Exercise 1

We used our personal laptops as the client (SAMSUNG) and server (DELL Inspiron) machines and connected them using standard Ethernet cables to a NETGEAR Router which we used as a switch. Other details about the setup -

- Max read bandwidth of the disk on Server ~ 65 MBps (found using disk1.c from Lab1 and ‘iotop’)
- Equivalent number of requests per second = ~ 32 req/sec (65 MBps / 2 MB/req)
- Maximum network bandwidth ~ 94 Mbps = 5.87 req/sec (94 Mbps * 1MB / 8Mb / 2 MB/req). This was found by running iperf.

In our network setup, the router was the bottleneck resource (with maximum capacity capped at ~100 Mbps). The cables used to connect to the router also have capacities of 100Mbps . The laptop network cards as well are capable of transferring more than 100 Mbps of data per second.

Exercise 2

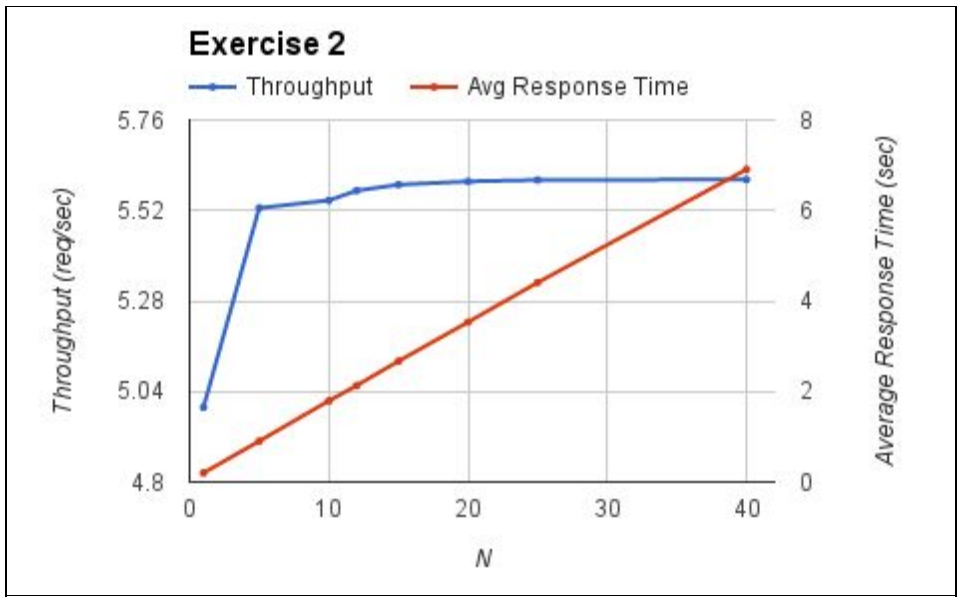
2a) Server saturates at N = 15 and we get highest throughput = 5.59 req/sec

2b) As N increases from 1, the throughput increases till it reaches saturation. After saturation, the throughput becomes constant (at the maximum possible throughput). The response time of the server increases for increasing values of N. Before the optimal value of N, the response time increases linearly.

2c) When the server operates as saturation, the bottleneck resource is the network. At saturation, the CPU utilization was very low and the hard disk utilization was about 20% (found using iotop). Hence, neither CPU, nor Hard Disk was the bottleneck resource.

As mentioned in Exercise 1, the maximum capacity of the network is ~5.87 req/sec and we are achieving 95% utilization. This maximum capacity was found using iperf.

2d) Server throughput at saturation is 5.59 req/sec. This is ~95% of the maximum possible network bandwidth as mentioned in 2c.

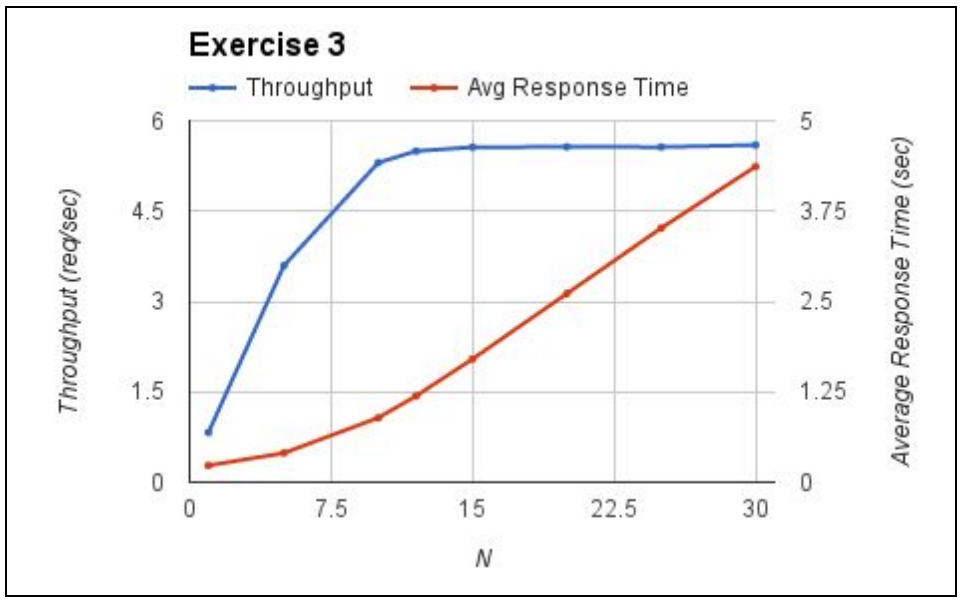


Exercise 3

3a) Server saturates at N = 25(Not so clear from the graph as y-axis scale is not small enough, evident from the data) and we get highest throughput = 5.56 req/sec.
3b-3d) Answers same as 2b-2d. The bottleneck resource is still the network.

The reason for a higher N to achieve saturation is due the sleep time. The fact that the client threads sleep means they send fewer requests to the server as compared to Q2, averaged over time.

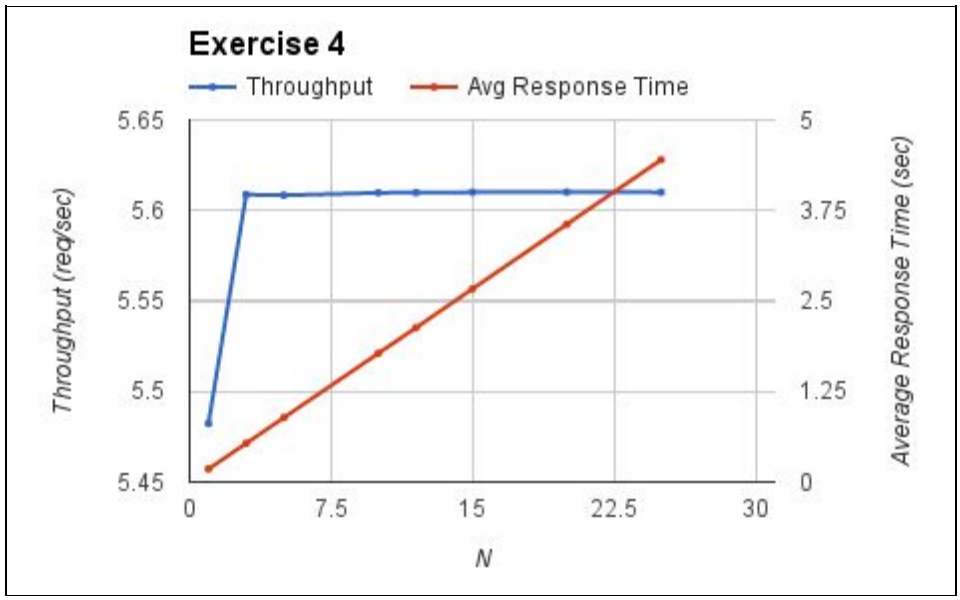
The response times also are lesser across all N values. This can be explained by the sleep time because a lower number of request are sent and hence, lower amount of work is pending at the server, compared to Q2.



Exercise 4

4a) Server saturates at N = 3 and we get highest throughput = 5.60 req/sec.
4b-4d) Answers same as 2b-2d. The bottleneck resource is still the network.

As only a fixed file is requested, the disk utilisation at the server is very low due to caching. This means that the server doesn't need to go to disk and hence, fetches the file very quickly. Therefore, even for a very small N a large throughput is observed and so we approach saturation even faster.



Notes:
If we had a more capable network setup or larger files then it can be conceived that in Q2,3 the Disk might turn out to be the bottleneck whereas in Q4 the network will most probably still be the bottleneck.