International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

An efficient ranked multi-keyword search, fuzzy keyword Search for multiple data owners over encrypted cloud data

Ms. Shireen I. Kudle[1] and Dr. Swapnaja A. Ubale[2]

[1]Department of Computer Engineering , [2]Department of Information Technology Zeal College of Engineering & Research, Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

With the happening to distributed computing, it has ended up being giving security to data. In the existing system, data users can access them without authentication Cipher text-Policy Attributebased Encryption (CP-ABE) can be utilized to conduct fine-grained and owner-centric access control. In a cloud computing system, we have developed the system providing security for information. In this system, the data owner can upload the different files using the AES algorithm in the encrypted format for maintaining security. For insurance concerns, secure endeavour over scrambled cloud data has propelled a couple of research works under the single proprietor model. In our system, we developed this system for multiple owners' models with different functionality. In this system, implemented plans to tree-based ranked multikeyword search scheme for multiple data owners (TBMSM), to efficiently develop novel search protocol based on bilinear pairing, which enables different data owners to use different keys to encrypt their keywords and trapdoors. In the cloud server module, view all users, data owners, and all encrypted files also. The user also views the attacker of the system. The data user can search over encrypted data using the hash value md5 algorithm. Data Users can also fuzzy keyword algorithm search technique also used moreover; Users can download the file at a particular place only as well as at particular times only. Also, find out an attacker of the system if any user enters three times wrong key.

Keywords: *Advanced Encryption Standard, Attacker, Fuzzy Keyword Search, Hash Value, Ranked keyword Search*

Introduction

Encryption on touchy information before re-appropriating can save information protection. Be that as it may, information encryption makes the customary information usage administration dependent on plaintext keyword search a difficult issue. The category of the search function includes secure ranked multi-keyword search, and similarity search. A different data owner can upload this any file in an encrypted format then encrypted index is generated. This encrypted index goes to the administrator system. Different data owners can upload files on a cloud so every file is stored in an encrypted format. In this system, users can search that file with different searching techniques like fuzzy keyword search, Hash value search and multikeyword search. Data owners uploaded file store on a cloud server an answer for this issue is to download all the hidden information and make the first information utilizing the hidden key, yet this is not practical because it makes additional overhead. In this system, the Data owner can file upload in a different file in encrypted format using AES 128 bit. The user can search any file then after checking authentication users get file. If the user wants to download that file then data user requests to the data owner. After getting the request, the user can send the key to download the file. Hence, propose when user search keywords that time give the security and demonstrate the bring about positioning structure to make simple cloud servers to perform safe excluding knowing the real value of both keywords and trapdoors, We proposed fuzzy keyword search, using this we can easily search the information. We also introduced any file that can download from a particular location only. Also, find out an attacker of the system if any user enters three time wrong key.

Objective and Scope

- File search using multikeyword search as well as search using hash value over encrypted data.
- File upload in different format like in encrypted format.
- Users can search the encrypted data using the fuzzy keyword.
- User file download at a particular place and particular time so the system becomes more secure. Find out the attacker of the system.

Problem Statement

Encryption of sensitive data before outsourcing can preserve data privacy. Nonetheless, information

encryption makes the conventional information use administration dependent on plaintext watchword search a difficult issue. The category of search function contains a secure ranked multi-keyword search and similarity search. Nonetheless, every one of these plans is constrained to the single-proprietor model. Search over encrypted data using hash value and attacker problem occurred in the existing system.

Review of Literature

Ravindra R. et.al [1] state that as cloud computing is exceptionally commanding innovation lately, whole delicate data is being put away onto the cloud. For keeping up information secrecy, touchy information is by and large scrambled, which makes viable information usage an extremely mind-boggling task. The Existing accessible encryption plans give a clear way to deal with secure hunt over encoded information utilizing catchphrases and recovering the vital documents of intrigue. While these methods bolster just careful fuzzy keyword search. That is, there is no acknowledgment of slight mistakes and organization irregularities which are run of the mill client looking through conduct. On account of this disadvantage, the current procedures get inconsistent in distributed computing, influencing the framework convenience. This makes the client looking through encounters extremely baffling and brings about low framework effectiveness. This system incorporates the formalization and arrangement of the issue of viable fuzzy keyword search over encoded cloud information just as saving catchphrase protection. Conquering the downsides of customary inquiry strategies, the fuzzy keyword search supports the framework convenience by creating the coordinating and applicable records when clients' looking through sources of info precisely coordinate the predefined catchphrases or the nearest conceivable coordinating or important documents dependent on catchphrase closeness semantics, when definite match falls flat.

Sofiane Mounine Hemam et.al [2] proposed that in this system, research the heap adjusting between hubs in the volunteer distributed computing. We propose another methodology which depends on cloning a cloud administration on at least one hubs when the quantity of the client solicitations will be significant at a given time. Our answer permits a superior framework unwavering quality and lessens the reaction time of the clients by appropriating their solicitations between the volunteer hubs. Shockingly, the replication of cloud administrations limits the extra room limit. Therefore, we propose a second calculation that chooses and erases the imitations of a cloud administration without the corruption of the heap adjusting, utilizing for this the Markov Chain Models. The trial results, because of the PeerSim test system, show that the proposed calculations can viably accomplish great execution (load adjusting) and improve the reaction time.

Wan-Ni Shih and Tai-Lin Chin [3] describe the developing ubiquity of distributed computing, an ever increasing number of information proprietors re-appropriate their information to cloud stockpiles due to the accommodation for the executives. Since cloud stockpiles are normally run by outsider specialist organizations, information security is unquestionably a significant issue. A basic method to ensure the security is to scramble the information before re-appropriating them to the cloud. Some current procedures give watchword search over encoded information. Be that as it may, the greater part of the strategies utilizes a long record for the archives and do correct watchword search over the dataset. Those procedures may experience the ill effects of the calculation and correspondence time just as the extra room. Right now, catchphrase search over scrambled information plot is proposed to improve the hunt execution as far as calculation time and required space. Basically, the relations between the reports and chose catchphrases are first investigated. At that point, the key data in the watchword archive relations is extricated to diminish the record files. The pertinence of clients' question and the records is quantitatively assessed to choose the last reports of intrigue. Analyses on genuine world dataset show that the proposed conspire adequately improve the calculation execution.

W. Zhang et al [4] explain the various advantages and uses of cloud storage. It helps in a large amount of storage space without the use of the resources in real-time. The resources used to carry out storage operations are never owned by users and hence the users have to pay peruse. This strategy helps in ecofriendly and green computing as well. This is possible as the resources that were used earlier for the storage, such as the servers, systems, space, cooling systems and many more are no longer required. The capacity frameworks are for all intents and purposes present for the client and consistently present in an alternate area. The framework additionally talks about the diverse assistance models; specifically, Infrastructure as a Service, Platform as a Service and Software as a Service. It also discusses the different types of clouds such as public cloud, private cloud, hybrid cloud, and the community cloud.

Hongwei Li et.al [5] introducing utilizing cloud computing, individuals can store their information on remote servers and permit information access to open clients through the cloud servers. As the outsourced data are likely to contain sensitive private information, they are commonly encoded before transferred to the cloud.

This, in any case, significantly restrains the ease of use of redistributed information due to the difficulty of looking over the encoded information. In this system, we address this issue by building up the fine-grained multi-keyword search plots over encoded cloud information. Our unique commitments are three-overlay. To begin with, we present the importance scores and inclination factors upon keywords that

empower the exact catchphrase search and customized client experience. Second, we build up a useful and very efficient multi-catchphrase search conspire. The proposed plan can bolster convoluted rationale search the mixed "AND", "OR", and "NO" operations of keywords. Third, we further utilize the classified sub-word references system to accomplish better efficiency on file building, trapdoor producing and question. In conclusion, we examine the security of the proposed plans as far as confidentiality of records, protection insurance of file and trapdoor, and unlink capacity of the trapdoor. Through broad investigations utilizing this present reality dataset, we approve the exhibition of the proposed plans. Both the security examination and trial results exhibit that the proposed plans can accomplish a similar security level contrasting with the current ones and better execution as far as usefulness, inquiry unpredictability and efficiency.

Wei Zhang et.al [6] state that cloud computing provides abundant benefits including simple access, diminished expenses and flexible asset the executives. For security concerns, touchy information must be scrambled before re-appropriating, which obsoletes conventional information usage dependent on plaintext catchphrase search. Hence, building up a protected pursuit administration over encoded cloud information is of principal significance. There are a few examine worried about this issue. Be that as it may, every one of these plans depends on a solitary cloud model that has the danger of single purpose of disappointment, misfortune, and defilement of information, loss of accessibility and loss of security. In this system, we investigate the issue of secure appropriated keyword searches in a multi-cloud worldview. In light of this model, we propose two plans. In the plot I, we propose to cross-store all encoded file cuts, keywords, and keys. In plot II, we deliberately develop a catchphrase circulating technique and a file conveying methodology. Further, we expand the two plans with Shamir's mystery plans to accomplish better accessibility and heartiness. Broad investigations on genuine world datasets confirm the efficacy and efficiency of our plan.

Zhiyong Xu et.al [7] proposed that cloud processing is getting progressively pervasive as of late. It acquaints an efficient route to accomplish the board's exibility and monetary investment funds for distributed applications. To exploit registering and capacity assets offered by cloud specialist organizations, information proprietors must re-appropriate their information onto open cloud servers that are not inside their confided in spaces. In this manner, information security and protection become a major concern. To forestall data revelation, touchy information must be scrambled before transferring onto the cloud servers. This makes plain content keyword inquiries inconceivable. As the aggregate sum of information put away in broad daylight mists collects exponentially, it is trying to help efficient catchphrase based inquiries and rank the coordinating outcomes on encoded information. Most current works just consider single catchphrase inquiries without proper positioning plans. The multi-keyword inquiry issue was being viewed as of late. MRSE [1] is one of the first inquire about attempts to define and address the issue of compelling yet secure positioned multi-keyword search over encoded cloud information. Nonetheless, the catchphrase word reference utilized in MRSE is static and must be remade when the number of keywords in the lexicon increments. It additionally has extreme out-of-request issues in the coordinating outcomes and doesn't consider the catchphrase get to frequencies, which extraordinarily influences its convenience. In this system, we propose a novel methodology, called MKQE, to address these issues. Just minor changes in the lexicon structure must be done when additional keywords are presented.

Qin Liu et.al [8] proposed that distributed computing as a

developing innovation pattern is required to reshape the advances in data innovation. In this system, we address two central issues in a cloud situation: protection and efficiency. We first survey a private catchphrase based file recovery conspire proposed by Ostrovsky et. al. At that point, in light of an aggregation and distribution layer (ADL), we present a plan, named efficient information retrieval for ranked query (EIRQ), to additionally decrease questioning expenses brought about in the cloud. Inquiries are classified into numerous positions, where a higher positioned inquiry can recover a higher level of coordinated files. Broad assessments have been led on an explanatory model to inspect the viability of our plan.

Proposed System Approach

This proposed system consists of mainly three modules of data owners, data users and cloud server. In our proposed system first data owner do registration with login with proper authentication. Data owner upload files using the AES algorithm in an encrypted format, this file is stored on the cloud and also upload the file with hash value using the MD5 algorithm.

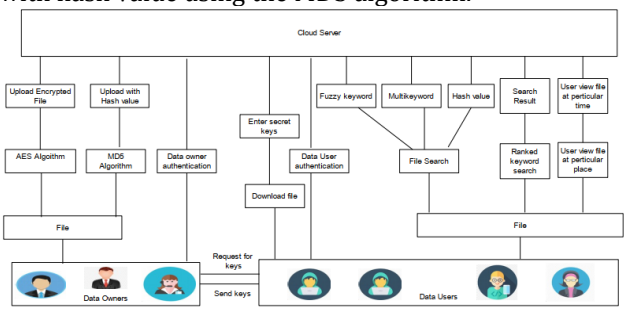


Fig.1 Block Diagram of Proposed System

Data User registration and login with proper authentication, after login user search different file with Multikeyword search, Fuzzy Keyword search and Search using hash value also. Searching the user view

the file and send the request to a particular data owner. Data owners accept the request and send secret keys to the user. Data user enters secret keys and downloads file at a particular time and particular place. If the user entered 3 times wrong key user becomes attacker also cloud server view the attackers. Users can view ranked multikeyword search also.

Mathematical Model

Mathematical Model in Equation format Notation

- **TFU=**Total number file upload
- **FU1=**Number of file upload 1
- **FU2=**Number of file upload 2
- **FU3=**Number of file upload 3
- **TDF=**Total number file download
- **DF1**=Number of file download 1

DF2=Number of file download 2

- **DF3=**Number of file download 3
- **FSK=**Total file search by keyword
- **FSK1=**Total file search by keyword 1
- **FSK2=**Total file search by keyword 2
- **FSK3=**Total file search by keyword 3

For calculating total number of file uploaded following equation 1 is used

Total number of file upload= Number of file upload 1+ Number of file upload 2+…..+ Number of file upload N ∑ TFU = ∑ FU1+ ∑ FU2+…+∑ FUN …… equation 1

For calculating the total number of file downloaded following equation 2

Total number of file download= Number of file download 1+ Number of file download 2+….. +Number of file download N

∑ TDF= ∑ DF1+ ∑ DF2+…+∑ DFN …… equation 2

For calculating total number of file searched by keyword following equation 3 is applied

Total number of file search by keyword= Number of file search by keyword 1+ Number of file search by keyword 2+…..+Number of file search by keyword N ∑ FSK= ∑ FSK1+ ∑ FSK2+…+∑ FSKN ……equation 3

Algorithms In Pseudo Code

1. AES Algorithm for Encryption.

AES is an iterative instead of Feistel cipher. It is based on two common techniques to encrypt and decrypt data knows as substitution and permutation network (SPN). SPN is a number of mathematical operations that are carried out in block cipher algorithms. AES can manage 128 bits (16 bytes) as a fixed plaintext square size. These16 bytes are represented in 4x4 matrixes and AES operates on a matrix of bytes. Likewise, another urgent element in AES is number of rounds. The quantity of rounds is depended on the length of key. There are three diverse key sizes are utilized by AES calculation to encode and unscramble information, for example, (128,
192 or 256 bits). The key sizes choose to the quantity of rounds, for example, AES utilizes 10 rounds for 128piece keys, 12 rounds for 192-piece keys and 14 rounds for 256-piece keys. **Input:**

- 128_bit /192 bit/256 bit input (0, 1)
- Secret key (128_bit) +plain text (128_bit).

Process:

- 10/12/14-rounds for-128_bit /192 bit/256 bit input
- Xor state block (i/p)
- Final round:10,12,14
- Each round consists:sub byte, shift byte, mix columns, add round key.

Output:

- cipher text(128 bit)

2. MD5 (Message-Digest Algorithm)

MD5 or "message digest 5" calculation was structured by Professor Ronald Rivest. Rivest is an educator at MIT who likewise developed RSA, RC5 and the MD-message digest hashing capacities. MD5 is a single direction hashing capacity. So by definition, it ought to satisfy two properties. One, it is one way which implies one can make a hash an incentive from a message however can't reproduce the message from the hash esteem. Two, it ought to be without impact that is two unmistakable messages can't have a similar hash esteem.

- **Steps 1:** A message digest calculation is a hash work that takes a piece grouping of any length and delivers a piece succession of a fixed little length.
- **Steps 2:** The yield of a message digest is considered as an advanced mark of the information.
- **Steps 3:** MD5 is a message digest calculation creating 128 bits of information.
- **Steps 4:** It utilizes constants inferred to trigonometric Sine work.
- **Steps 5:** It circles through the first message in squares of 512 bits, with 4 rounds of activities for each square, and 16 tasks in each round.
- **Steps 6:** Most present day programming dialects gives MD5 calculation as inherent capacities.

3. Fuzzy Keyword Search:-

Fuzzy keyword search significantly upgrades frameworkconvenience by restoring the coordinating records whenclients looking through data sources precisely coordinate the predefined catchphrases or the nearest conceivablecoordinating documents dependent on keyword similitude semantics, when careful match comes up short.

Inputs:-

- C= (F1, F2… Fn)

- W= {W1, W2… Wn}
- Edit distance *d*
- A searching input (*w, k*) (k<=d)

Process:-

For Normal Search Set Up

- Π=(Setup(1λ), Enc(*sk,* ·), Dec(*sk,* ·))
- *Twi = f* (*sk, wi*)

For Fuzzy Keyword

The wildcard-based fuzzy set of wi with edit distance d is denoted as

Swi,d={Swi,0, Swi,1, · · · , Swi,d}.

- d=1 (2L+1)*26+1
- d=2 C1L+1+C1L*C1L+2C2L+2

For Searching Input:-

- Π= (Setup(1λ), Enc(*sk,* ·), Dec(*sk,* ·))
- *Twi = f* (*sk, wi*) *Tw'i= f* (*sk, w'i*) for each *w'i*∈ *Swi,d*
- **Step 1:**FIDwi Enc(sk,FIDwi||wi){({Tw'i}w'i ∈Swi,d, Enc(sk,FIDwi||wi))}wi∈W
- **Step 2:** {Tw'} w' ∈ Sw,k
- **Step 3 :**Enc (sk, FIDwi||wi)

Output:-

Get Expected result which is search by the user.

Comparative Results

In our experimental setup, table no. 8.1, shows the number of file upload and file download. In our system 50 total number of files. In that 30 were number file upload and 20 were downloading of files.

Table 8.1: No. of Upload and download files

Sr. No.	Number of File Upload	Number of File Download
1	30	20

In our experimental setup, in table no.8.2, User can search different file with different keywords so get information about how to user can search any files. In that 35 users search by 1st keyword, 25 users search by 2nd keyword and 28 users search by 3rd keyword.

Table 8.2: No. of File Search by keyword

Sr. No.	No. of File Search by Keyword 1	No. of File Search by Keyword 2	No. of File Search by keyword 3
1	35	25	28

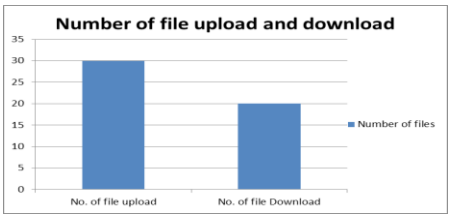
In our experimental setup, table no. 8.3, shows the different encryption algorithm and time for that file for encryption. In our system comparison between AES algorithm with DES algorithm, MD 5 algorithm, SHA 256 algorithm and blowfish algorithm is shown.

Table 8.3: No. of Upload and download files

Sr. No.	Algorithm Name	Time for Encryption (Millisecond)
1	AES	10
2	DES	15
3	MD5	12
4	SHA 256	17
5	Blowfish	20

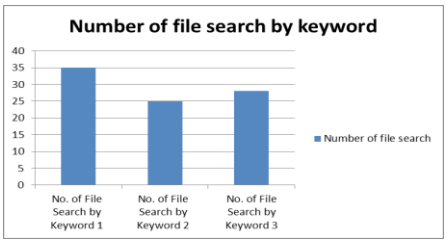
Results

From the above data, as shown in graph 9.1, the total numbers of files were 50. The numbers of files found to be uploading were 20 and downloading files were 30.



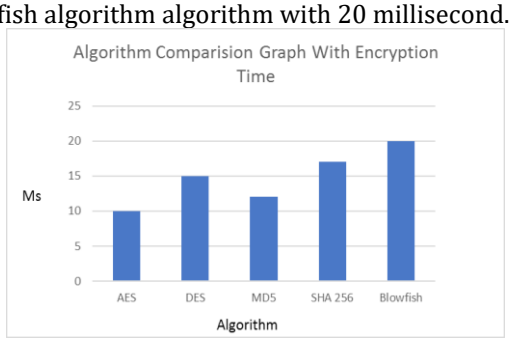
Graph 9.1: Number of file upload and download

In our experimental setup, as shown in graph 9.2, From above table data, In graph, we can see the no. of file search keyword by keyword 1, no of file keyword 2 and no of file keyword 3 in the graph; we see 35 files search by keyword 1, 25 files search by keyword 2 and 28 files search by keyword 3 by different users are shown in the graph.



Graph 9.2: Number of file search by keyword

In our experimental setup, graph no. 9.3, shows the different encryption algorithm and time for that file for encryption. In our system comparison between AES algorithm with 10 millisecond, DES algorithm with 15 millisecond, MD 5 algorithm with 12 millisecond, SHA 256 algorithm algorithm with 12 millisecond and blowfish algorithm algorithm with 20 millisecond.



Graph 9.3: Algorithm comparison graph with encryption time

Conclusion

In this study, we consider a multiple data owners model in cloud computing and propose an efficient ranked multikeyword search scheme over encrypted data. In the existing system, data user can access the without authentication Cipher text-Policy Attribute-based Encryption (CP-ABE) can be utilized to conduct finegrained and owner-centric access control. In this system, user can search using different searching techniques like multikeyword search, Fuzzy keyword search, and Hash Value search. Upload a file in encrypted format for maintain the security. User can download any file in particular place and particular time only.

Future Work

In future, we can upload data with images and videos also.

Acknowledgment

This work is supported in a multi-keyword search, a fuzzy keyword search for multiple data owners over encrypted cloud data fields in India. Authors are thankful to the Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

References

[1] Ravindra R. Ghugare, Pranjuli Yavatkar, Nikita Patil,Sneha Kale , Fuzzy Keyword Search over Encrypted Data in Cloud Computing, International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 04, Issue 01; January - 2018

[2] Sofiane Mounine Hemam, Ouided Hioual, Abbes Laghrour, Load Balancing Between Nodes in a Volunteer Cloud Computing by Taking into Consideration the Number of Cloud Services Replicas, 3rd International Conference of Cloud Computing Technologies and Application (CloudTech) 2017

[3] Wan-Ni Shih and Chin, Approximate Multi-Keyword Rank Search on Encrypted Cloud Data, IEEE Global Communications Conference 2017.

[4] H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou, and X. Shen, Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data, in IEEE Transaction on dependable and secure computing, vol 13, no. 3, May/June 2016.

[5] W. Zhang, Y. Lin, S. Xiao, J. Wu, and S. Zhou, Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing, IEEE Transactions on computers, vol. 65, no. 5, May 2016.

[6] Wei Zhang Sheng Xiao Yaping Lin, Ting Zhou Siwang Zhou, Secure Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing, 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks 2014.

[7] Zhiyong Xu, Wansheng Kang, Ruixuan Li, KinChoongYow, and Cheng-Zhong Xu,Efficient Multi-Keyword Ranked Query on Encrypted Data in the Cloud, IEEE 18th International Conference on Parallel and Distributed Systems 2012

[8] Qin Liu, Chiu C. Tan, Jie Wu, and Guojun Wang, Hunan Province, P. R. China, Efficient Information Retrieval for Ranked Queries in Cost-Effective Cloud Environments, The 31st Annual IEEE International Conference on Computer Communication 2012.