

## DATA DEDUPLICATION IN CLOUD STORAGE WITH PROOF OF OWNERSHIP

Dr. P. Sai Kiran<sup>1</sup>, J. Bharathwaj Reddy<sup>2</sup>, K. Gopi Sainadh<sup>3</sup>

<sup>1</sup>Professor, K.L. University, India

<sup>2,3</sup>Student, K.L. University, India

[nathg422@gmail.com](mailto:nathg422@gmail.com)

**Abstract:** Deduplication is the technique is to compare the data for removing the same type of the files in the cloud server. The main function of this technique is to utilize the storage capacity in the cloud server in more efficient way. Authentication is also one of the problem in the cloud technology. To provide confidentiality and proof of ownership for the data, some encryption technique is used to convert the plain text to cipher text before uploading into the cloud server. So, we introduce a technique to address the above issues.

**Keywords:** Deduplication, SHA-256, Random number generation, Proof of Ownership.

### 1. Introduction

Cloud computing is a technology where clients utilize the resources like compute, network and storage in which are present at various locations which can be access through high speed internet connection. The five major characteristics of the cloud are:

**On demand service:** This character mainly talks about whenever the request made by the cloud user for any type of resource. Only that resource will be given by the cloud service provider.

**Broad network access:** This character enables us handle the resources spread over the world through internet connection.

**Resource pooling:** Based on the requirements of the cloud user. The cloud service provider will produce the resources to the user through internet connection.

**Rapid elasticity:** This character based on the requirement of cloud user it will enables to expand the cloud or shrink the cloud.

**Measured service:** Based on the usage cloud user will be charged [3].

There are three major types of cloud. Private cloud: In this cloud number of users will be restricted to a particular organization or company. Public cloud: It is an environment which is available to all type of user where security concerns are raised compared to private and hybrid environments.

**Hybrid cloud:** It is a combination of two or more types of cloud environments.

#### 1.1 Convergent Key encryption

Convergent key encryption is an encryption technique in which the data [7], [8] a item is encrypted with the hash of the same file. Hashing technique may vary according to the security concern.

#### 1.2 Proof of Ownership (PoW)

The main aim of the POW technique enables user to prove himself as an owner for the data copies present on the cloud server by providing required information. PoW is developed by the interactive algorithm by the cloud server and the user [6].

#### 1.3 Existing System

The existing model stores data files in the public cloud and keys in the private cloud. User need to request the respective keys from the private cloud to access the files from the public cloud [9]. As the request is processed by the private cloud and the keys are given to the user, after retrieving keys from the private cloud the user access the files present in the public cloud by using the private keys [7].

#### 1.4 Limitations

After receiving keys from the private cloud, if there is any security breach at the user side, these keys are used by the attackers to access

the data files from the public cloud because there is no authorization check at the public cloud.

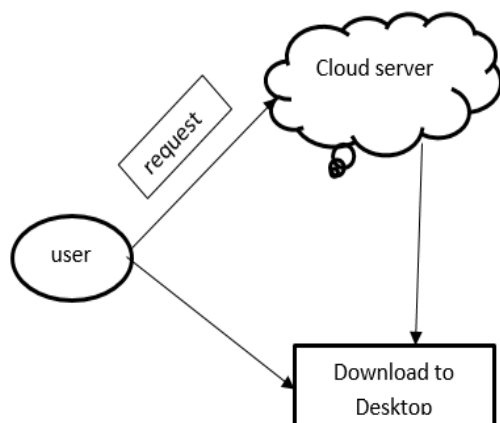
## 2. Proposed System

In our project user requires to generate hash value of their file that they want to upload into the cloud using SHA-256 technique. The hash value is verified in data base to make sure that file is not re-uploaded by the user. The hash value available on the data base it simply means that the file is present in the server. The user is now required to provide some kind of meta data to the server so that this meta data is encrypted and used as one of the security detail about the file.

The security of the file is not compromised by having the hash value and the encrypted text. So, to make it complete we are generating one time password(OTP). To generate this OTP, we use random number generator. The OTP is generated and stored in data base until the verification is done.

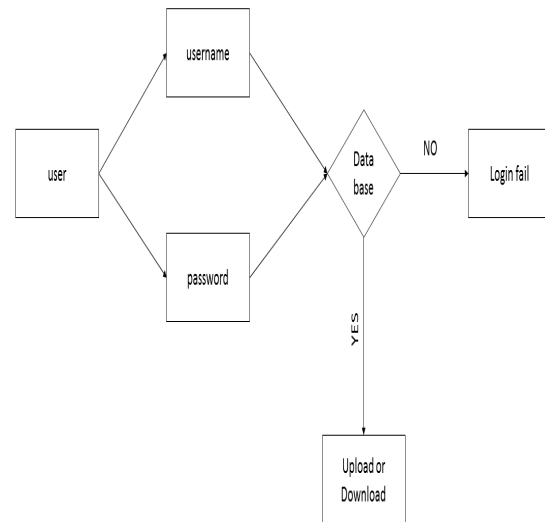
As the user need to login into their account. To download the file by their request, the user is required to enter the hash value as input and the encrypted data text for accessing it. After the user enters the hash value and encrypted text, the user is required to submit the one-time password to the server.

This method provides confidentiality and reliability to the user data. Even when the user account is compromised the user data is still secure as the attacker needs the OTP and encrypted text which is generated from the meta data.

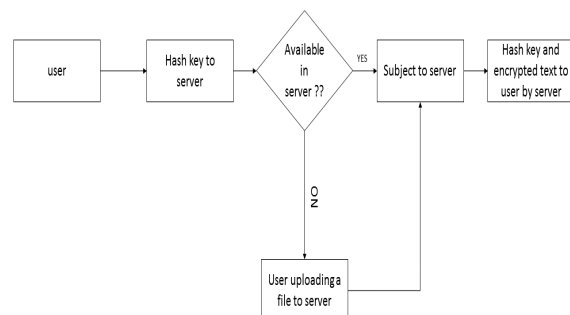


**Figure 1.1** System Overview

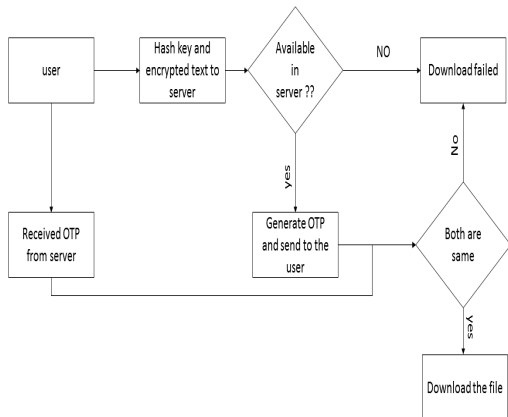
### 1.6 Architecture



**Figure 1.2** User Login Form



**Figure 1.3** Uploading a File in Server



**Figure 1.4** Downloading a file from Server

### 1.7 Algorithm

#### *Algorithm to upload a file*

```

require(user_info);
if(is_avail(user_info()))
{
require(hash(file));
if(avail(hash(file)))
printf ("\n file already exists in the data base");
else
{
printf ("\n file doesn't exist in the data base");
require(upload_file());
produce_enc(uploaded_file);
}
}
else
printf ("\n user doesn't exist");

```

#### *Algorithm to download a file*

```

require(user_info);
if(avail(user_info))
{
require(hash(file));
require(ency_text);
if(check_db(hash(file))&&check_db(ency_text))
{
produce(randm_num);
update_db(randm_num);
//send the generated random number to the user
if(check_db(user_randm))

Printf ("grant access to download the file");
download_file(file);

```

```

}
else
printf ("given otp is wrong");
}
else
{
Printf ("given information is wrong");
require(hash(file));
require(ency_text);
}
}
}

```

### 3. Related Work

**Client Side Deduplication:** It is the technique many of the organizations or users are using in cloud technology. It is used to back up the user data and eliminates the same copy of the data during backup time [4]. It minimizes the capacity of the file in the server by the user through local area network. The security for the data in this method is very high compared to other methods.

**Random Number Generation:** Using this method we can generate the random number sequence which is unpredictable. There is a less chance to identify the exact number in this method. In the above model this method is applicable in the development of the one-time password.

#### *Algorithm for Random Number Generation:*

```

Static unit32_r P [4098], d=362437;
initial_random(unit32_r z)
{
int b;
P[0] =z;
P[1] = z + J;
P[2]= z + J + J;
for ( b =5; b<4098 ;b++)
P[b]= (((P[b-3])^P[b-2])^I)^b);
}
unit32_r random(void)
{
unit64_r r, s=18782LL;
Static unit32_r b=4097;
unit32_r z, f= 0xfffffffffe;
b = (b+1) & 4097;
r=s *P[b] +d;
d=(r>>32);
z= r + d;
if(z<d)
{
z++;
d++;
}
}

```

```
return (P[b] = f-z);
}
```

#### 4. Conclusion

From the model, we are able to say that data deduplication is well performed by using convergent encryption technique compared to other algorithms. To ensure the authentication of the user we used the concept of OTP which leads to satisfy the requirements of Proof of Ownership.

In future, we would like to add some additional features like some security questions which yields higher security to the data.

#### References

- [1] B. Tirupati Reddy, M.V. P Chandra sekhar, L.S.S Reddy, V. Krishna Reddy, P. Sai Kiran, "A Survey on Assured File Deletion in Cloud Environment", International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 9, Number 23 (2014) pp. 19899-19907
- [2] B. Tirupathi Reddy, Ramya U, M.V.P. Chandra Sekhar, "A comparative study on data deduplication techniques in cloud storage".
- [3] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
- [4] A.T. Clements, I. Ahmad, M. Vilayannur, and J. Li, "Decen-tralized Deduplication in San Cluster File Systems," in Proc. USENIX ATC, 2009, p. 8.
- [5] libcurl. <http://curl.haxx.se/libcurl/>.
- [6] "Proofs of Ownership in Remote Storage Systems," by Shai. Halevi et al.
- [7] Message-locked encryption and secure deduplication by M. Bellare, S. Keelveedhi, and T. Ristenpart in the year of 2013.
- [8] Reclaiming space from duplicate files in a serverless distributed file system by J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer in the year of 2002.
- [9] S.Rajeswari and M.S.Sushmitha," Augmented reality and its application", International Innovative Research Journal of Engineering and Technology, March. 2016.
- [10]"A hybrid Cloud Approach for Secure Authorized Deduplication", by Jin li et al at iee transactions Year 2014.
- [11]"secure deduplication with efficient and reliable Convergent key management" by Jin li et al, iee transactions in the year 2014.
- [12]"Boosting efficiency and security in proof of ownership for deduplication" by R. Di Pietro and A. Sorniotti.



