# Team 55 (yixin10, yangt2) - MP 1 Report

## 1. Querier Design

**Design of Client:** Client creates multiple threads where each thread handles query to one particular remote machine. Once all threads finish execution, results from distributed log files will be printed on the terminal. Within each thread, first try to connect to the server through socket. If it fails, report the failure and return immediately. If it succeeds, determine whether the server to be connected is itself. If it is, directly search local log file. Otherwise, send to remote server the object containing pattern and option of grep command. Then thread receives the list including the results of that particular server, and write the results into a local file, which will then be read by the main thread.
**Design of Server:** Use infinite loop to listen connection request from client. When server connect the client successfully, server receive the object transmitted from client, and the object contains the pattern or regexp and option for grep command. Server uses Unix4j to implement grep instruction, server reads local log file to generate a list of String, and the list is the result of grep, and we send the list back to client.
**Design of Unit Test:** Generate test patterns including frequent pattern, infrequent pattern and regular expression to match on demo log files, and compare the results with true results.

## 2. Query Latency Discussion

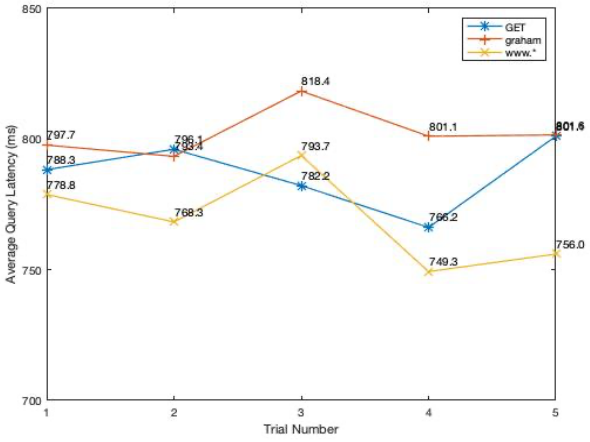


Figure 1 Average query latency

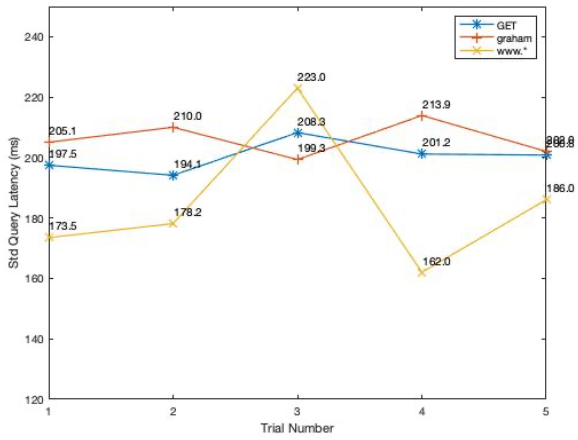


Figure 2 Standard deviation of query latency

Figure 1 illustrates the average query latency for each of the three patterns (frequent “GET”, in frequent “graham” and regexp www.*), Figure 2 illustrates the standard deviation of query latency for each of the three patterns, and each data point is executed five trials. It can be seen from Figure 1 that there is not much difference among the average query time of three different patterns, which means “grep” command spends almost the same amount of time on searching difference patterns. Specific to one pattern, there is some time gaps among five trials, which means network transmission has certain impact on the efficiency of distributed log query. Additionally, it can be seen from Figure 2 that the standard deviation is relatively large for all three patterns, which means log query on different machines has a big difference, either because grep command itself, or network transmission, or both.