TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4164-4167

Cloud Sanitization and Auditing

[1]Sai Sharan P, [2]Ridaa Rauf, [3]Nandini D, [4]Ritesh V R, [5]Gopinath R

[1,2,3]School of C & IT, REVA University, Bengaluru, India
[4]B.Tech Student, School of C & IT, REVA University, Bengaluru, India
[5]Assistant Professor, School of C & IT, REVA University, Bengaluru, India
[1]tonyhawk@yahoo.in, [2]ridaarauf@gmail.com, [3]nandinidevraj23@gmail.com, [4]riteshvr20@gmail.com, [5]gopinath.r@reva.edu.in

Article Info
Volume 83
Page Number: 4164-4167
Publication Issue:
May-June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

Storage services provided by the cloud gives users the ability to reserve their data in an efficient manner in the storage provider/cloud. In addition to this, by the use of the proposed integrity auditing scheme, the integrity of the files sent by the client to the server can be checked. In some basic cloud storage structures, like student records or office records, some vulnerable data may be present in the cloud file. A third party or other users cannot access this vulnerable data. As a result, if the entire data is encoded, the delicate data would be concealed but it would be rendered useless to other users. In order to address the above issue, we offer a strategy that acknowledges information imparting to cover up vulnerable data. At the moment, a sanitizer is employed to sanitize the given data blocks associated with the vulnerable information and this sanitizer converts the signatures of the physical records into valuable ones. Hence the idea proposed in this paper would give users the ability to store and utilize the stored documents, and they can easily rely on the fact that their vulnerable data is hidden, assuring that the data integrity is sustained. The assessment of the suggested idea is shown to be functional.

Keywords**:** *Auditing, Cryptography, Decryption, Encryption, Sanitization.*

1. Introduction

With the globally expanding growth in the magnitude of data in the everyday world, it arises challenges. Storage and security are the few main concerns to address. People find it convenient to store their data in an online cloud platform rather than other storage means due to its obvious advantages such as easy accessibility and more. Cloud platforms are even able to provide security for their users but the features we want to bring the area of focus onto is hiding of confidential (or "sensitive") information and means for checking its integrity. Building a cloud storage system that is able to achieve these features is the goal. Such a system would prove useful to many organizations, health and education institutions, companies, business corps and more that utilizes cloud storage systems. Such organizations require a system that enables them to hide specific information in the entire file pertaining to a user that would be confidential to that user in the organization. Encryption strategies might also help in hiding these specific information but that also means to encrypt the file as a whole rather than the proposed strategy. It is both economically less suitable and conceptually different strategies that are appreciatively used for different scenarios or situations. Storing data in the cloud is pretty easy and simple but, so are the chances of it getting altered or changed due to reasons such as unauthorized user access, upload failures, or manual errors. Organizations require their data to be correct and a system that could check for the same. Online and offline auditing strategies are adopted to allow these checks in the system. The word "Sanitization" here refers to a strategy which is used to achieve data hiding i.e, hiding certain words that are matters of confidence. It makes use of a "sanitizer" which converts a file in normal form into one that's in a "sanitized" form. Keys are generated for each users which makes it a user based and user specific system. This system can be accessed from anywhere and is more viable which makes it commendable.

2. Literature Review

[1]This approach is mainly based on proving the data integrity stored at unreliable servers. The client sends the data to the server for storage. These files may contain some meaningful information which the server may use or modify. This technique prevents that from happening by maintaining the integrity and confidentiality of data. [3]

May-June 2020
ISSN: 0193-4120 Page No. 4164-4167

Security of data is maintained in this technique but the entire data sent by the client needs to be encrypted whereas in our work, along with maintaining integrity of the data, only specific or sensitive data blocks are encrypted which requires less storage. We've taken the idea of data integrity from this technique.

[2] In this mechanism, the client stores his files on a storage provider or a cloud so that the auditing protocol is efficiently executed in order to spot check that the client's data is present with the server when needed in the future.[2]

Proofs of retrievability provides light-weight storing and proving but verification time is longer because the entire block of data needs to be encrypted and must be maintained by the server while our approach provides third party auditing along with data sharing with sensitive information hiding.

3. Proposed Work

As shown in figure 1.1, we see a method of achieving data protection by hiding sensitive information while ensuring accurate integrity checking of the data. In this model, the file uploaded will be kept protected all the while giving control access to check for the file's integrity easily. The outcome of this model exhibits the model's true potential once implemented and placed in a system.

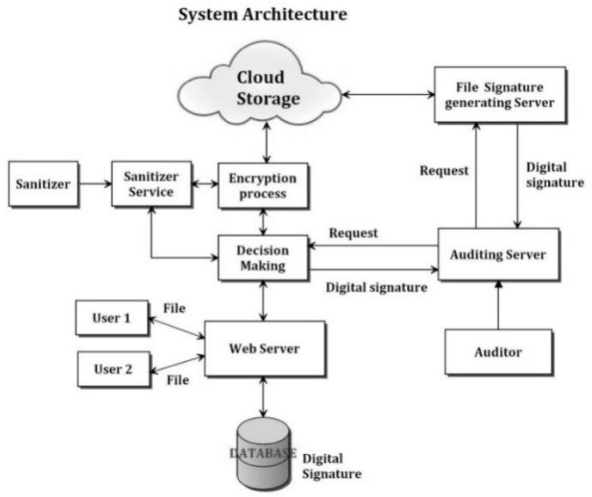


Figure 1.1: System Architecture

The below sheds light into the new and unique aspects of the proposed system, namely:

- Online and Offline Auditing.
- Pre-tag Sanitization.
- Concept of De-sanitization.

The system helps us do audit checks not just by online means using signatures as proposed in the previous system but also offline by the use of self-auditing techniques. The sanitized file after storing in the cloud, has to be made available back to its original form which wasn't something that was acknowledged in the previous system for whenever the user whom the file belongs to wants to download, in order to make use of the file and for this system to be more acceptable commercially

A. Modules:

1) Admin Module

The Admin is the one with authority to this system's base configurations. On landing onto the login page, the admin will be able to carry out the following:

a) Can login and out of his or her profile.

b) Can view and edit his or her own profile (with name and other details).

c) Can view, edit and delete users from the list of users in the system.

d) Can view, edit and delete auditors from the list of auditors in the system.

e) Can change password.

f) Can view the transaction logs of every entities (users and auditors) logging in the system.

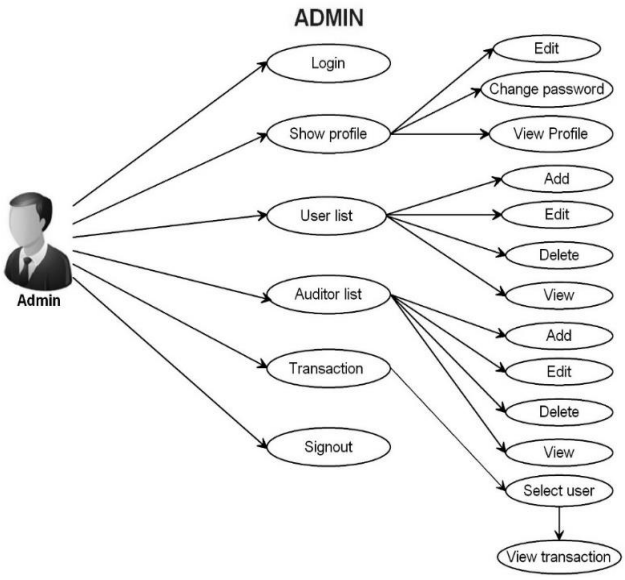


Figure 2.1: Use case diagram for Admin

2) User module

The User is the one who uses the system for storing and accessing the file from the cloud. On landing onto the login page, the user will be able to carry out the following:

a) Can login and out of his or her profile.

b) Can register self as a user of the system by filling out the specifics (email id, name, id and more).

c) Can view and edit his or her own profile (with name and other details).

d) Can upload file:

May-June 2020
ISSN: 0193-4120 Page No. 4164-4167

- Select the file and click upload.
- Pre-processing steps include removal of special symbols and words of less meaning (like the, it, was and so on).
- Compare the words with the sensitive words dataset of sensitive words.
- If matching, generate the signatures.
- DNA algorithm.
- Store signatures in database of the respective user.
- Transaction logged, successful.

e) Can download file:

- Select file.
- Download.
- Transaction logged, successful.

f) Can send request for file integrity verification.

g) Can change password.

h) Can view his or her own transaction logs.

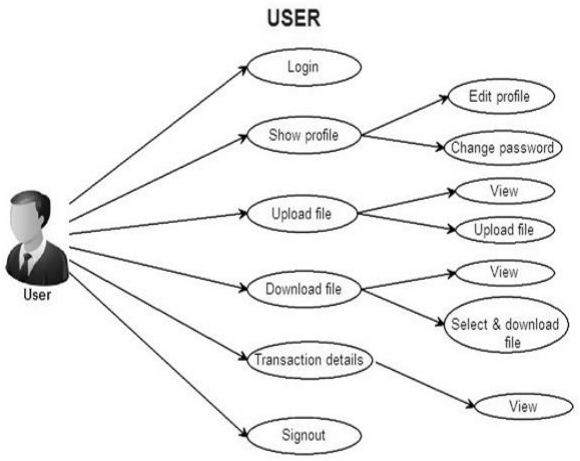


Figure 2.2: Use case diagram for User

3) Auditor module

a) The Auditor is the one who handles online auditing of files in the cloud storage. On landing onto the login page, the auditor will be able to carry out the following:

b) Can login and out of his or her profile.

c) Can change password.

d) Can view and edit his or her own profile (with name and other details).

e) Can view the list of users that applied for file verification.

f) Can audit the file:

- Select file.
- Click verify.
- Get the key from server.
- Generate new key.

g) Compare keys and display results.

Auditor

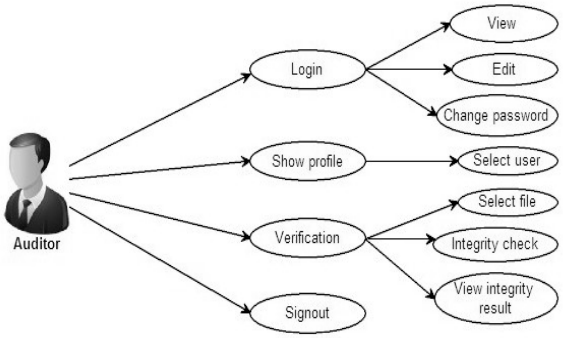


Figure 2.3: Use case diagram for Auditor

4. Appendices

A. Algorithm 1: DNA

1) Algorithm 1.1: Algorithm for Encryption

Step 1: Convert binary data to DNA sequences:

A=00, T=01, C=10, and G=11.

Step 2: Use Complementary pair rule.

Step 3: Representing DNA sequences as numeric data.

Example:

Assume the reference sequence to be:

CT1GA2TC3CC4GC5AT6TT7.

Then the numerical representation will be 040602.

In this implementation, the admin assigns a unique DNA sequence and a unique key to every user which is later used to create the DNA reference sequence for the user using rand() function and hash set constructs.

2) Algorithm 1.2: Algorithm for Decryption

Step 1: Conversion to DNA sequences from numeric data.

Step 2: Complementary pair rule.

Step 3: Convert DNA sequences to binary data.

Example:

DNA Reference Sequence:

(TG, TA, AT, GC, CT, GA, CA, AC, AA, GT, CG, AG, CC, TT, TC, GG)

[TG00,TA01,AT02,GC03,CT04,GA05,CA06,AC07,AA08,GT09,CG10,AG11,CC12,TT13,TC14,GG15]

- M´´´=0706 (Input)
- By referring the DNA sequence:
  - M´´= ACCA.
- Use Complementary rule
- Use Base Pair Rule

B. Algorithm 2 : Sanitization

Step 1: Read the message as a string.

Step 2: Split the string with respect to space → " "

Step 3: Compare each word with the list of sensitive words.

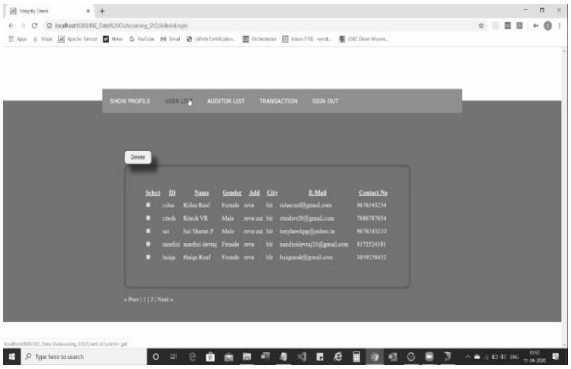
Step 4: If matched, then encrypt (DNA Algorithm.)

Step 5: Else, compare the next word.

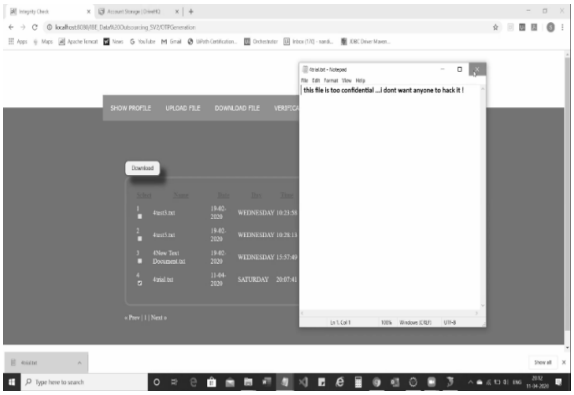
Step 6: Continue till the last word.

5. Result

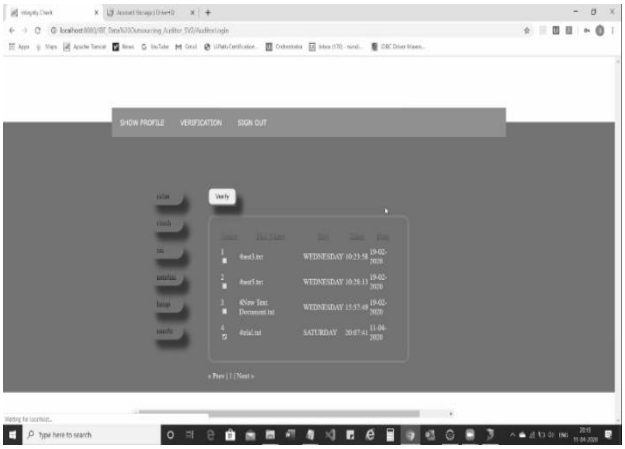
- Admin home page with all admin functionalities.



- User Home Page With File Selected To Upload/Download



- Auditor Home Page With Verification



6. Conclusion

We have proposed a unique user-based auditing scheme to ensure accurate integrity checking of the data for reliable cloud memory systems, which also realizes sensitive information hiding in the uploaded data records/file. In the idea introduced above, only if the susceptible content of the record/file is in safe hands or is secure, the file stored in the cloud can be explored or employed. Furthermore, in order to systematically accomplish the integrity checks, an on the go online and offline monitoring or auditing scheme has been proposed. The analysis of the suggested idea is shown to be functional i.e. the security is guaranteed, capability and efficacy distribution of a steady, and a lot more efficient cloud based storage and management system.

7. Future Work

A feature that can be added to this proposed idea is data sharing i.e. letting other users conveniently access our files directly over the cloud only if the vulnerable content of the file is secured. This feature can either be in the encrypted or the decrypted form, depending on the data owner, data viewer's needs and access rights, and their convenience.

References

[1] Hitesh Marwaha, Rajeshwar Singh, "The Secure Migration of Data to Cloud using Data Sanitization and MAC address based AES", ISSN: 2277-3878, Volume-7, Issue-6, March 2019.

[2] C. Guan, K. Ren, F. Zhang, F. Kerschbaum, and J. Yu, "Symmetric-key based proofs of retrievability supporting public verification," in Computer Security – ESORICS 2015. Cham: Springer International Publishing, 2015, pp. 203–223.

[3] C. Chris Erway, Alptekin Küpçü, Charalampos Papamanthou, Roberto Tamassia, "Dynamic Provable Data Possession", 2015.

[4] Wenting Shen, Jing Qin, Jia Yu, Rong Hao, Jiankun Hu, "Enabling Identity Based Integrity Auditing and Data Sharing with Sensitive Information Hiding for Secure Cloud Storage",2018.

[5] David Sánchez, Montserrat Batet," Privacy-preserving data outsourcing in the cloud via semantic data splitting", 2017.