

An Efficient Approach To Cease Duplication And Increase Storage Space On Cloud

¹Poonam Pathak, ²Suyash Benudas Petkar, ³Sidhesh Madhukar Gite, ⁴Nishank Prasad Kadam

¹Professor, ^{2,3,4}BE IT Students
Pillai HOC College of Engineering and Teechnology
Navi Mumbai, India

Abstract: In today's world where data storage, maintenance, and data integrity are a necessity of the moment. Cloud Computing and its services play a significant role for this purpose. Cloud Computing permits delivering of services where resources are retrieved in a "pay as you go" by manner to the user. It has been found that data deduplication creates a crucial issue for regular users needing transactions of data. The same file can be stored by totally different users multiple times causing piling of similar files on the cloud. This results in a vast amount of duplicate data on the cloud. We propose a system that could eradicate inefficiencies such as data duplication. In this system data within the file is compared with a similar looking file on the cloud here data duplication of the similar looking file can be removed using hashing technique which restricts the user to upload an identical file. The file could also be a multimedia file on the cloud.

Keywords: Cloud, Deduplication, Storage

I. INTRODUCTION

Cloud computing is a versatile, cost-effective and proven delivery platform for providing business or consumer IT services over the Internet. Cloud computing supports distributed service-oriented design, multi-users, and multi-domain administrative infrastructure, it is more liable to security threats and vulnerabilities. At present, a significant concern in cloud adoption is its privacy, storage, and security. Intrusion prospects within a cloud environment are many and with high gains. Duplication issues are of more concern to cloud service providers who are actually hosting the services. In most cases, the provider must guarantee that their infrastructure is secure and user's data and applications are safe by implementing security policies and mechanisms. While the cloud customer must ensure that the provider has taken proper security measures to protect their information. The issues are organized into several general categories: trust, architecture, identity management, software isolation, data protection, availability Reliability, Ownership, Data Backup, Data Portability and Conversion, Multiplatform Support and Intellectual Property.

II. PURPOSE

The purpose of the system is to create a system where a user uploads a file on the cloud as per his/her use. The consequence of this action is duplication on files on the cloud which wastes valuable cloud space and reduces efficiency. So we propose a system where this drawback is reduced. The system encrypts the user's data on their device before it is synchronized to the cloud. Thereby, the protection of the data is assured.

III. CONSTRAINTS

We here define the constraint using the triple constraint of project management:



Cost: The cost of developing a system like this is quite minimal. For just this demo purpose we are deploying the project on a rental cloud service offered online, we are implementing this project in java as everything we need related to is freely available on internet.

Time: The time required for developing the project is depended on the complexity of the project. According to the current specification of the system it will require almost 8-9 months for deploying the final product for the user.

Scope: The main work begins after analyzing the whole process of working of the cloud. In the analyzing phase, we collected most of the data that is required by the process. This has helped in listing out the risks and milestones in our project. Based on the

information we were able to agree on the final scope and dates of the project. After getting all the required data we had first started developing the schema of the database that will hold most of the data. After creating the schema, we will check for the proper methods that are required to transfer data in various modules of the system. Next step will be coding of various phases data transfer between the browser and the server. We will probably use Java, PHP for our application. The final step will be the deployment of could with the features.

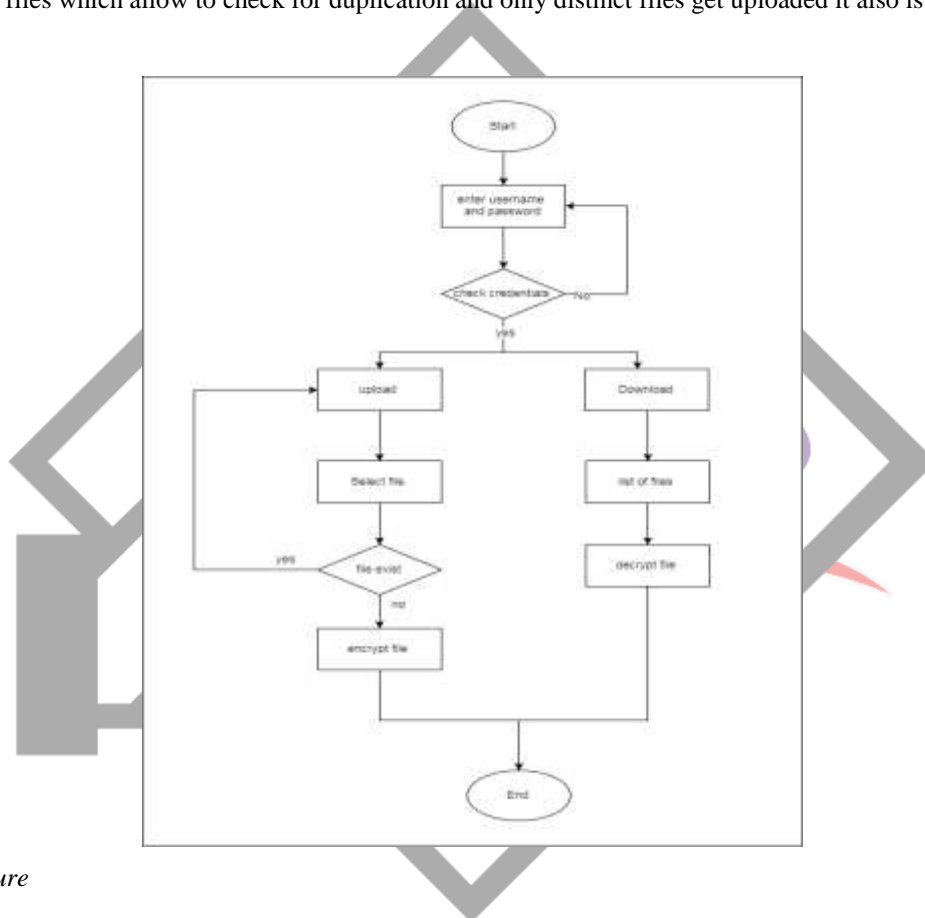
IV. OVERALL SYSTEM DESCRIPTION

A. Existing System

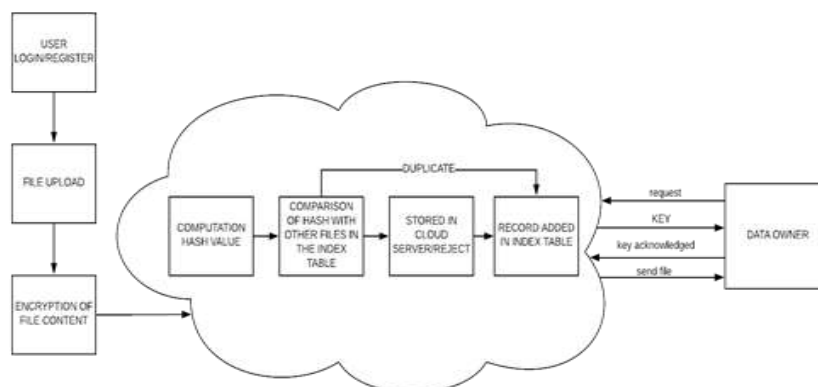
Google Drive is a service that lets the user store their personal files within the “cloud”. The drive is a direct competition to different services like Dropbox, iCloud, SpiderOak, and Microsoft OneDrive.

B. Implemented System

This system encrypts the user’s data on their device before it is synchronized to the cloud. Thereby, the protection of the data is assured. The software can be used with Microsoft OneDrive, SpiderOak, iCloud, and many more cloud providers. In order to keep control of the content of data, the data goes through encryption which is provided by cloud providers. The system encrypts it with AES. Thus, data integrity is maintained. We conjointly use SHA algorithm (one of the SHA-2 family of hash functions) to create hash functions of the files which allow to check for duplication and only distinct files get uploaded it also is used for strengthening the security.



C. System Architecture



The main components of the proposed system are:

User: The user uploads, downloads the data. This data needs to be encrypted before uploading and decrypted after downloading from the cloud. This ensures the integrity of the file also security assurance to the user.

Cloud: The cloud (SaaS) has operations such as deduplication check that is done by comparing hash values of comparable files and duplicate files are rejected. An index record of stored files is created.

CONCLUSION

As data duplication is a major issue both for cloud service providers and users the system provides security, privacy, storage space and vital efforts of users who need to store, retrieve and process a file on the cloud which could be a duplicate file. The system enhances the storage capacity of every file which could be otherwise duplicated on the cloud, thus increasing storage efficiencies for the cloud service provider.

FUTURE SCOPE

The system has a scope to be deployed on SAAS platform of the cloud. In the future, we could produce the same system to be deployed on IAAS platform of cloud, which could be highly beneficial for cloud service provider for companies and its users.

ACKNOWLEDGMENT

We remain immensely obliged to Prof. Poonam Pathak for providing us with the moral and technical support and guiding us. We would also like to thank our guide for providing us with her expert opinion and valuable suggestions at every stage of the project. We would like to take this opportunity to thank Prof. Monisha Mohan, Head of Information Technology for her motivation and valuable support. This acknowledgment is incomplete without thanking teaching and non-teaching staff of the department of their kind support. We would also like to thank Dr. Madhumita Chatterjee, Principal of Pillai HOC College of Engineering and Technology, Rasayani for providing the infrastructure and resources required for the project.

REFERENCES

- [1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou" A Hybrid Cloud Approach for Secure Authorized De-duplication" in vol: pp no-99, IEEE, 2014
- [2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
- [3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message blocked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.

