

1 Objectives

- Implement a distributed *grep* command that can fetch certain informations from log files stored on multiple machines.

2 Analysis and Design

Since the *grep* command requires searching in certain log files, we adopted a client/server system design, where servers being in charge of searching and client (client itself is also a server) will receive the result.

The difficult part of this project is to make the "grep" command fast. Each server stores an approximately 100M log files, which will take a lot of time to find them one by one. For each server there will be three main procedures:

- Client: Create a thread to build a request
- Server: Find out all the lines that matches the pattern
- Server: Return the result to the client

The client will **concurrently** make socket connection to servers and wait for responses for servers. This will largely improve the efficiency by saving the time waiting for the response one by one.

In order to verify the correctness of the system, we designed 7 **unit test** cases, where 5 cases are designed for case in regular expression, frequent pattern, and infrequent patterns, and another 2 cases are designed for fault tolerance

3 Performance Analysis

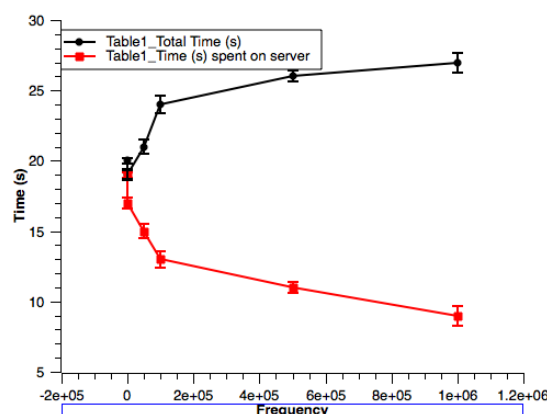


Figure 1: Performance of our System. We can see that as the frequency increase, the time spent on server decrease. This is because the pattern with high frequencies requires longer time to match. The total time goes up because it takes more time to transmit the data.