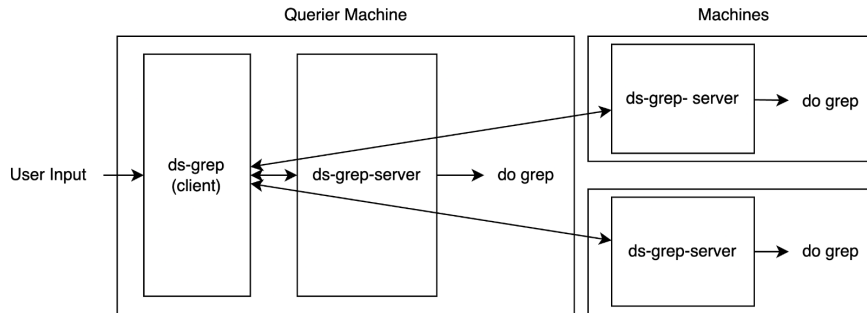# Machine Programming 1 – Distributed Log Querier

Che-Kuang Chu(ckchu2), Jhih-Wei Lin(jhihwei2)

Design

Our design follows a client-server model. On the client side (ds-grep), its responsibilities include receiving user input, sending it to multiple servers concurrently (using goroutines), receiving results from the servers, and displaying them on the terminal. The server side (ds-grep-server) has two primary functions. The first is to listen for TCP connections from the client side, and the second is to execute the 'grep' command and return the results to the client.

Regarding fault tolerance, we have implemented a connection timeout constant of 5 seconds on the client side. Consequently, if a server fails to respond within this timeframe, the client side will categorize the machine as non-responsive and proceed to its next tasks.

Unit Testing

Here are the scenarios of the unit tests we implemented:
- Grep: query one machine and multiple machines with frequent, somewhat frequent, infrequent patterns, regular expressions, and patterns that only appear in one or some machines.
- Simultaneously query: query one machine with multiple clients simultaneously.
- Fault-tolerant: query one normal machine and a malfunctioning machine simultaneously, and solely query a malfunctioning machine.

Performance

We executed the query on VM1 and queried VM[1-4]. Each VM contains a log file of size 60MB.
Five patterns: rare, somewhat frequent, frequent, and patterns that only appear in one or some machines.

We expected that frequent and somewhat frequent patterns might perform similarly, while rare patterns might be the slowest. However, we were surprised to find that somewhat frequent patterns were the slowest, while frequent and rare patterns had similar performance.

Besides, we observed that patterns appearing in just one file were slower than those appearing in some files. This was in line with our expectations, as lower occurrence patterns generally require more time to locate. However, we were astonished to discover that it was nearly two times slower than the frequent pattern appearing in all the log files.