TITLE

Optimized data transfer scheduler for Onedatashare

GROUP MEMBERS

AKASH KUMAR ROY, OSBAN ANIL CEREJO, SWAPNIL KISHORE,
VISHAL SINGH

DESCRIPTION

This project will involve implementing an optimized data transfer protocol for end-to-end communication with the goal to improve throughput by finding a balance between the factors affecting throughput. The main factors being considered are pipelining, parallelism and concurrency. Optimization of big data transfers over inter-cloud and intra-cloud networks is a challenging task that requires joint-consideration of all of these parameters.

PIPELINING: Pipelining deals with transferring large numbers of small files [1]. It has two major goals: first, to prevent the data channel idleness and to eliminate the idle time due to control channel conversations in between the consecutive transfers. Secondly, pipelining prevents TCP window size from shrinking to zero due to idle data channel time if it is more than one Round Trip Time (RTT).

PARALLELISM: With parallel streams, portions of a file are sent through multiple TCP streams and it is possible to achieve multiples of the throughput of a single stream. Setting the optimal parallelism level is a very challenging task and several models have been proposed in the past.

CONCURRENCY: Concurrency is especially good for small file transfers, and overcoming end system bottlenecks such as CPU utilisation, NIC bandwidth, parallel file system characteristics. The Stork data scheduler [20], [21] has the ability to issue concurrent transfer jobs and in most of the cases, concurrency has proven itself over parallelism.
As shown in the above figure, this project will be dealing with implementing the 'Scheduling and Optimization Service' as part of the user services of OneDataShare.