IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10673

Secure Multi-keyword Retrieval system over Encrypted Data

Saurabh Patkar[1], Mohini Patil[2],Pratiksha Khape[3],Geeta Shinde[4], Prof. Shrishail Patil[5]

Department of Computer Engineering,

JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi[1-4]

Assistant Professor, Department of Computer Engineering,

JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi,

Pune, Maharashtra, India[5]

Abstract— Cloud computing has become a major computer trend which allows users to upload their data on a cloud server for flexible services on request. A great deal of outsource information is preserved and collected worldwide. in order to address those circumstances, new and updated security protocols and procedures must be enforced. Until confidential data is moved, it is encrypted on the server to make sure that it cannot be read by someone. Data that is encrypted in a secure cloud typically does not have data-specific and unambiguous information, or entity names. To deal with these problems, we recommend an Expanded Multi keyword Search and Top-k Retrieval algorithm. It provides good precision and high performance. Through calling for a set of query keywords that are part of the document's validity formula, the cloud customer is able to locate information from cloud storage that matches the document's keywords in the query. In addition, we increase the speed of the indexing process we use by extending it so that it produces the most important documents at the top of the records. We show by research that the scheme can bring high precision and efficiency to the process, but less data entry and faster retrieval.

Keywords— cloud computing, encrypted data, Multi-keyword search, Ranking, Top-k retrieval.

I. INTRODUCTION

On-demand high-quality software, networks, and facilities from varied resources are supported by the cloud environment. Cloud computing advantages include quick resource elasticity, omnibus access to the network, cost-based billing, stand-alone pooling of resources, etc. Cloud provides economic savings which encourage people and businesses to outsource their data to the cloud. In addition to a shared interest in cloud computing, the main challenges in outsourcing sensitive and confidential data are data storage and data privacy. Data can be stored in the cloud by consumers and they can sacrifice data confidentiality across the network. Amazon, Google and Drop box are well established cloud computing services. Various encryption approaches are available to protect data security. The data should be encrypted before it outsourced to the cloud in order for the safety of classified information, email messages, personal health reports, tax records, financial transactions etc.

In actuality, most cloud servers do not serve a single data owner; instead, they frequently support several data owners to share the benefits of cloud computing. For example, many volunteers who are pleased with the quality of health care and the government's treatment of patients will give their records to the cloud, thereby helping the governments enforce policy and encourage hospitals in their search for better treatments. They want to protect their own anonymity, so they can encrypt their own personal health data with their private keys. Under this case, only the organisations with the correct credentials have the right to run a search encrypted data scans on data from different data owners. an intricate framework where numerous partners are all taking part in building and preserving a Personal Health Record, I am worried about the complexities of a true multi-owner strategy because it would generate a lot of challenges that will not have been present with the single-owner plans so far.

First of all, the data owner has to stay online to build trapdoors for data users in the single owner scheme. But where a vast number of data owners are included, automatically requiring them to remain online will significantly hinder the system's versatility. No one of us is able to reveal our passwords, but various people choose to use their own. If there are many key pairs on an encrypted file, it is a very time-consuming and resource-intensive to check for them both keys, which also has many security implications. Third, since many data owners are concerned, we must ensure that our infrastructure enjoys optimal protection and scalability. Protecting data privacy in the cloud is not straightforward, as encryption alone can limit cloud's usage in computation. Another key utility feature is data sharing, that is, sharing of data files amongst themselves.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10673

Data users should be able to access their top-of-the-line files in personal health records system, such as patients, on a particular case of different data owners (e.g., health monitors, hospitals, doctors). Similarly, the staff of a company should be able to scan data that was generated by their co-workers as well as third-party suppliers' recent work in a multi-owner model (PRMM) has suggested the solution to solve the multi-word search challenge, which includes a method for protecting privacy in a the PRK search. In the other hand, the operation is inefficiency and costliness can be a problem because different owners need to pay for different cipher text for each question.

II. LITERATURE SURVEY

In this article [1], the author presents a new method called linear, multi-value confidentiality sharing that can greatly improve access policy speech. In addition, each attribute, namely the name and its meaning, is divided into two parts. The most apparent advantage of this scheme is, however, that it is possible to hide sensitive attribute values. And the rights of consumers can be well secured in PHR.

The IB-CPRE-FG security [2] model was developed and IND-CCA safety demonstrated by the authors. Access policies are specified by an access framework in this system. First of all, it is worth building an IB-CPRE scheme to directly benefit AND or gates. Second, there have been proposals for several proxy encryption programmers to catch the key-private property

An author has implemented [3] an adaptive CCA-proxy-attribute-based encryption method in this paper and demonstrated a primary policy attribute-based CCA ciphering scheme that fits well with it. This scheme broadens the concept of proxy re-based encryption, and applies policy attributes to the keys as well. the proposal of the author made is to use policy-attributive key decryption for securing all users' keys supports the theory of the claim that anyone with a key can get access This scheme is considered to be secure in the adaptive model in case of a chosen-cipher attack.

When applied [4] to fine-grained searches, this system helps the owner of the data to limit searches to certain categories of users, permissions and/actions. The premise of this concept is that the data owners encrypt such index keywords only if their attributes meet the search criteria. The definition of a cryptographic attribute-based searchable encryption scheme is introduced by the author who implemented the cipher-text-policy technique.

This is proposed [5] as a technique for data that is not meant to be related so that each user can add search data attributes in his or her own environment This also works in the under Trace-Expand Illustrated how to implement ability can be accommodated with Traceability (more precisely, a traceability function can be made available in this framework.

In this paper [6], authors propose a PHR safeguarding privacy, which promotes thorough control of access and effective revocation. When PHRs are coded, an expressive tree structure can be associated with a cypher text and fine-grained access control is achieved. Data conservation by anonymous protocol is also achieved by authors. This paper implements access structure, fast decryption, data verifiability.

In this paper [7], Author solves the problem left by Fang, Susilo, Ge and Wang by proposing a KP-ABPRE scheme without random oracles. This system enhances the safety model by improving the key query re-encoding and re-encoding query. The author proposes a CCA-based KP-ABPRE system.

An algorithm [8] has finally been mentioned in this article in depth, and while, and also a feasible refinement of the principle of DFA-based FPE has been suggested for the first time. Finally, Lewko et al. [have shown] that the encryption method to be expandable and flexible in the first schemes, which were both unique, in the sense that they exist in the standard model, employ two forms of it. Expanded: Cipher text association requires no formula with significant number of index string length but does have DFA; in this system, a message is decrypted with a key sufficient to decrypt it if and valid cipher text is in the associate able. An encryption algorithm that has been released to a re-keyed/transformed by a semi rusted proxy will generate another cipher text whose key is unique to that proxy, though the resulting string is unique.

In this paper [9], Author addressed the challenging fuse search issue for the encrypted data in a multi-keyword way. Appointed and integrated several innovative concepts to solve search and the complex search problems for multiple keywords. By using locality sensitive hash technique, Author proposes a new multi keyword, fuzzy search scheme. This arrangement makes it possible to match algorithmically instead of extending the index file.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10673

Author proposes [10] a new primitive cryptography known as verifiable keyword-based cloud storage quest over outsourced encrypted files. This primitive helps a data owner to track the search by means of an access management strategy for its outsourced encrypted data while registered data users can outsource the search to the cloud and pressure the cloud to conduct the search consistently (as a cheating cloud can be held accountable.

III. PROPOSED SYSTEM

Proposed system will provide security to data. Secure search protocol is proposed in which cloud server can perform secure search without knowing the actual value of keywords and trapdoors. In multi owner and multi user cloud computing model, three entities are involved such as group manager, group members, cloud server, group manager have collection of files. Group manager build secure searchable index of keyword set. Group manager submit keyword index to server. Group manager encrypt files and outsource encrypted files to cloud server. When group member's wants to search over files from cloud server, he first computes the corresponding trapdoors and submits them then encrypted trapdoors and submits them to cloud server. Cloud server searches encrypted index of group manager and returns top-k relevant encrypted files to the group members. When a group member receives top-K files from cloud server, then group members download files and decrypt these files.

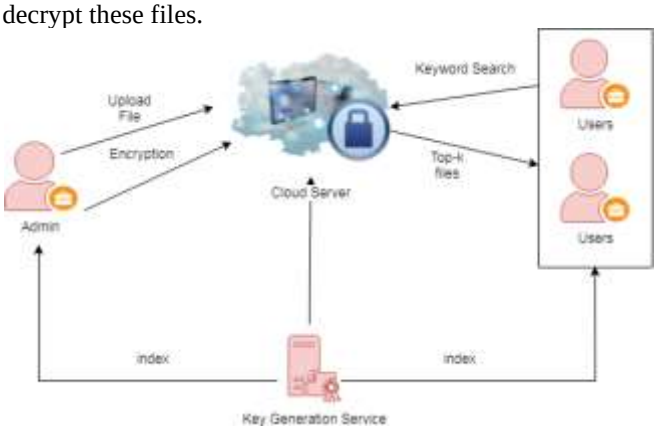

Figure 1. System Architecture

1. TFIDF Algorithm:

Terminology:

1. w — word
2. d — Document
3. N — count of corpus
4. corpus — the total document set

- TF: Term Frequency, which measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (aka. the total number of terms in the document) as a way of normalization:

tf(t,d) = count of t in d / number of words in d

- IDF: Inverse Document Frequency, which measures how important a term is. While computing TF, all terms are considered equally important. However, it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

idf(t) = log(N/(df + 1))

$$\mathrm{tf} - \mathrm{idf}(I_i^j)\log(\mathrm{tf}(I_i^j, d_j) + 1) * \log(\frac{|D|}{1} + \mathrm{df}(I_i^j, D))$$

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10673

2. Encryption and Decryption Algorithm:

The equation of an elliptic curve is given as,

$$y^2 = x^3 + ax + b$$

Few terms that will be used,
E -> Elliptic Curve
P -> Point on the curve
n -> Maximum limit (This should be a prime number)

Key Generation

Key generation is an important part where we have to generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key.
Now, we have to select a number 'd' within the range of 'n'.
Using the following equation, we can generate the public key

$$Q = d * p$$

d = the random number that we have selected within the range of (1 to n-1). P is the point on the curve.
'Q' is the public key and 'd' is the private key.

Encryption

Let 'm' be the message that we are sending. We have to represent this message on the curve. This has in-depth implementation details.
Consider 'm' has the point 'M' on the curve 'E'. Randomly select 'k' from [1 – (n-1)].
Two cipher texts will be generated let it be C1 and C2.

$$C_1 = k * p$$
$$C_2 = M + k * Q$$

C1 and C2 will be send.

Decryption

We have to get back the message 'm' that was send to us,

$$M = C_2 - d * C_1$$

M is the original message that we have send.
How does we get back the message?

$$M = C_2 - d * C_1$$

M can be represented as '$C_2 - d * C_1$

$$C_2 - d * C_1 = (M + k * Q) - d * (k * p) \; (C_2 = M + k * Q \text{ and } C_1 = k * p)$$
$$= M + k * d * P - d * k * P$$
$$= M(\text{Orignal Message})$$

CONCLUSION

We propose a top-k search and retrieval, an Extended Multikeyword system to search for multi-keyword queries for top cloud data recovery. The updated similarity score system is used to calculate and extract related data from cloud data outsourced to increase the precision of the scan.
The proposed procedure effectively improves data access, which decreases the search and recovery time relative to other processes. Through the study of results, we prove that the device proposed guarantees significant security and a high-speed efficient retrieval.

ACKNOWLEDGMENT

Express my true sense of gratitude, sincere and sincere gratitude to my guide to the project Prof. Shrishail Patil for his precious collaboration and guidance that he gave me during my research, to inspire me and provide me with all the laboratory facilities, This it allowed me to carry out this research work in a very simple and practical way. I would also like to express my thanks and thanks to our coordinator, Prof. Yogendra Patil, HOD. Prof. Gayatri Bhandari and Principal Dr. T. K. Nagaraj and all my friends who, knowingly or unknowingly, helped me during my hard work.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10673

REFERENCES

[1] L. Zhang, G. Hu, Y. Mu, and F. Rezaeibagha, "Hidden ciphertext policy attribute-based encryption with fast decryption for personal health record system," IEEE Access, vol. 7, pp. 33202–33213, 2019.

[2] C. Ge, W. Susilo, J. Wang, and L. Fang, "Identity-based conditional proxy re-encryption with fine grain policy," Computer Standards & Interfaces, vol. 52, pp. 1–9, 2017.

[3] C. Ge, W. Susilo, L. Fang, J. Wang, and Y. Shi, "A cca-secure key-policy attribute-based proxy re-encryption in the adaptive corruption model for dropbox data sharing system," Designs, Codes and Cryptography, pp. 1–17, 2018.

[4] H. Yin, J. Zhang, Y. Xiong, L. Ou, F. Li, S. Liao, and K. Li, "Cp-abse: A ciphertext-policy attribute-based searchable encryption scheme," IEEE Access, vol. 7, pp. 5682–5694, 2019.

[5] Y. Miao, X. Liu, K.-K. R. Choo, R. H. Deng, J. Li, H. Li, and J. Ma, "Privacy-preserving attribute-based keyword search in shared multi-owner setting," IEEE Transactions on Dependable and Secure Computing, 2019.

[6] C. Ge, W. Susilo, J. Wang, Z. Huang, L. Fang, and Y. Ren, "A key- policy attribute-based proxy re-encryption without random oracles," The Computer Journal, vol. 59, no. 7, pp. 970–982, 2016.

[7] K. Liang, M. H. Au, J. K. Liu, W. Susilo, D. S. Wong, G. Yang, T. V. X.Phuong, and Q. Xie, "A dfa-based functional proxy re-encryption scheme for secure public cloud data sharing," IEEE Transactions on Information Forensics and Security, vol. 9, no. 10, pp. 1667–1680, 2014.

[8] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in IEEE INFO COM 2014-IEEE Conference on Computer Communications, pp. 2112-2120, IEEE, 2014.

[9] Q. Zheng, S. Xu, and G. Ateniese, "Vabks: verifiable attribute-based keyword search over outsourced encrypted data," in Infocom, 2014 proceedings IEEE, pp. 522–530, IEEE, 2014.

[10] K. Liang and W. Susilo, "Searchable attribute-based mechanism with efficient data sharing for secure cloud storage," IEEE Transactions on Information Forensics and Security, vol. 10, no. 9, pp. 1981–1992, 2015.