Department of Computer Science & Engineering

University Of Moratuwa

LBRSync

Low Bandwidth Replica Synchronization for Cloud Storage Systems

Initial Project Proposal

Project Supervisor: Prof. Gihan Dias

Group Members: 090022B -   P.D.D.A. Anicitus

090236J   -   Y. Jenananthan

090324D -   M. Mayuresan

090471A -   S. Sathyavarathan

Introduction

File Synchronization is the process of ensuring that files in two or more separate locations contain identical up-to-date details. If a file is added, changed or deleted from one location, the synchronization process will perform the same action (add, change, or delete the same file) at the other location/locations. Files are copied in both directions, keeping the two or more locations in sync with each other.

Wherever files may be, when two locations are synchronized, the most current version of a file is available at both locations, regardless of where and when it was last modified. A better file synchronization avoids duplicates and do not delete any file by accident.

It is used in software distribution mirror sites, backup and restores systems, versioning systems and content distribution networks and also in cloud computing for keeping the exact copy of the local machine file, folder or folders in any remote server at any time. The two locations of the syncing process are connected with network such as internet, and data is sent through it.

Importance of the project

Cloud computing has become an essential part in current computing and users prefer having copy of their selected important folders in a remote server, enabling access to them from anywhere at any time and sharing them with anyone, providing selected privileges. A consultant constantly traveling between various sites may want to sit down and access the same project files from every location. File synchronization, which uses networks for data transfer, does a major role in making it successful.

However people often have occasion to work over slower networks. Even with the broadband Internet access, a person working from home usually only has a fraction of a Megabit/sec of upstream bandwidth. A company which has offices in several locations may have many users collaborating over a single slower connection. For video streaming and for web applications, interactive programs sufficient network bandwidth is essential. So the file synchronization process which should work in those environments should be capable to perform. Given the size of the collections, large amount of data transferring will degrade the performance when the bandwidth of the connection between the source and the destination is low and will be more expensive if the connection provider charges based on the data consumption.

Objectives of the project

PRIMARY OBJECTIVE
- Perform literature survey, compare & contrast different methods/techniques/implementations of file synchronization processes, and identify methodologies that can synchronize files efficiently and cost effectively in a low bandwidth connection environment.
- Optimize the methodologies to design an appropriate architecture to perform the file synchronization with a minimum amount of data transfer over the network.
- Implement the proposed architecture

SECONDARY OBJECTIVE
- Compare the developed solution with the existing solutions.
- Adding features to the product such as version control

Assumptions/Limitations

- The two machines are connected by a low-bandwidth high-latency bi-directional communications link
- One of the machines has high bandwidth connection with the internet.

Scope

The project will be for low bandwidth connections and for two way synchronization in cloud storage systems.

Methodology

PHASE 1

First we need to have an in-depth understanding of the file synchronization methodologies/techniques to accomplish the primary objectives. In order to study these aspects we plan to do a literature survey on the existing implementations, solution technologies and protocols. Further we are going to evaluate the existing approaches by comparing and contrasting the implemented algorithms/protocols, their characteristics and drawbacks. In parallel to the above, fulfilling the secondary objectives, some research work will be also conducted in finding out efficient file synchronization techniques that have been implemented and used in low bandwidth connections.

PHASE 2

In this phase we will work on coming up with an efficient algorithm and a design that could be either a modification of existing implementations or a completely new one, which suits development of the file synchronization tool working in low bandwidth connection environments.

PHASE 3

This third phase is the implementation of file synchronization tool in a selected platform from previous phases. Here rather than implementing the chosen design various different

techniques will be tried out to improve efficiency. Finally the tool will be evaluated on how well it works in expected environment and as a result satisfies the objectives of the project.

Deliverables

- Initial Project Proposal
- Project Website
- Implementation of File synchronization plug-in or software
- Analysis results of the efficiency of the developed software and other currently available software
- Project Report
- 2 Videos(2 minute video -which is for marketing purpose of the product & another 10 minute video-which will describe the features of the product)
- Research Paper for the approach used in the project
- User Manual
- System Requirement Specification
- System Design Document

Project Plan for the first phase