John Stiben Perez Brand – 225446
Matteo Romiti – 225907
Daniel Velez Duque – 225908

LAB 1

TASK 1

M/M/1 QUEUE

In this simulation, we compared the average time spent by a customer in the system considering the mathematical model, reported in Figure 1 with the label "*THEORETICAL*", and the measured average time obtained by our experiment. Observing the same figure, we can deduce that a proper value for the simulation time is 10^9 seconds, since a smaller simulation time such as 10^5 seconds would give erroneous results. The arrival rate λ spans between 0.01 and 0.09 and its inverse is used as mean for the generation of the random variable used in the arrival scheduling. The service time has a mean value equal to $1/\mu$, and it is set to 10 seconds. Hence, the load ρ ranges from 0.1 up to 0.9, interval in which the system is ergodic.

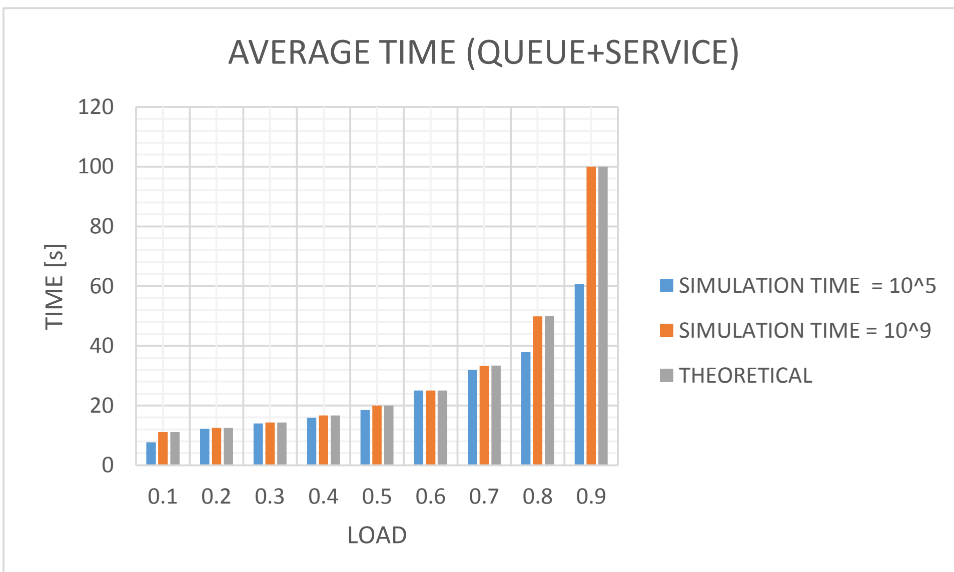


Figure 1

According to the analytical model, the probability of having a server idle decreases linearly with the increment of the load. In Figure 2, the expected behavior of the probability is obtained with a simulation time of 10^9 seconds.

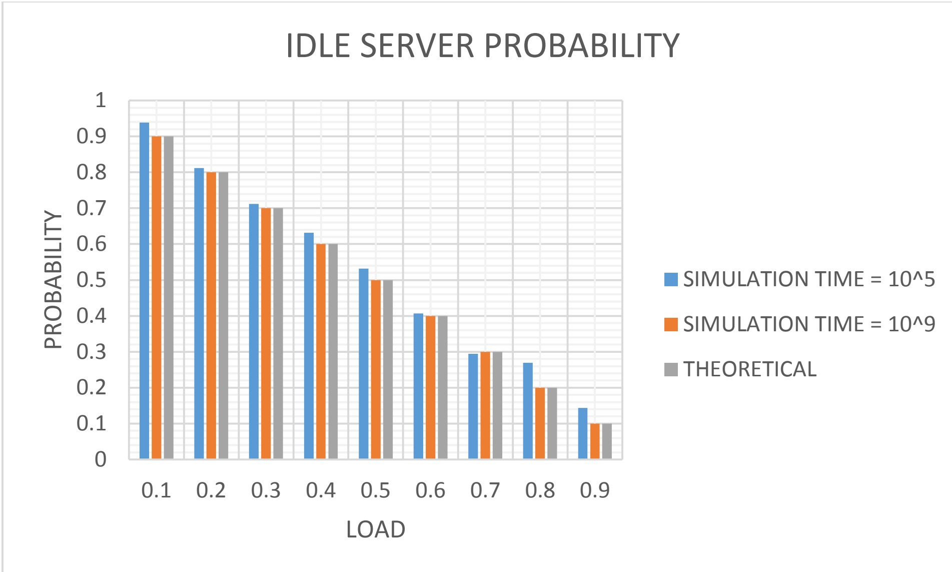


Figure 2

M/G/1 QUEUE

For this queue, we used a service time with a geometric distribution starting from 1, exploiting the function *geometric1*(*double mean, long *seed*). Also in this simulation, we set the argument *mean* equal to $1/\mu$, which is fixed to 10, while the arrival rate varies from 0.1 to 0.9. From now on, we will use a simulation time of 10^9 seconds. In figure 3, we compare the M/M/1 queue with the introduced M/G/1 queue. We also verify the trends of the measured average delay and average number of users with the P-K formula.

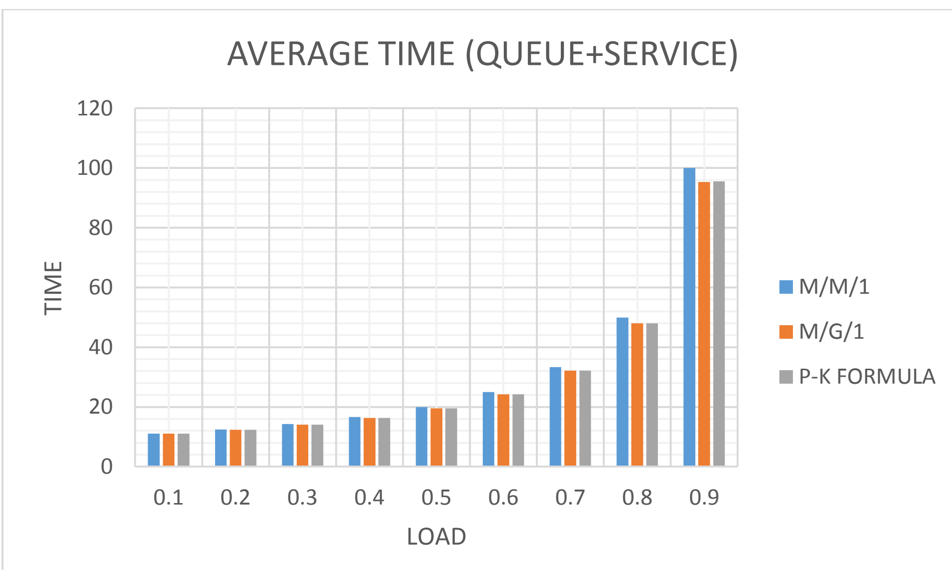
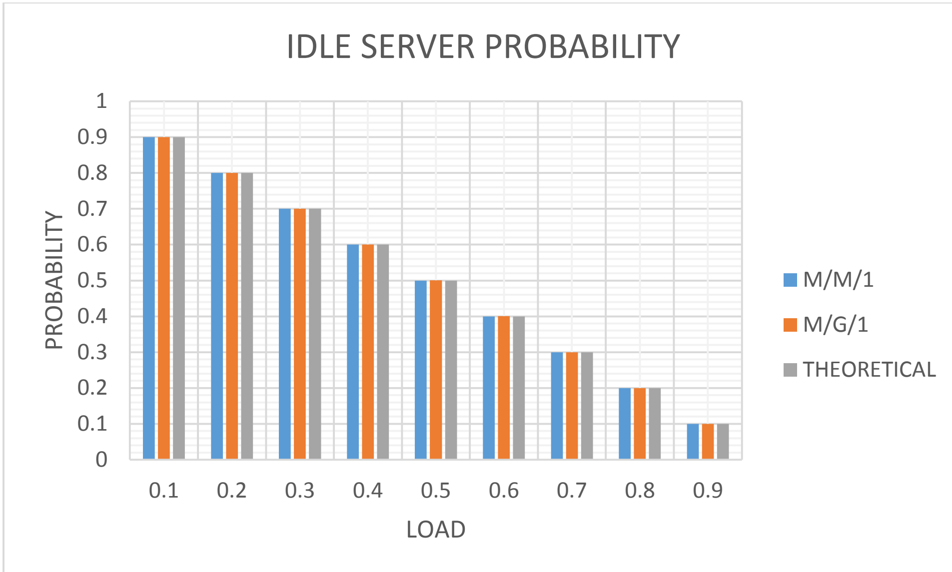
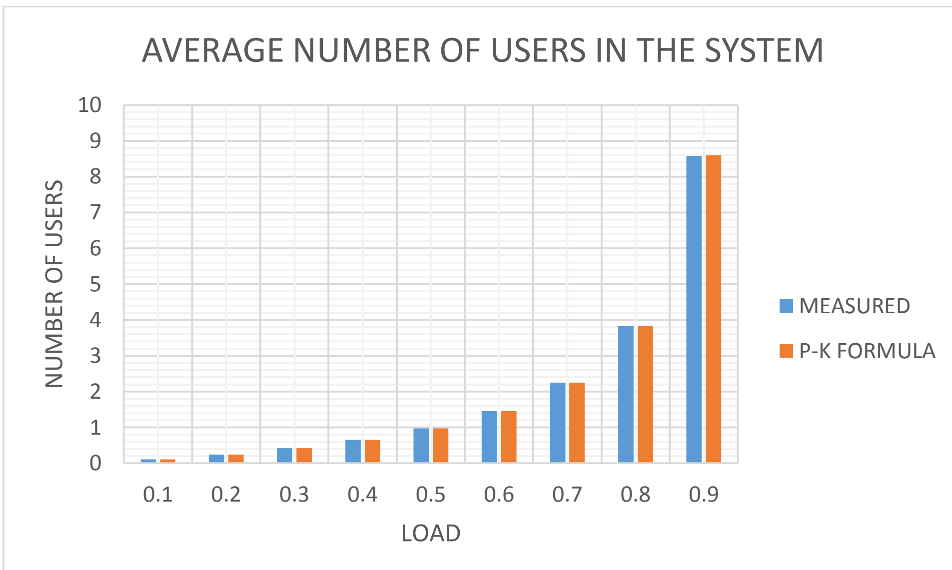


Figure 3





TASK 2

M/G/k/B QUEUE

For this scenario, we used three servers (k = 3) while the total number of users B in the system is set to 5, meaning that the queue has a size of 2. Also here, we modeled the service time with a geometric distribution starting from 1, but the range of the load is not limited to (0;1) because the system is always ergodic, since we have a limited queue. We decided to serve customers in the following way: first try to use the server 1, if it busy, try with server 2 and then server 3. This means that the workload distribution is not balanced, but, a part from the simpler structure, the main benefit is that the last server could be also turned off if long idle times occur and the energy consumption could be reduced. Looking at Figure 4, we note that the average delay is almost constant to 10 seconds for $\rho < 1$ and this is due to the presence of three servers. This value can be compared with the case of M/G/1 queue, where we had an average delay of 95.5 seconds for $\rho = 0.9$. As expected, we see in Figure 5 that the server 1 is the busiest, while the server 3 is the least loaded. The correct behavior of the system can be also verified in Figure 6, where an increment of the load implies a higher average of busy servers with an upper bound of 3. With the chosen values of queue size and number of servers, we obtain a loss probability that remains under 8% for $\rho < 2$, and is negligible for $\rho < 1$, as reported in Figure 7.

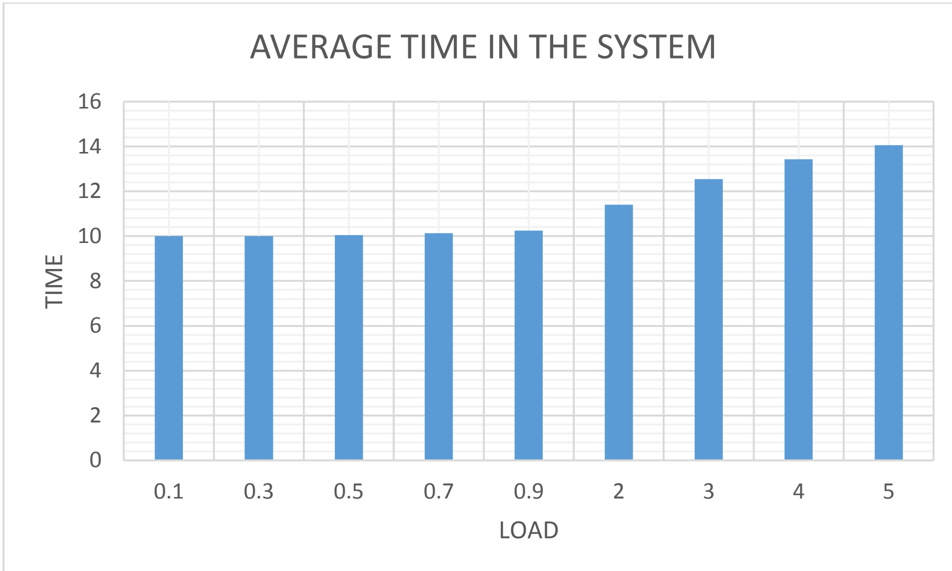


Figure 4

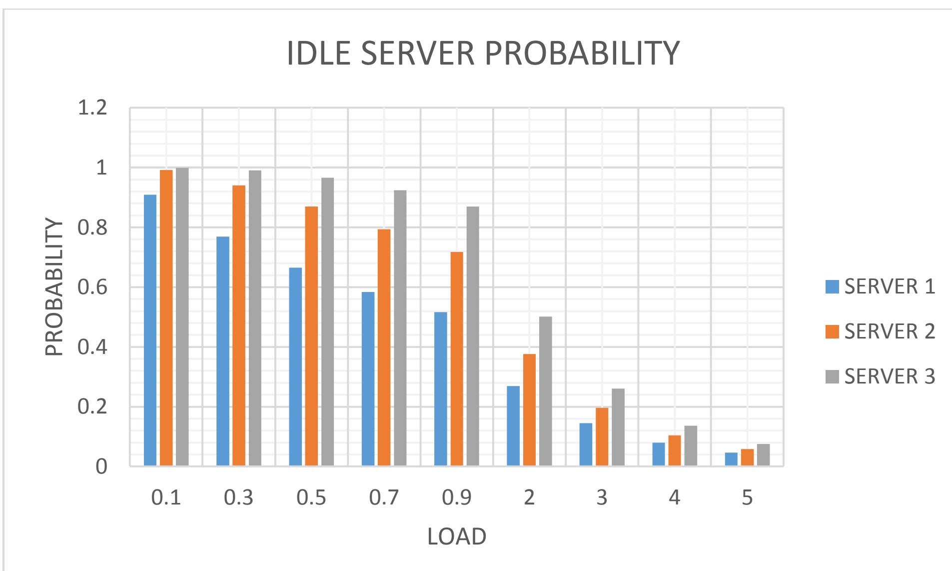


Figure 5

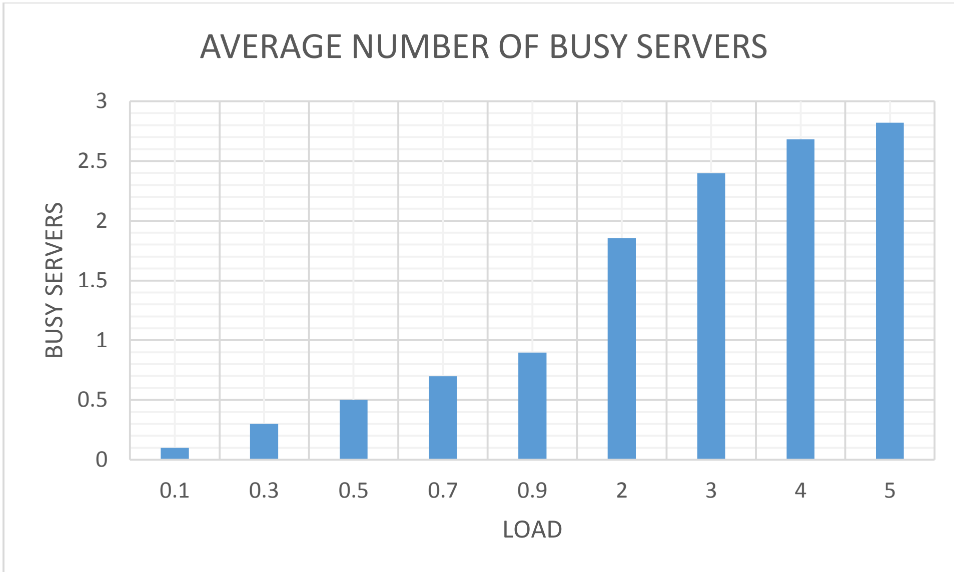


Figure 6

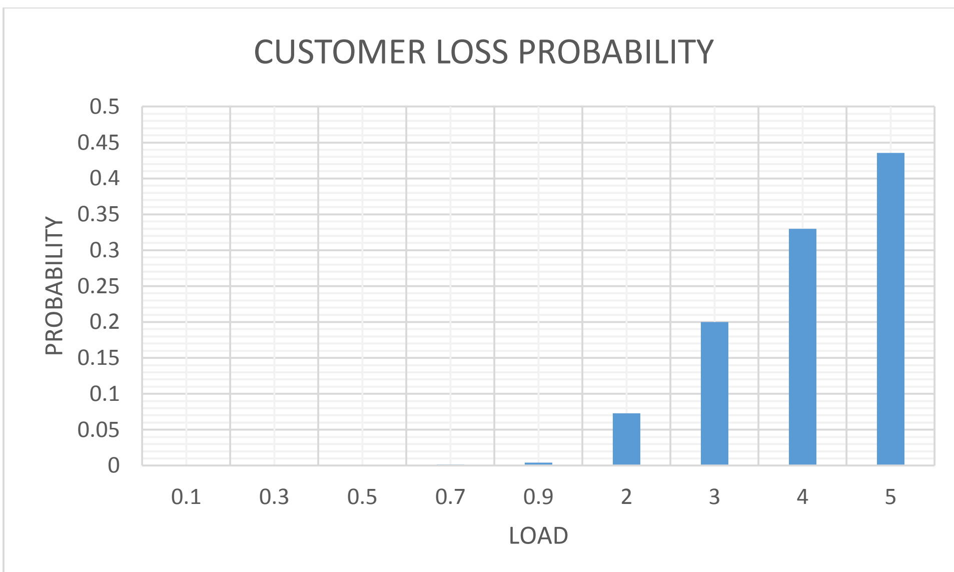


Figure 7