

ORIGINAL ARTICLE

Dynamic Multi-Keyword Search for Secure Data Storage in the Cloud using Cuckoo Filter

Gadu Srinivasa Rao ^{#1}, Tanishq Ravi ^{#2}, Gogineni Venkata Sai Preetham ^{#3}, Jacinth Attada ^{#4}, Bojja Pranitha ^{#5}, Srikanth Reddy Baikadi ^{#6}, Sai Joshitha Chitikala ^{#7}, Vignesh Rayani ^{#8}

^{#1} Assistant Professor, Department of Computer Science and Engineering, Gitam Institute of Technology, Visakhapatnam-530045.

^{#2, 3, 4, 5, 6, 7, 8} Students, Department of Computer Science and Engineering, Gitam Institute of Technology, Visakhapatnam-530045.

Received: 10 March 2021; Accepted: 15 May 2021; Published: 20 July 2021

Abstract

In current days cloud domain gained a tremendous increase of user's attention by several small and large scale companies including Software, BPO, Medical, Schools and a lot more. Since there was very less security in the primitive clouds for storing and accessing the information from the remote locations, still there was a lot of demand for the data which is to be stored in the cloud. As we know that in primitive clouds, there are no concepts like privacy for the data in terms of encryption and message digest in order to provide data authorization. In current days cloud servers are almost dishonest in nature by omitting intentionally some qualified results to save computational resources and communication overhead. In this paper, we try to design a new protocol and analyze the importance of Cuckoo filter over encrypted cloud data to provide data search accurately by the cloud users. Here we try to encrypt the data using AES algorithm and try to generate MAC key for the data which is uploaded and downloaded by the cloud users. For data authorization we use MD5 algorithm and with the help of MD5 algorithm, we can generate a short signature to find out the data verification and finally we try to apply Cuckoo filter for data integrity. This cuckoo filter will try to store the data in duplicated manner and try to provide end user with original data.

Keywords:

Cloud Computing, Computational Resources, Communication Overhead, Data Owner, Data User, Message Authentication Code, Data Integrity, Cuckoo Filter.

1. Introduction

Cloud Computing is one of the fascinating technology which will help in reducing maintenance and development cost, in contrast to produce a very high performance services. This has become an emerging trend and is also becoming popular in very less time due to its different services which are deployed inside. All the information or applications are stored into the centralized storage place within the boundaries and can be accessed at any time is known as Data Centers[1]. The cloud data centers are mainly having a facility to execute the tasks in very fast manner with very low overhead and less time complexity. The clients data will be received and stored in separate locations rather than on the client machines and can be accessed from anywhere by using fast internet. The data which is outsourced to centralized cloud servers may be influenced by various security and privacy issues by the un-authorized users who try to gain illegal access. Hence the data privacy plays a main role for designing the protocol to store and access the information in a secure manner.

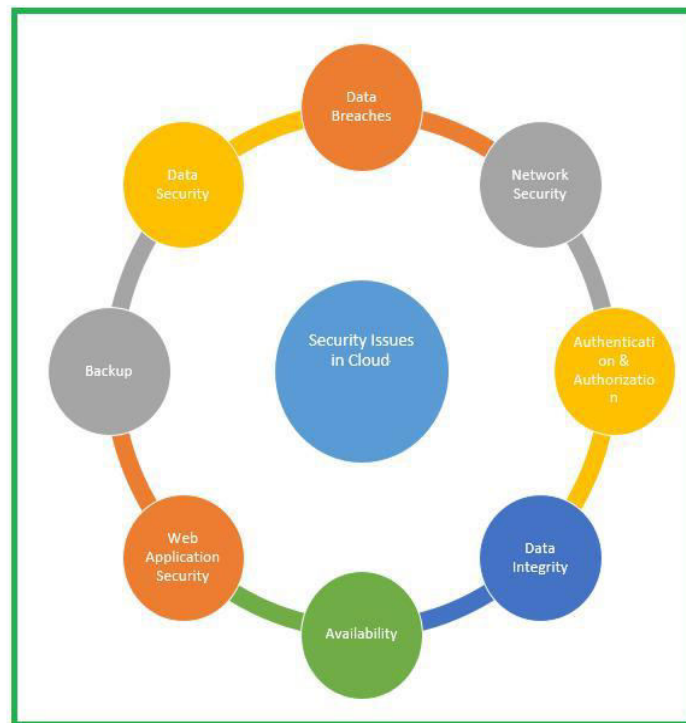


Figure. 1. Represents the Security Issues Influenced in Cloud Computing

From the above figure 1, we can clearly see there are several security issues which are influencing the cloud server. One among the several issues is data integrity, as the data which is stored inside the cloud server is not stored in the encrypted manner the data integrity is not achieved by the cloud server[2]. The clients try to outsource the valuable information directly into the centralized server and there are no facilities like encryption or MAC, hence the data can be clearly viewed or modified by the end users who try to gain illegal access on the sensitive

data. In general there are various cloud service providers present in the real world to store and access the data from remote locations and to reduce the maintenance overhead of physical server.



Figure. 2. Represents the Best Cloud Service Providers in the Real World Environment

There are several types of cloud service providers for striking business and storing the sensitive information from remote locations and try to share the centralized data to remote users. Some of the CSPs are Amazon, Google, Microsoft, IBM, Kamatera and Adobe Creative Cloud which is clearly shown in figure 2. In general the cloud data users might not be able to identify the true machines which are best to host the information in remote location, while they are enjoying the advantages which are brought by this novel cloud technology by ignoring the importance of data loss which is occurred in the cloud server [3]. Almost each and every cloud organizations try to concentrate more on the storage space to store the useful information on their individual memory locations [4] rather than concentrating on the data integrity factors. Hence this motivated me to design the current application in which we mainly concentrate on both storage as well as on data integrity.

2. LITERATURE SURVEY

Literature survey is that the most vital step in software development process. Before developing the new application or model, it's necessary to work out the time factor, economy and company strength. Once all these factors are confirmed and got an approval then we can start building the application. The literature survey is one which is mainly deal with all the previous work which is done by several users and what are the advantages and limitations in those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

MOTIVATION

A Well-known authors. Ren, C. Wang [6] has discussed a paper on "Security challenges of public cloud". In this proposal paper the authors first discussed on the cloud computing and its features. They discussed about the privacy issues which arise while using public cloud server. They investigated more on public cloud servers and gave an outline about several critical securities challenges and try to motivate further investigation of security solutions for a trustworthy public cloud environment.

S. Kamara and K. Lauter [5], has written a paper on "Cryptographic cloud storage". In this paper the authors initially discussed about the factors which effect on the problem of building and establishing a secure cloud server among the several cloud service infrastructures. First they try to identify the security preferences which are more pertained by the cloud users and also they discussed about the recent cryptographic primitives. These two authors survey a lot about all the cryptography techniques which were used for providing security for the data encryption and decryption and then came to a conclusion about accuracy and performance of each and every individual cryptography algorithm. From this proposed application we try to gain more knowledge about the working of cryptography algorithms and which algorithm gives best accuracy and efficiency compared with other algorithms.

A well-known author D. Wagner [7] has written a paper on "Practical techniques for searches on encrypted data". In this proposed article the author discussed mainly on the security and privacy risks present on the cloud server. They concentrated more on the cloud server efficiency for storing and accessing the information. They concluded that one need to sacrifice functionality of their server usage for cloud security. For instance, a user who wishes to retrieve the documents from cloud server needs to remember the search keyword, and also the circumstances about that file. Here we try to define the cryptography function for enabling the cloud data security and also we try to provide a proof of security to achieve the challenges faced by the cloud computing.

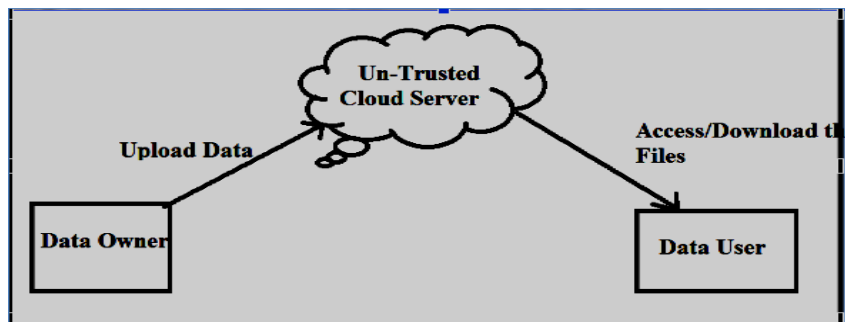
3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing cloud servers, there is no concept of multi keyword search for extracting the data from the cloud server. Almost all existing clouds used to store and access the data in plain text manner, hence privacy of the data is not completely preserved in the primitive clouds. Also there was no facility like message digest function to identify the integrity of data. The current cloud storage is almost centralized and all the data which is stored along with details of data owners and data users is clearly visible by the cloud server department, which is almost a big problem in the current cloud service. In the current cloud, if any data is modified by the hacker or intruder inside the cloud storage area, the integrity of data is not recovered by the end users, hence this plays a main contribution to design the proposed system in which the integrity of data plays a main role.

LIMITATIONS OF THE EXISTING SYSTEM

1. All the existing cloud servers are limited up to single keyword search.
2. All the data is stored and accessed in plain text manner, rather than in a encrypted manner.
3. The existing cloud servers are almost operated in a centralized manner, where all the access can be viewed and monitored by the cloud service providers.
4. There is no facility to achieve the data integrity for the sensitive data.
5. There is no proper verification mechanism to achieve the data accuracy and to retrieve the data in efficient manner.

EXISTING CLOUD ARCHITECTURE

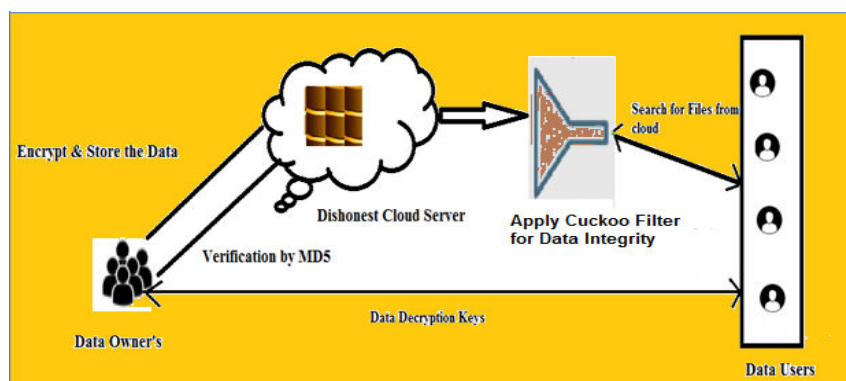


4. PROPOSED SYSTEM AND ITS ADVANTAGES

In this proposed work, we extend our work to make it more applicable in the cloud environment and more secure to against dishonest cloud server. The main contributions of this paper are

1. We designed a new filter like cuckoo filter for downloading the data in a secure manner from the untrusted cloud server.
2. We propose a short signature technique based on certificate less public-key cryptography to guarantee the authenticity of the verification objects themselves.
3. Here we used message digest algorithm MD5 in order to maintain data integrity for the uploaded cloud data.
4. Here the data integrity plays a main role and all the sensitive data achieved this principle.
5. Also we try to store sensitive data in encrypted manner so that un-authorized users cannot able to access the file in plain text manner.

5. PROPOSED ARCHITECTURE DIAGRAM



In this proposed application we try to construct a network with a set of multiple data owners, multiple data users with dishonest cloud server. The data owners and data users are connected to the centralized dishonest cloud server for storing and accessing their valuable information. Here the term dishonest means even though the data is stored in form of encrypted manner, still the CSP try to view and edit the sensitive information illegally by omitting the useful content which is present in the cloud servers. Here the cuckoo filter is applied to the cloud sever in which the data which is uploaded by the data owner will be initially encrypted [8] and that will be uploaded to the cloud server. Now the conformation of uploaded documents will be send to the sender and if any data user who wishes to access any file from the cloud server, the data user need to request the cloud server with its filename[9]. If the file is found with that corresponding filename, the data user needs to request the owner for getting the decryption key. If the data owner approves the key request then only user can open the file in plain text manner or else the file cannot be opened in plain text manner. During the process of file upload, the owner need to encrypt the file by using AES cryptography encryption algorithm and then he need to generate a message authentication code (MAC) on that encrypted data[10]. This MAC key is used for verification purpose in order to check whether the data is maintaining its integrity or not.

Notations of Proposed Model

Let $D = (D_1, D_2, \dots, D_n)$ be a set of documents which are in plain text manner.

Assume $K = (k_1, k_2, \dots, k_m)$ be the set of keywords which are collected from the documents for searching the documents in D ,

Where $\forall i \in [1, m] k_i \in \{0, 1\}^*$.

Let us assume that $C = \{C_1, C_2, \dots, C_n\}$ be the encrypted documents which are encrypted using AES algorithm for storing and accessing the file in an encrypted document [9].

I_i is a searchable index associated with the corresponding encrypted document C_i .

If A is an algorithm then $a \leftarrow A(\dots)$ represents the result of applying the algorithm A to given arguments.

Let R be an operational ring, we write vectors in bold, e.g. $v \in R$.

The notation $v[i]$ refers to the i -th coefficient of v .

We denote the dot product of $u, v \in R$ as $u \otimes v = \sum_{i=1}^P u[i] \cdot v[i] \in R$.

We use $|x|$ to indicate rounding x to the nearest integer, and $\lfloor x \rfloor, \lceil x \rceil$ (for $x > 0$) to indicate rounding down or up[10].

6. CUCKOO FILTER FUNCTIONALITY

In this section we try to discuss about the cuckoo filter when compared with other several other filters in real world cloud. We compare the general bloom filters with some advanced filter such as cuckoo and then find out the performance of each and every filter while storing the data inside the table.

CUCKOO HASHING BASICS

A basic cuckoo hash table mainly contains set of buckets where each and every bucket contains two keys generated by hash functions $h_1(x)$ and $h_2(x)$. In the lookup procedure we try to check if the data is contained in any of these buckets. For example if we try to insert a new item x in to a hash table of 8 buckets, this x value can be placed where x can be placed in either buckets 2 or 6. In one case it will see if either of two buckets are empty, then the algorithm try to insert x value to the available free bucket and then insertion procedure will be completed. If there is no space in the list, then the list will chose randomly one of the candidate buckets and re-inserts the victim item to the location, which is clearly shown in figure 3.

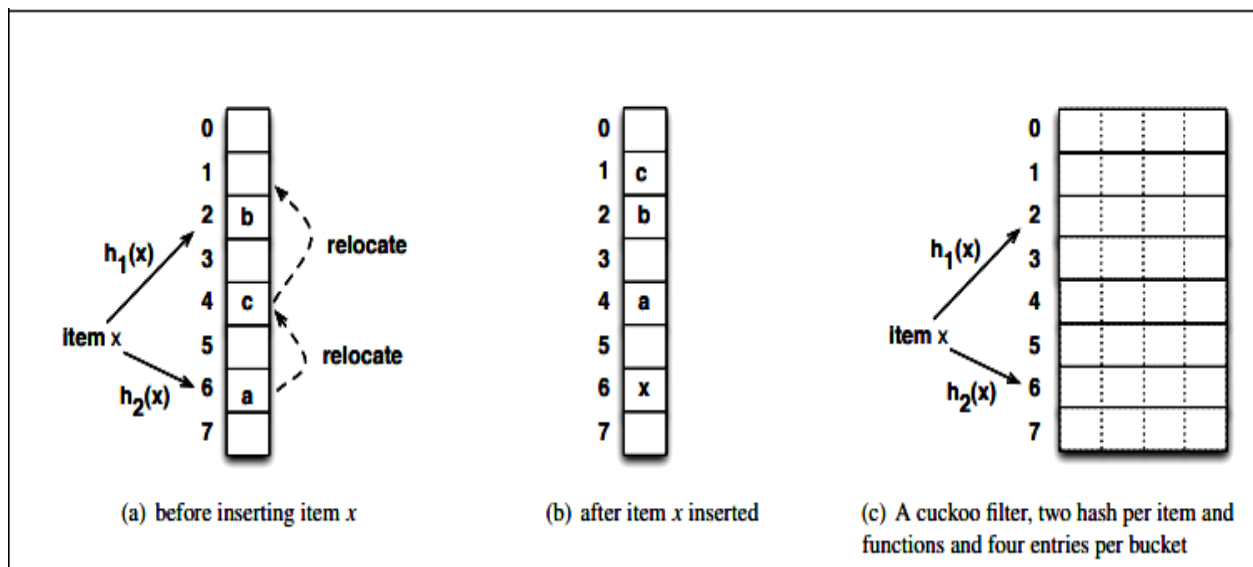


Figure. 3. Represents the Functionality of Cuckoo Filter

In general the cuckoo hashing technique has high space occupancy compared with all primitive filters because this is mainly used to refine the items which are placed in the process of insertion. The cuckoo filter is one kind of hash technique which improved earlier item-placement decisions when inserting new items. By conducting several experiments on all the primitive filters along with cuckoo filter, we finally got a conclusion that cuckoo gives filling of 95 % data with high accuracy and probability compared with all other primitive filters.

TABLE : Represent the Comparison of Cuckoo Filter with Several Other Filters

filter type	space cost	cache misses per lookup	deletion support
Bloom	1	k	no
blocked Bloom	1x	1	no
counting Bloom	3x ~ 4x	k	yes
d -left counting Bloom	1.5x ~ 2x	d	yes
quotient	1x ~ 1.2x	≥ 1	yes
cuckoo	$\leq 1x$	2	yes

7. EXPERIMENTAL RESULTS

Implementation is a stage where the theoretical design is converted into programmatically manner. In this proposed application we try to use JAVA as programming language in which HTML, JSP is used as front end technologies and as a back end we try to use MY-SQL.

1) HOME PAGE

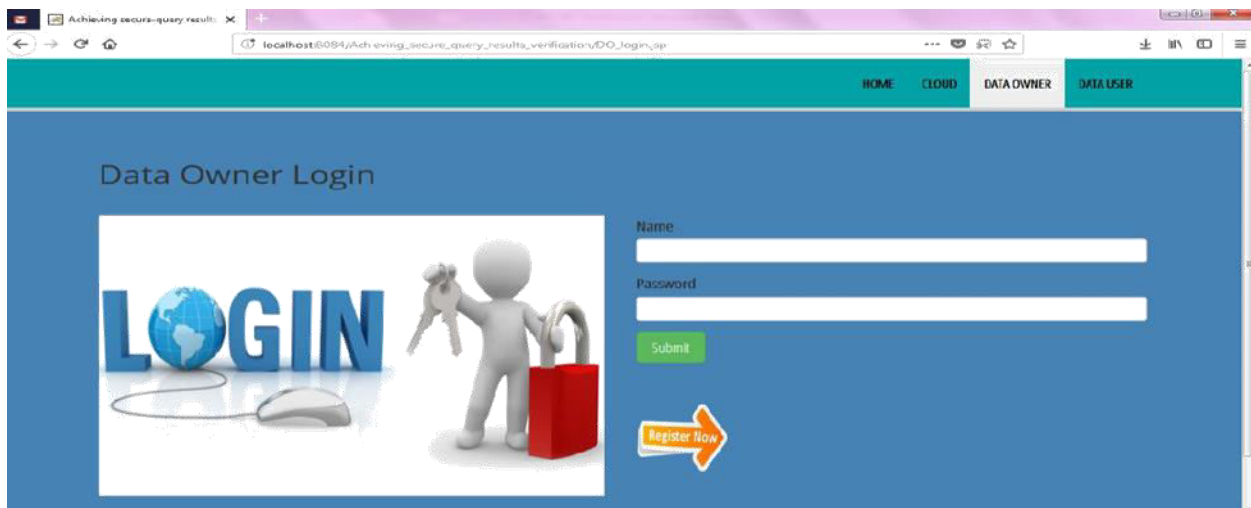


Figure . Represents the Home Page with Cloud, Data Owner & Data User

2) DATA OWNER UPLOAD A FILE

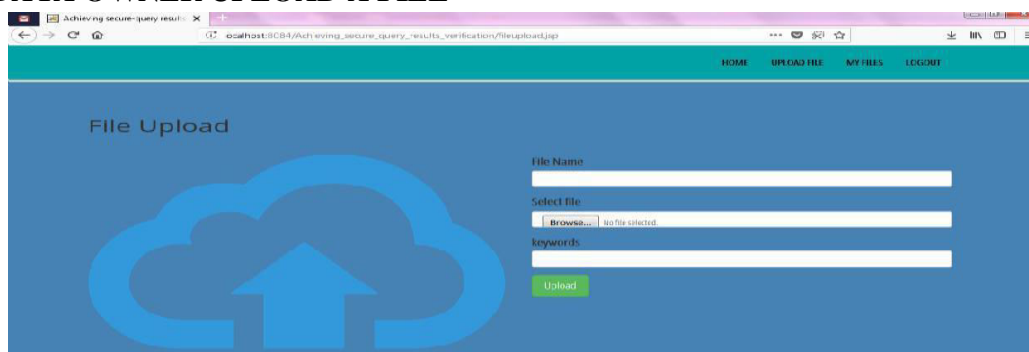
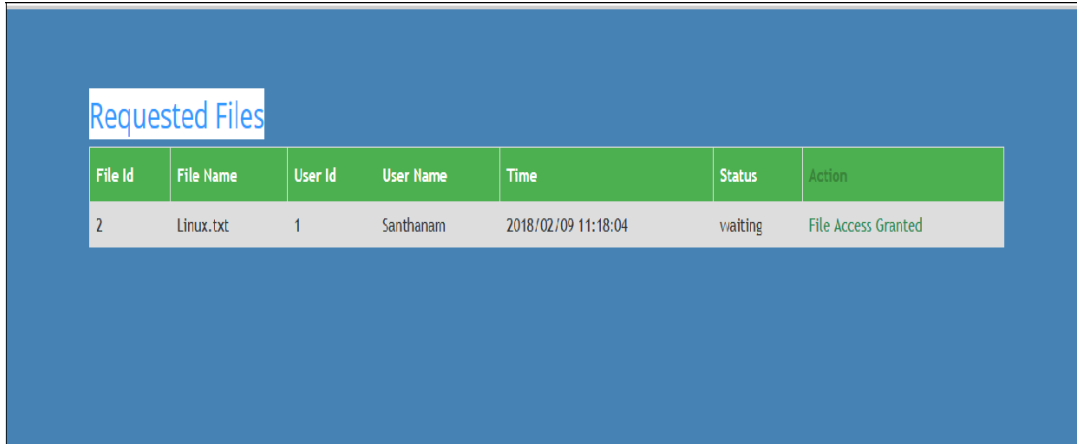


Figure . Represents the Data Owner try to upload sensitive document

From the above figure we can clearly see that data owner try to upload a valid file as input and try to enter filename, browse the file and enter the search keywords. Here the file will be encrypted and then upload into the cloud server.

3) DATA USER SEARCH FOR FILE



File Id	File Name	User Id	User Name	Time	Status	Action
2	Linux.txt	1	Santhanam	2018/02/09 11:18:04	waiting	File Access Granted

From the above figure ,we can clearly see that data user try to search for that sensitive document by substituting the filename or search keywords.Once if the file is found he can see the file displayed in encrypted manner with owner request.Once if the data owner wish to give decryption key for the end user he will grant the key permission.

4) DATA USER RECEIVES DECRYPTION KEYS TO HIS REGISTERED MAIL ID



Figure. Represents the Data User received the Decryption Keys

From the above figure ,we can clearly see that data user received the downloading keys from the owner and now he need to substitute these keys while downloading the data from the cloud server.

5) DATA USER TRY TO SUBSTITUTE THE KEYS AND VERIFY



Figure. Represents the Data User Substituted the Keys

From the above figure, we can clearly see that data user substituted the keys and he can see the two verification objects are same which means the data is not edited or altered by the cloud server. Hence the cuckoo filter will try to generate original information by decrypting the data .

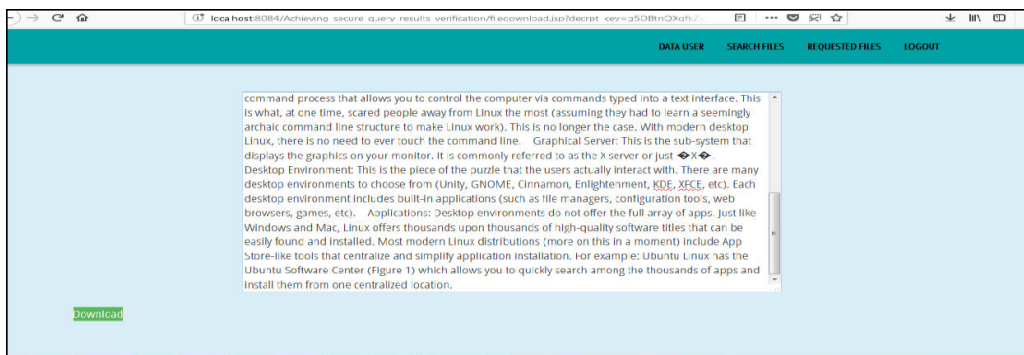


Figure . Represents the File in Plain Text Manner

From the above figure we can see the file is retrieved in plain text manner and can able to download the file.

6) DATA USER CAN VIEW THE DIFFERENCE IN KEYS IF THE CONTENT IS MODIFIED

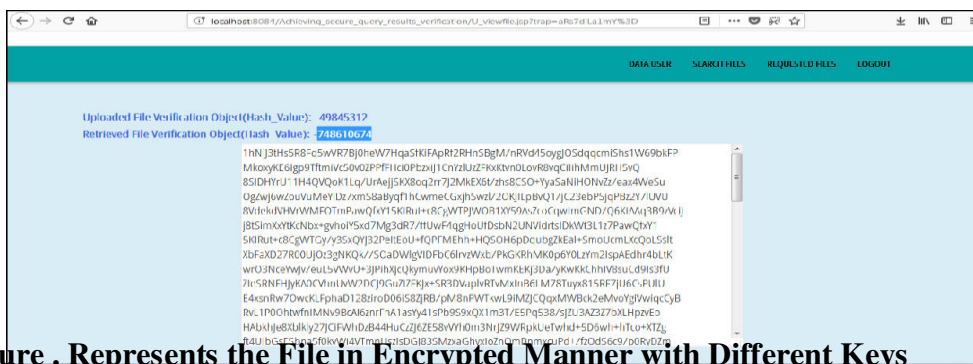


Figure . Represents the File in Encrypted Manner with Different Keys

From the above figure we can see the file is retrieved with two distinct key objects. This clearly represent the data is modified by the cloud server department and hence the MAC keys are altered, the cuckoo filter will be invoked and retrieve the original information without disturbing the original content.

7) DATA USER RETRIEVE THE ORIGINAL CONTENT BY USING CUCKOO FILTER

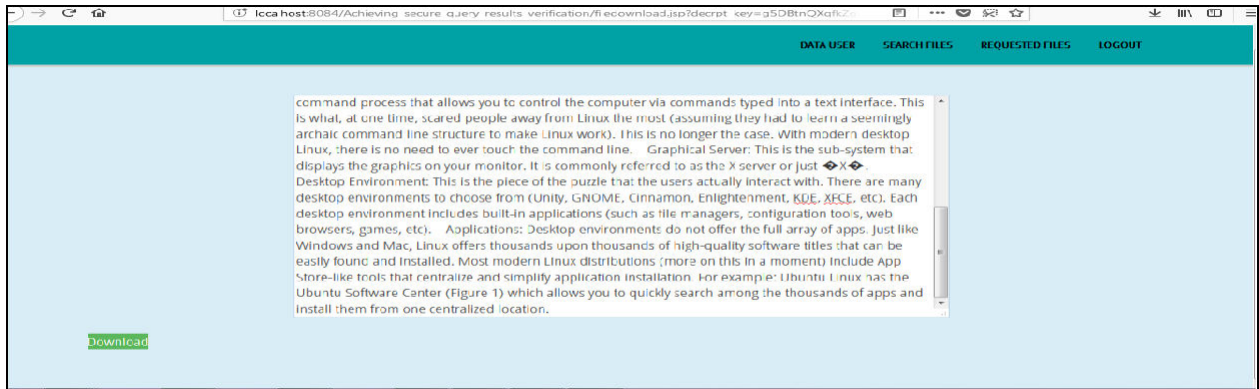


Figure . Represents the File in Plain Text Manner

From the above figure we can see the file is retrieved in plain text manner and can able to download the file.

8. CONCLUSION

In this current work we for the first time designed and implemented a secure protocol for storing and retrieving the data in an efficient manner by maintaining data integrity using cuckoo filter. Here we attempt to use MD5 algorithm for generating MAC keys for that sensitive documents then attempt to store them inside the dishonest cloud server. In this proposed application we launched a new filter like Cuckoo filter in which the filter will give data integrity for the data users who attempt to download the sensitive data from the cloud server. By conducting various experiments on our proposed model, we finally came to a conclusion that our proposed approach is best in providing data integrity and security over sensitive data under the encrypted cloud data.

9. REFERENCES

- [1] Two Well-known authors K. Kurosawa and Y. Ohtaki, has written a paper on “Unsecure searchable symmetric encryption,” published in Lecture Notes in Computer Science, vol. 7397, pp. 258–274, 2012.
- [2] Two Well-known authors P. Xu and W. Wang, has written a paper on “Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack,” published in IEEE Transactions on Computers, vol. 62, no. 11, pp. 2266–2277, 2013.
- [3] Two Well-known authors P. Mell and T. Grance, has written a paper on “The nist definition of cloud computing,” published in <http://dx.doi.org/10.602/NIST.SP.800-145>.
- [4] Two Well-known authors D. Song and A. Perrig, has written a paper on “Practical techniques for searches on encrypted data,” published in IEEE Symposium on Security and Privacy, vol. 8, 2000, pp. 44–55.
- [5] A Well-known author E.-J.Goh, has written a paper on “Secure indexes,” published in IACR ePrint Cryptography Archive, <http://eprint.iacr.org/2003/216>, Tech. Rep., 2003.
- [6] Two Well-known authors D. Boneh, and G. Persiano, has written a paper on “Public-key encryption with keyword search,” published in EUROCRYPT, 2004, pp. 506–522.

- [7] Two Well-known authors R. Curtmola, and R. Ostrovsky, has written a paper on “Searchable symmetric encryption: improved definitions and efficient constructions,” published in ACM CCS, vol. 19, 2006, pp. 79–88.
- [8] Two Well-known authors M. Bellare, and A. O’Neill, has written a paper on “Deterministic and efficiently searchable encryption,” published in Springer CRYPTO, 2007.
- [9] Two Well-known authors K. Ren, C. Wang, and Q. Wang, has written a paper on “Security challenges for the public cloud,” published in IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.
- [10] Two Well-known authors S. Kamara and K. Lauter, has written a paper on “Cryptographic cloud storage,” published in Springer RLCPS, January 2010.