulation C and the encrypted searchable tree index I for data owner. In the wake of tolerating the trapdoor TD from the data user, look over the index tree I, in conclusion gives back the relating gathering of top-k situated encoded reports. Also, in the wake of tolerating the update information from the data owner, the server needs to update the index I and document gathering C as per the received information.

5. IMPLEMENTATION

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors. This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable were declared. The definition-use chain method was used in this type of testing. These were particularly useful in nested statements. In this type of testing all the loops are tested to all the limits possible. The following exercise was adopted for all loops:

- All the loops were skipped at least once.
- For nested loops test the inner most loop first and then work outwards.

6. CONCLUSION

In this paper discussed several methods used for secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. Some of the important issues to be handled by the searching technique for providing the data utilization and security are keyword privacy, data privacy, fine grained search, scalability, efficiency, index privacy, query privacy, result ranking, index confidentiality, query confidentiality, query unlink ability, semantic security and trapdoor unlink ability. The limitations for all the searching techniques mentioned in this paper are discussed as well. From the above survey, we can say that a secure, efficient and dynamic search scheme can be developed, which supports not only the accurate multi keyword ranked search but also the dynamic deletion and insertion of documents. Different types of features and classification algorithms can be combined in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the performance.

References

[1] D M. Kuzu, M. S. Islam, and M. Kantarcioglu, Efficient similarity search over encrypted data, in Proc. IEEE 28th Int. Conf. Data Eng., 2012, pp. 11561167.

[2] C. Orencik, M. Kantarcioglu, and E. Savas, A practical and secure multi-keyword search method over encrypted cloud data, in Proc. IEEE 6th Int. Conf. Cloud Comput., 2013, pp. 390397.

[3] K. Ren, C. Wang, and Q. Wang, Security challenges for the public cloud, IEEE Internet Comput., vol. 16, no. 1, pp. 6973, Jan-Feb.

[4] S. Kamara and K. Lauter, Cryptographic cloud storage, in Proc. Financ. Cryptography Data Secur., 2010, pp. 136149.C. Lin, Yulan He, R. Everson ―Weakly Supervised Joint Sentiment-Topic Detection from Text, IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 6, June 2012.

[5] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, "Secure ranked multi-keyword search for multiple data owners in cloudcomputing, in Dependable Syst. Networks (DSN), IEEE 44th Annu. IEEE/IFIP Int. Conf., 2014, pp. 276286.

[6] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, and Minglu Li, Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data, in dependable and secure computing, IEEE July/August 2013.

[7] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, Privacy-preserving multi-keyword ranked search over encrypted cloud data, In Proc. IEEE INFOCOM, Apr. 2011, pp. 829837.

[8] J. Feigenbaum, Y. Ishai, K. Nissim, M. Strauss, and R. Wright, Secure multiparty computation of approximations, ACM Transactions on Algorithms, vol. 2, pp. 435472, 2006.

[9] M. Atallah, F. Kerschbaum, and W. Du, Secure and private sequence comparisons, in Proc. of the WPES03, 2003, pp. 3944.

[10] B S. Yu, C. Wang, K. Ren, and W. Lou, Achieving Secure, Scalable, and Fine- Grained Data Access Control in Cloud Computing, Proc. IEEE INFOCOM, 2010.

[11] A. Rajaraman and D. Ullman, Jeffrey, Mining of massive datasets. Cambridge University Press, 2011.

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