Secure Data De-Duplication On 2PVC Algorithm In Web Services

N. Amutha[1], Dr. S. Sujatha[2]

[1]*Master of computer application,Anna university,Tiruchirappalli-620 024*

[2]*Associate professor,Department of computer application,Anna university,Tiruchirappalli-620 024*

Tamil Nadu, India

---***---

Abstract - *The rapidly increasing amounts of data produced worldwide, shared-network multi-user web services systems are becoming very popular. Distributed transactional database systems deployed over web services, entities cooperate to form proofs of authorizations that are justified by collections of certified credentials. Data de-duplication is a process by which a storage provider only stores a single copy of a file (or of its part) owned by multiple users. Define the notion of trusted transactions when dealing with proofs of authorization. Accordingly, we propose several increasingly stringent levels of policy consistency constraints, and present different enforcement approaches to guarantee the trustworthiness of transactions executing on web services.*

Key Words*:* Security, De-duplication, Policy Consistency, Web Authority,Web Services ect.
.

1. INTRODUCTION

Web Services has recently emerged as a computing paradigm in which storage and computation can be outsourced from organizations to next generation data centers hosted by companies such as Amazon, Google, Yahoo, and Microsoft. Such companies help free organizations from requiring expensive infrastructure and expertise in-house, and instead make use of the web services providers to maintain, support, and broker access to high-end resources. The data de-duplication is a process by which a storage provider only stores a single copy of a file (or of its part) owned by multiple users. There are four different de-duplication strategies, depending on whether de-duplication happens at the client side (i.e. before the upload) or at the server side, and whether it happens at a block level or at a file level. Client-side data de-duplication is more beneficial than server-side since it ensures that multiple uploads of the same content only consume network bandwidth and storage space of a single upload. De-duplication technology helps reduce storage costs and network bandwidth utilization by eliminating duplicate data blocks when you back up and transfer data.

Acronis Backup Advanced de-duplication helps you to:
1. Reduce storage space usage by storing only unique data

2. Eliminate the need to invest in data de-duplication-specific hardware

3. Reduce network load because less data is transferred, leaving more bandwidth for your production tasks

The 2-phase commit (2PC) protocol is a distributed Algorithm to ensure the consistent termination of a transaction in a distributed environment. Thus, via 2PC a unanimous decision is reached and enforced among multiple participating servers whether to commit or abort a given transaction, thereby guaranteeing atomicity.

1.1 ASP .NET Platform

ASP.NET is a set of Web development tools offered by Microsoft. Programs like Visual Studio, Of course, programmers can write their own code and scripts and incorporate it into ASP.NET websites as well. Though it often seen as a successor to Microsoft's ASP programming technology, ASP.NET also supports Visual Basic.NET, JScript .NET and open-source languages like Python and Perl.
In order for an ASP.NET website to function correctly, it must be published to a Web server that supports ASP.NET applications. Microsoft's Internet Information Services (IIS) Web server is by far the most common platform for ASP.NET websites. While there are some open-source options available for Linux-based systems, these alternatives often provide less than full support for ASP.NET applications.

1.2 SQL server

MS SQL Server is a relational database management system (RDBMS) developed by Microsoft. This product is built for the basic function of storing retrieving data as required by other applications. It can be run either on the same computer or on another across a network. This tutorial explains some basic and advanced concepts of SQL Server such as how to create and restore data, create login and backup, assign permissions, etc. Each topic is explained using examples for easy understanding. It is a software, developed by Microsoft, which is implemented from the specification of RDBMS. It is also an ORDBMS.It is platform dependent. It is both GUI and command based software. It supports SQL (SEQUEL) language which is an IBM product, non-procedural, common database and case insensitive language

2. LITERATURE SURVEY

"**Jan Stanek, Member, IEEE, and Lukas Kencl, Member, IEEE, "Enhanced Secure Thresholded Data De-duplication Scheme for Cloud Storage"** " As more corporate and private users outsource their data to cloud storage, recent data breach incidents make end-to-end encryption increasingly desirable. Unfortunately, semantically secure encryption renders various cost-effective storage optimization techniques, such as data de-duplication, ineffective. With the rapidly increasing amounts of data produced worldwide, shared-network multi-user cloud storage systems are becoming very popular. However, concerns over data security still prevent many users from migrating data to remote storage.

"**Deepali Choudhari1, R. W. Deshpande2," De-duplication Techniques in Storage System"**"" Cloud storage is getting popular more and more as it is low cost and on demand use of large storage. To make data management scalable in cloud computing, the de-duplication concept is used. According to the analysis report of IDC, the volume of data in the world is expected to reach 40 trillion gigabytes in 2020 [1] De-duplication can take place at 1) File level, in this it detects redundant data within the files or 2) Block level, in this it detects redundant data across the files and removes those data of identical data files.

"**K. Keerthana*, C. Suresh Gnanadhas, RT. Dinesh Kumar" A Survey On Managing Cloud Storage Using Secure Deduplication"**"" Data Deduplication is a method of data compression that eliminates the replicates of data. This process will only save the unique copy of data in a storage media. When a data redundancy occurs it provides a pointer to access that data and ignore the process of saving the redundant data in storage. This process automatically saves the storage space. Reducing the storage space requirements will reduce the cost on disk expenditures, bandwidth usage for transacting data over internet and helps us to use the storage space efficiently. It identifies the redundant data by comparing the data already present in the storage.

"**Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang," Secure Data De-duplication with Dynamic Ownership Management in Cloud Storage".** "Cloud storage services, de-duplication technology is commonly used to reduce the space and bandwidth requirements of services by eliminating redundant data and storing only a single copy of them. De-duplication is most effective when multiple users outsource the same data to the cloud storage, but it raises issues relating to security and ownership. Proof of- ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a robust way. However, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers de-duplication because of the randomization property of encryption.

3. EXISTING SYSTEM

To provide scalability and elasticity, web services often make heavy use of replication to ensure consistent performance and availability. Web services rely on the notion of eventual consistency when propagating data throughout the system. This consistency model is a variant of weak consistency that allows data to be inconsistent among some replicas during the update process, but ensures that updates will eventually be propagated to all replicas. Consistency problems can arise as transactional database systems are deployed in web services environments and use policy-based authorization systems to protect sensitive resources.

3.1 Disadvantages of Existing System

- Consistency problems can arise as transactional database systems are deployed in cloud environments and use policy-based authorization systems to protect sensitive resources.

- The system may suffer from policy inconsistencies during policy updates.

- It is possible for external factors to cause user credential inconsistencies over the lifetime of a transaction.

4. PROPOSED SYSYTEM

The Propose a Two-Phase Validation Commit (2PVC) protocol that ensures that a transaction is safe by checking policy, credential, and data consistency during transaction execution. Identifies transactions that are both trusted and conform to the ACID properties of distributed database systems. A transaction is safe by checking policy, credential, and data consistency during transaction execution. Most suitable in various situations.

4.1 Advantages of Proposed System

- Guarantee the trustworthiness of transactions executing on cloud servers.

- A transaction is safe by checking policy, credential, and data consistency during transaction execution.

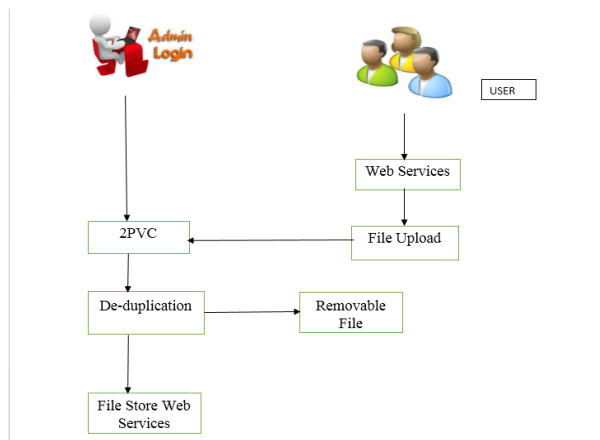- Most suitable in various situations.

5. ARCHITECTURE DIAGRAM

Fig -1: Architecture diagram of De-duplication

5.1. Module Description

There are five modules define in the system.

- User Registration

- File Upload

- Deduplicate file

- View and Download File

- Transaction Manager

User Registration

This module is designed for new users who visit this project. This module is main module that is user can select the file type here then only the use can upload the files later. The selected file type only can be uploaded by the user.

File Upload

This module is designed for the user to upload files for storage. The user can provide file details and select file to upload. After uploading files user can get the proper message. Uploaded files can be retrieved later for future reference.

De-duplication File

Data de-duplication is a process by which a storage provider only stores a single copy of a file (or of its part) owned by multiple users. Identify a duplicate file are detected the store in original file.

View and Download File

This page is designed for user. After proper authentication uploaded files can be viewed and download in this module

Transaction Manager

The transaction manager verifies the policies (i.e.) file type which is selected by the user. The selected file type only the user can upload and download.
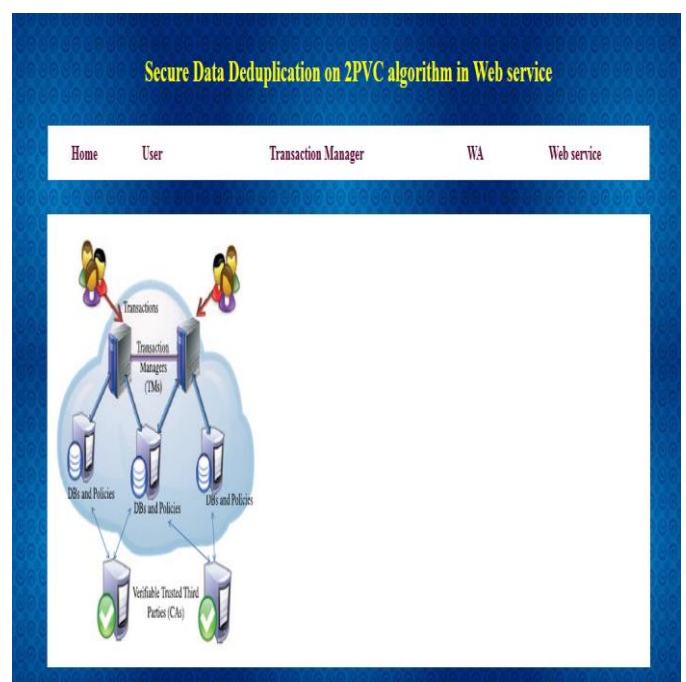
6. RESULT AND OUTPUT

Fig 2: Home page

Fig 3: New User Register

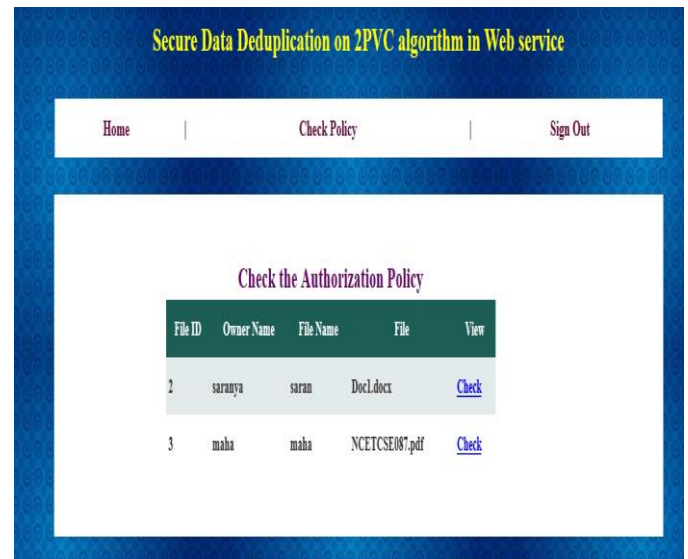**Fig 4:** Approve User

Fig 6: Check Policy
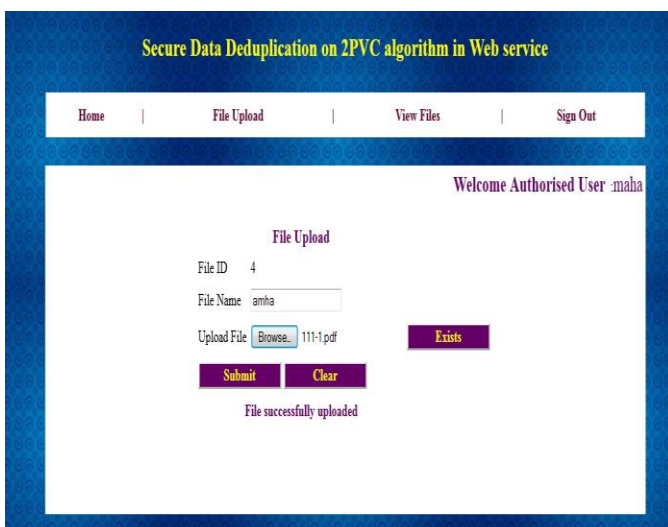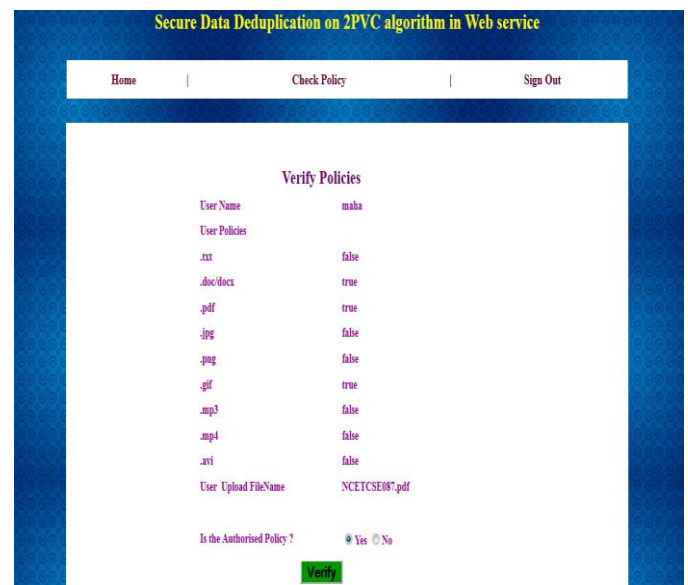

Fig 5: File Upload

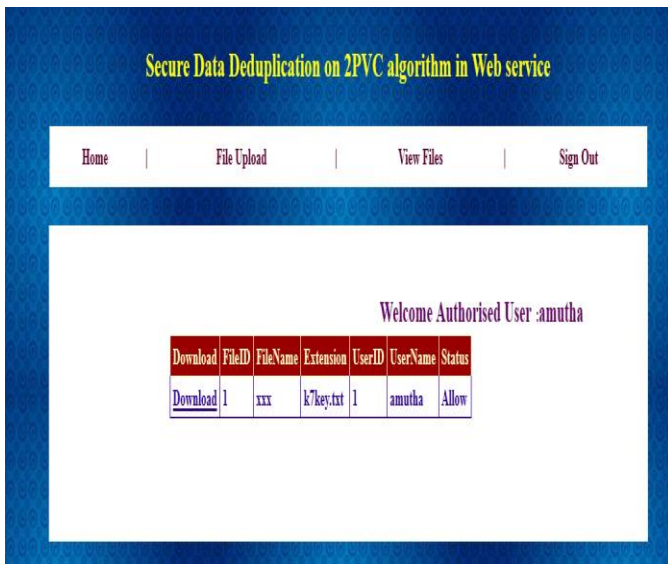**Fig 7:** Verify Policy

Fig 8: Complete Policy

Fig 9: View File

7. CONCLUSION

This work deals with the inherent tension between storage optimization methods and end-to-end processed de-duplication file detected. To evaluate space savings efficiency and the computational and communication overhead of the proposal, we present an extensive performance evaluation using real datasets and real-like environment. To provide comparison to other state-of-the-art secure de-duplication schemes we include their prototypes in the performance evaluation. The results show that there is no clear "winner" – none of the schemes outperforms others under all conditions, each has advantages and disadvantages with regards to a particular data mix, environment and requirements. This project conclude in detected the de-duplication file and store the web service.

REFERENCES

[1] Jan Stanek, Member, IEEE, and Lukas Kencl, Member, IEEE ," Enhanced Secure Thresholded Data Deduplication Scheme for Cloud Storage," in 2016

[2] Deepali Choudhari1, R. W. Deshpande2, ''Secure Data Deduplication for cloud storage,'', 2014.

[3] "K . Keerthana*, C. Suresh Gnanadhas, RT. Dinesh Kumar" A Survey On Managing Cloud Storage Using SecureDeduplication"in 2013.

[4] "Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang," Secure Data De-duplication with Dynamic Ownership Management in Cloud Storage",2015

[5] John Paul ANBU1 and Sridevi JETTY2,"Deduplication Techniques in Storage System," in 2015.

[6] Jian Liu, N. Asokan, Benny Pinkas," Secure Deduplication of Encrypted Data without Additional Independent Servers.