ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Detection of Content Replication on Cloud using Hash Key Generation Method

Avinash S B[1], Dr. Shambhavi B R[2]

P.G. Student, Department of Information Science Engineering, BMS College of Engineering, Basavanagudi, Bangalore, India[1]

Associate Professor, Department of Information Science Engineering, BMS College of Engineering, Basavanagudi, Bangalore, India[2]

ABSTRACT: Dynamic Proof of storage, (DyPos) is a cryptographic technique which permits the user to check the reliability of the data which is outsourced, it also enables the user to modulate the data in block level and finally it avoids storing of exact copy of the content which is already present in storage devices. For single user environment there are several techniques are available, but when considering the multi user environment many issues need to be solved efficiently.A secured content replication check technique has to be needed in the client side to avoid duplication by stopping the upload process of the replicated file content and make the best use of storage device capacity. As of the literature survey no existing systems are doing this job.This work introduces the idea ofdeduplicatable dynamic proof of storage and propose an efficient new tool called as Homomorphic Authenticated Tree (HAT) to overcome the issues of structure diversity and private tag generation. This work is trying to prove the safety of the structure, theoretical analysis and also experimental results are efficient for checking the deduplication of a file on cloud.

KEYWORDS*:* Cloud storage, dynamic Proof of storage, deduplication.

I. INTRODUCTION

Storing of file in a storage device wheresecurity [1], reliability, easy accessibility and sharing [2], are very important criteria's everyone need to considered. Outsourcing the file is becoming very effective and attractive in many fields like IT companies, education academics due to all the advantages of reliability, security, and easy sharing. Many organizations like Amazon, Google, Microsoftare providing these storage facilities to outsource our data's, where the end users can keep their data, modify them whenever needed and can also access them at any time irrespective of the place. Data reliability is one of the most important aspect when the user outsources their data to cloud storage. Service providers must convince the customers about the reliability of the data, availability and security as well. There are many techniques for maintaining the reliability of data such as Message Authentication Code (MAC) and digital signature. Here, if the users need to check the integrity of the data and to modify any content, the user need to download the entire file which incurs heavy communication cost. These techniques are not suitable for cloud storage where the users check the integrity of the data frequently. So, this work introduces the concept calledProof of storage,for checking the reliability of the data without downloading the entire file. The users may also perform many dynamic operations such as modifying the content of data, adding extra content and removing some content from file to update their file.Dynamic Proof of storage is proposed for such operations.Along with the Proof of storage dynamic Proof of storage, also employsauthenticated structuresuch as Merkle Tree**.** Here when the dynamic operations are performed on the file content, new tags are generated for themodified content**.** Later, Generated tags are used for integrity checking.

To understand it much clear,when the users upload their file to the cloud storage the new tag is generated for content of the file, which is then used for the deduplication check as well as for integrity checking. For storing file in server, a new tool called as Homomorphic Authenticated Tree (HAT) is proposed.As shown in Figure 1, HAT divides the entire

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

content of the file into n number of blocks, each block's content then encrypted using the AES algorithm and stored as a different block in cloud storage.

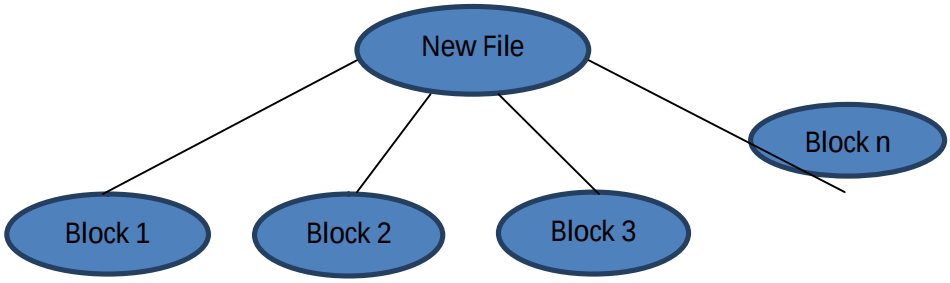

Figure 1: Initial blocks

In a single user environment once after the file stored in cloud the file is said to be secured as there is only a single user can access that file. But in the case of multi-user environment the file must be given access for trusted users for some operations.If the users need to retrieve the file they must require the generated tag, then that tag must be sent to server, depending on that tag value the server get the file from its storage and return back to the users. For modification process the users first access the file then they can do modification to the file in block levelt and stored in the server. New tag will be now used for integrity check and deduplication check.

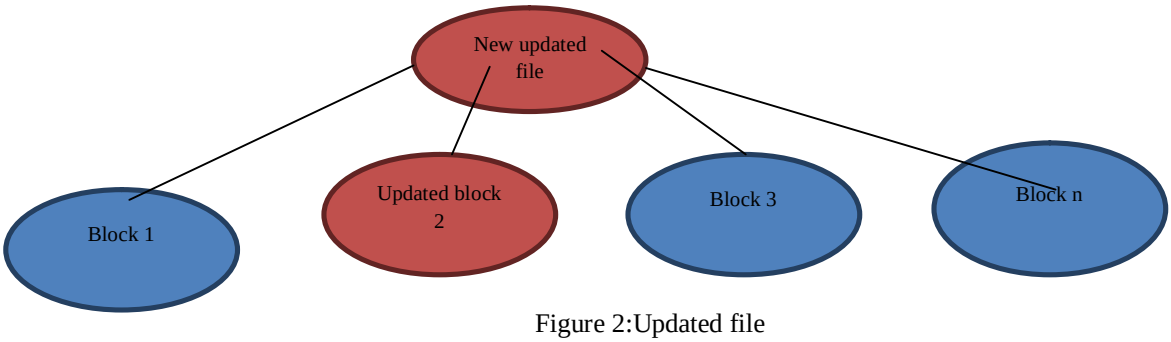

Figure 2:Updated file

As shown in Figure 2, the content of block 2 is modified; the updated content is re-encrypted using AES algorithm. Later the new encrypted block replaces the previous block and new tag value will be generated for the new file content. Later newly generated tag value will be used for integrity check and deduplication check.
This work considered two major issues of deduplication and reliability of the file which is in cloud storage. When come to the deduplication, if the user uploads the file with the same content which is present in the cloud storage then the tag generated for this file compare with the tag which is already present in server, if it matches the then the uploading process will skip and notify about the same.If the new tag generated for the file is different from all the other tags stored in server, then the user is allowed to store their file in cloud storage. Considering other issue of reliability, every user need to be guarantee that their files are safe without tampering and data lost.When an attackertries to tamper the data, notification will be provide to the user and the files are blocked until validating the file again, Once the files being tried to attack the files are not available to access and can't able to do any modification to the file until being validated again. So from the above processes both the issues of deduplication and reliability will be solved.As a summary the proposed system is much suitable for multi user cloud storage environment.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

II. RELATED WORK

Xiao, Zhifeng, and Yang Xiao [1], concentrated on the issues like security and privacy of the data or business application while outsourcing to third party. Authors have mitigated the security issues based on attribute driven approach. Integrity, secrecy, privacy-preservability, accountability and availability are the five attributes identified in this work. And this work also concentrated on the vulnerabilities that can be exploited by the attackers. Ardagna, Claudio A., et al [2], works on the interface between the cloud security, assurance and cloud security. Initially this work provide the overview of State of an art on the cloud security approaches. As cloud storage getting more popularity and are widely used by many organizations for storing the files, business applications and many things, because of its low cost, better performance, availability. Finally it provides the recommendations for development of the next generation cloud security, and assurance solution. Ateniese, Giuseppe, et al [3], introduces the model for Provable data possessions for the very first time and allows the tenant (client) to check whether the untrusted server possesses the original file without retrieving it. This work generate a probabilistic proof of possession by sampling the random sets of the blocks from the server, as a result it reduces the I/O costs. It uses Homomorphic verifiable tags for data checking on server. Hence the PDP technique for data verifying supports for large data sets in a widely distributed storage system. The disadvantage of this model is, it concentrated only on the static files.Erway, C. Chris, et al [4], considered the problem of proving the Integrity of data efficiently, which is stored in the untrusted server. In order to overcome the disadvantages of PDP, they introduces Dynamic PDP, it extends the PDP technique to support the updates to the stored data. This work introduces the new authenticated dictionary version based on the rank information.Li, Jin, et al [5], concentrated on protecting data security, they attempted to formally address the issue of authorized data deduplication. Apart from traditional techniques, the privileges of the uses are considered in deduplication check. The duplication check is accomplished by using the duplicate check tokens of files, which are generated by the private cloud server along with private keys.Yan, Zheng, et al [6], works on generating a encryption algorithm for deduplicate encrypted content/data stored in cloud and at the same time it supports the secure data access control. They finally used the Symmetric Encryption algorithm for data protection.

Jagtap, Pradeep Sarjerao, and A. D. Gujar [7], concentrated on securing the deduplication in cloud storage. Own computing devices which are using the cloud storage services for data backup will face the challenges because of the restricted system resources. In order to ensure the secure deduplication they introduced the application called as ALG-Dedupe. This application improved the efficiency of the deduplication.Leesakul, Waraporn, Paul Townend, and Jie Xu.[8], proposed the new dynamic deduplication scheme for the cloud storage, which aims to improve the efficiency of storage and will maintain the redundancy of fault tolerance. Finally this scheme balances the efficiency in changing storage and requirement of fault tolerance.

Shin, Young-joo, JunbeomHur, and Kwangjo Kim [9], proposing Proof of storage, along with Deduplication. They present the insecurity of the model under new model attack, where the malicious clients exploits the manipulated keys. Authors have mitigated the above mentioned issues based on blending client's key. Zheng, Qingji, and Shouhuai Xu [10].are trying to say that the security and efficiency, the two major issues of cloud storage are co-exist in the same framework. In the current work they took the concept of "the public verifiability which is offered by the Proof of Data Possession or Proof of retrievability techniques can be exploited naturally to achieve the Proof of Ownership. They also tried to design the concrete scheme that will secure the Random Oracle Model, based on the Diffie-Hellman assumption. Naor, Moni, and Moti Yung [11], concentrated on proving that the one way hash functions exists if ony of the 1-1 hash function exists. In this work they used the public key cryptographic technique and some randomizes algorithms to generate the hash key.

III. SYSTEM ARCHITECTURE

As shown in Figure 3, proposed system consist of two entities, Cloud Server and Client/User. For each file, original user is the user who uploaded the file to the cloud server, while subsequent user is the user who proved the ownership of the file but did not actually upload the file to the cloud server due to deduplication. This system consist of five phases, pre-process phase, Upload phase, Deduplication phase, Update phase, Proof of storage [9][10], phase.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

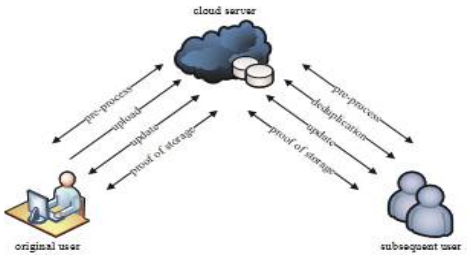

Figure 3:System Overview

In the very first phase the users intend to upload their local files to the cloud storage. The cloud server decides whether these files should be uploaded. If the upload process is granted, go into the upload phase; otherwise, go into the deduplication phase. In the second phase the files to be uploaded do not exist in the cloud server. The original users (owners) encodes the local files and upload them to the cloud server. In third phase the files to be uploaded already exist in the cloud server. The subsequent users possess the files locally and the cloud server stores the authenticated structures of the files. In fourth phase users may modify, insert, or delete some blocks of the files. Then, they update the corresponding parts of the encoded files and the authenticated structures in the cloud server. The users should upload the files in the upload phase or pass the verification in the deduplication phase, if he fails in the first phase he can't upload as well as update the files. In the last phase users only possess a small constant size metadata locally and they want to check whether the files are faithfully stored in the cloud server without downloading them.

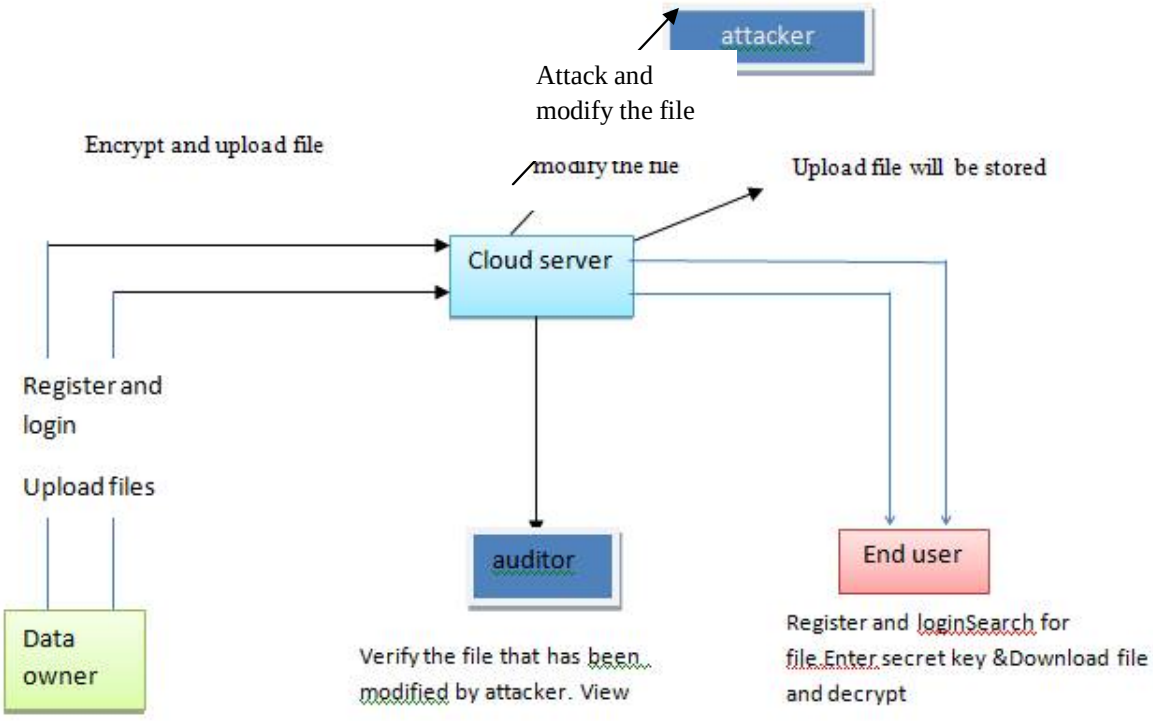

Figure 4:System Architecture

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

As shown in Figure 4, the system architecture of the proposed systemconsistsof five entities, data owner, cloud server, attacker, auditor and end user. Data owner is the one who owns an account in cloud storage and can upload the file to the server. And has the permission to modify the data when needed. Clouds server will maintain the records in cloud, it is the one which allow the file to upload or reject. Auditor plays an important role in auditing the file when it is being attacked. End user is also can access the file by downloading it but does not have the permission to upload or to modify the data. End user need the secret key (tag) value to download the file which is generated when the file is uploaded to cloud server (cloud storage).

IV. METHODOLOGY

Homomorphic Authenticated Tree (HAT):
To implement an efficient deduplicatable dynamic PoS scheme, proposed system design a new novel authenticated structure called homomorphic authenticated tree (HAT). A HAT is a binary tree in which each leaf node corresponds to a data block. Though HAT does not have any limitation on the number of data blocks, for the sake of description simplicity, assume that the number of data blocks n is equal to the number of leaf nodes in a full binary tree. When the data owner uploads the file to the cloud storage the entire file is divided into n blocks. Before storing those blocks in cloud server each block content is encrypted using a AES algorithm for security, purpose. Whenever the data owner need to update the file they can update the file in block level so only the updated block will re-encrypted and replace the previous block in server. So here there is no need of encrypting the whole file content again and again. This is one of the major advantage of the HAT.

Advanced Encryption Standard Algorithm:
AES algorithm is based on substitution and permutation design. This algorithm has a fixed block size of 128bits and a key size of 128,192 or 256 bits. The key size indicated the number of repetitions of transformation rounds that will convert the plaintext into ciphertext.

Construction of DyPoS:
The proposed system introduces a concrete scheme of deduplicatable dynamic PoS called DyPoS. It consists of five algorithms Initialization, Upload, Deduplication check, Update, and Proof of storage,Check.
Initialization Phase: In this phase the data owner who intended to upload the file to cloud storage ready with the file to upload, have idea about the cloud storage (server) to which they intended to upload and be aware about the reliability, anytime access, security and many other criteria's.
Upload Phase:In this phase data owner upload the file to the server (cloud storage)**.** Before saving the file to the server a tag be generated for the file content using the hash function. For every different content different tag will be generated. And the file is split into n number of blocks, each block content is encoded using AES algorithm for security purpose, later on the these the file is stored as a different blocks of encrypted data along with the unique tag value generated.
Deduplication check:This is the main phase of the proposed system where the redundant content will be checked.As mentioned in the upload phase tag value is generated for every unique file content, when the subsequent user tries to upload the file,tag will generated for that file. If it matches the tag which is already present in the server, then it skips the upload process of the new file. And notify about the same to the subsequent user who tries to upload the file.
Update Phase:Once after successfully passes the deduplication phase the file stored in the server. If the authorized owner wants to do some modifications, they will access the file from server using the tag key generated and modify the file content in block level. It reduces the communication cost as because there is no need to encrypt the entire file content. Only the modified block will re-encrypted and replace the previous block stored in the server.
Proof of storage, Phase:This is the phase where the data owners or users verify the integrity of the data. Users can check the Integrity of data any number of times.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

V. RESULTS

Case 1: When the users trying to upload the file with same content which is already present in the server, the server prevent or skipthe upload process of that file and notifying about the existence of the same content in the server.Figure 5, shows the popup window alert when the server finds the redundant copy of file while uploading new file.

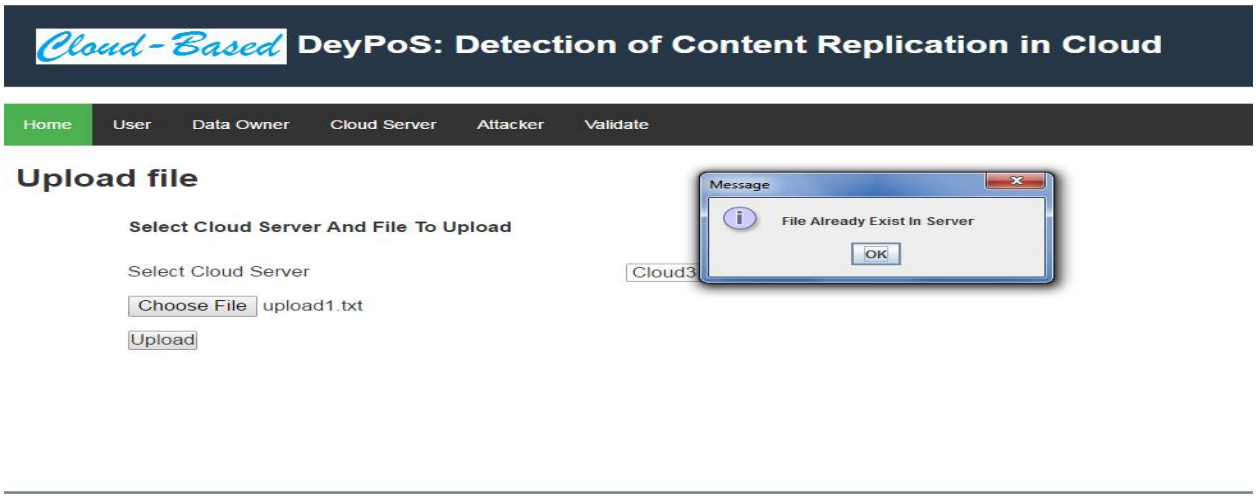

Figure 5: deduplication

Case 2: When the file is being attacked by the intruder, the proposed system changes the status of the file in the server and blocks all the access to that particular file.

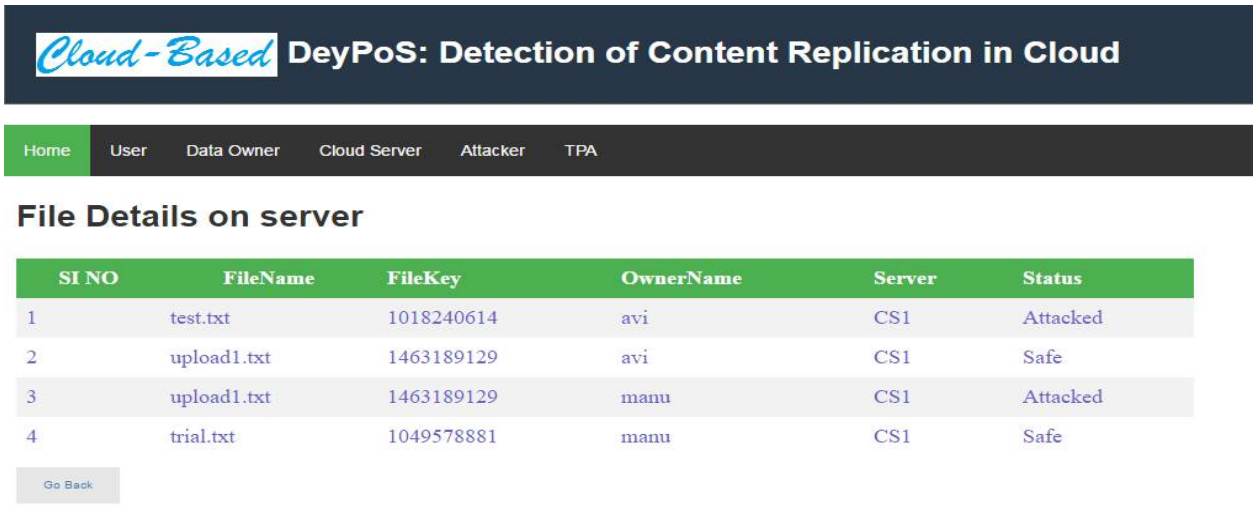

Figure 6: file status

Figure 6, shows the detailed information about the file stored in server.
Once after the auditor validate the file, it will regain its normal status and will be available for the end users for multiple operations.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

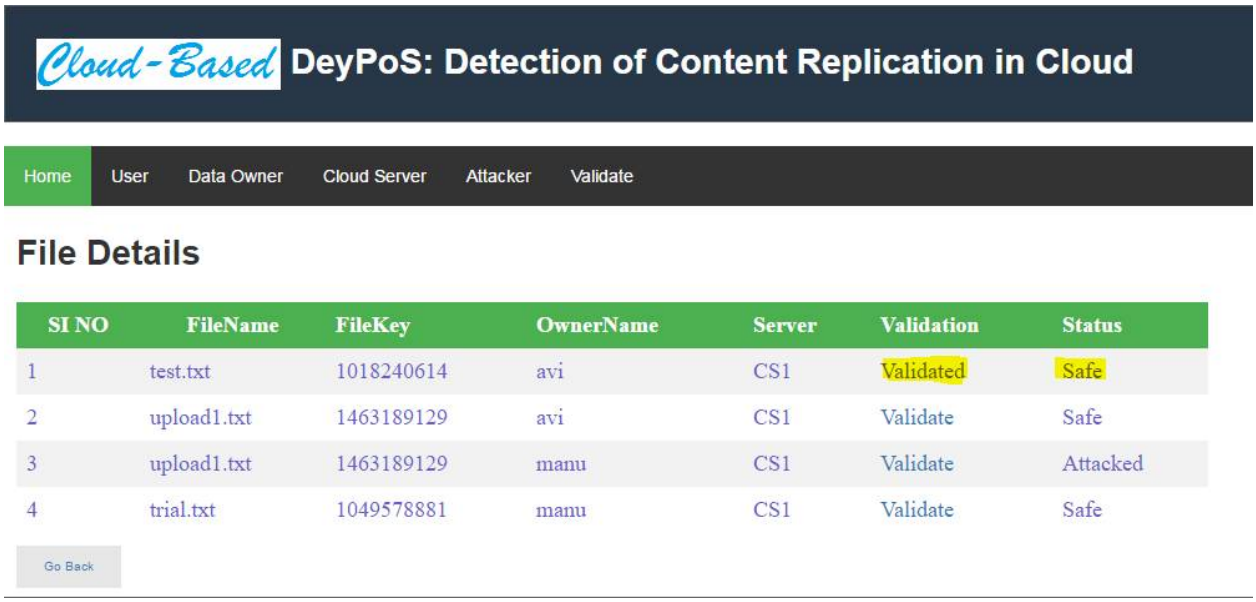

Figure 7: Validated file

As shown in figure 7, the file is validated by the auditor, later the file get back to its safe state.

VI. CONCLUSION

In this paper, a Deduplicatable dynamic PoS model is proposed to satisfy the comprehensive requirements in the multi-user environment cloud storage system .The proposed work designed a novel tool called HAT, which is an effective authenticated structure. Based on HAT, the work implements the Deduplicatable dynamic PoSalgorithm and proved that theproposed work is proficient in multi user cloud storage environment. The theoretical and experimental results show that the DeyPoS implementation is efficient, especially when the file size and the number of the challenged blocks are large. Finally the work is best suitable for better utilization of storage space.

REFERENCES

[1]. Xiao, Zhifeng, and Yang Xiao. "Security and privacy in cloud computing." *IEEE Communications Surveys & Tutorials* 15.2 (2013): 843-859.
[2]. Ardagna, Claudio A., et al. "From security to assurance in the cloud: a survey." *ACM Computing Surveys (CSUR)* 48.1 (2015): 2.
[3]. Ateniese, Giuseppe, et al. "Provable data possession at untrusted stores." *Proceedings of the 14th ACM conference on Computer and communications security*.Acm, 2007.
[4]. Erway, C. Chris, et al. "Dynamic provable data possession." *ACM Transactions on Information and System Security (TISSEC)* 17.4 (2015): 15.
[5]. Li, Jin, et al. "A hybrid cloud approach for secure authorized deduplication." *IEEE Transactions on Parallel and Distributed Systems* 26.5 (2015): 1206-1216.
[6]. Yan, Zheng, et al. "Encrypted data management with deduplication in cloud computing." *IEEE Cloud Computing* 3.2 (2016): 28-35.
[7]. Jagtap, Pradeep Sarjerao, and A. D. Gujar. "Encrypted Deduplication for Efficient Cloud Backup." *International Journal* 4.6 (2016).
[8]. Leesakul, Waraporn, Paul Townend, and Jie Xu. "Dynamic data deduplication in cloud storage." *Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on*. IEEE, 2014.
[9]. Shin, Young-joo, JunbeomHur, and Kwangjo Kim. "Security weakness in the Proof of storage, with Deduplication." *IACR Cryptology ePrint Archive* 2012 (2012): 554.
[10]. Zheng, Qingji, and Shouhuai Xu. "Secure and efficient Proof of storage, with deduplication." *Proceedings of the second ACM conference on Data and Application Security and Privacy*.ACM, 2012.
[11]. Naor, Moni, and Moti Yung. "Universal one-way hash functions and their cryptographic applications." *Proceedings of the twenty-first annual ACM symposium on Theory of computing*.