Karl Munson
CSE 130

WRITEUP

Testing:

In order to test the load balancer I used primarily whole system testing in testing functionality and then testing in specific circumstances to repair bugs. I used bash scripts provided for testing assignment 2 to also test assignment 3 and making sure that the correct was being produced. I also used the provided http server in order to guarantee that all errors that appeared were coming from the load balancer.

First I built the getopt() portion to correctly parse in input from the arguments testing all combinations of optional flags and having flags in any order. After I tested in a similar way to assignment 2 starting with getting a single request of all types to be correctly forwarded between the client and a single http server by the load balancer. After getting functional forwarding I implemented the error cases for a request such as if all servers were done or if a server stopped responding. After ensuring correct functionality of forwarding a request I built the healthcheck function to allow for multiple servers to be chosen from. I then checked that healthcheck could recognize immediately if a server was offline and that it could be activated in the event of the correct number of requests or after the timer.

At this point all of the functions were built I started piecing them together making sure that they all worked together to select the correct server. I tried taking down and bringing back servers at different times to ensure that this portion of the functionality worked. I also tested which server would be selected by testing servers with the same number of entries but different numbers of errors as well as comparing operational servers to dead servers.

After checking to make sure selections were correct I implemented the parallel connections for the load balancer and tested the queue of file descriptors to check that if all of the threads were busy that it would still work.

Question 1:

Considering performance attributes for the machine running the server would be pointless in this case as all of the servers are running on different ports of the same machine which would mean they all have the same performance attributes.

Question 2:

If you allowed some processing of the client request you could catch some errors and invalid requests meaning that servers would not need to handle these issues and the load balancer would not need to assign these bad requests to servers. The cost of doing this is a more complex implementation of the load balancer and also losing some modularity as some of the functions of the httpserver are spilling over into the load balancer meaning it could be more difficult to debug.

Question3:

There is a difference in performance with the load balancer running 2 http servers performing the task in about half the time as the server that is running a server and one instance of nc.