Network modelling: Theory and Simulation

Zoppi Samuele 202769
Ferro Demetrio 207872
Minetto Alex 211419

Laboratory 1

Task 1: The M/M/1 queue

Consider an M/M/1 queue.
Plot the performance of the queue versus load in terms of:

- Average queuing delay
- Probability that the server is idle (i.e., fraction of time that the server is idle)

Verify Little's law. Compare simulation and analytical results.

M/M/1 Queue

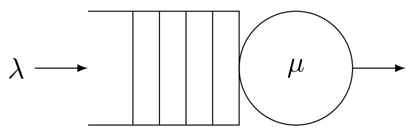


Figure 1: Model of an M/M/1 Queue

The queueing system is modeled as follows:

- Arrivals are Poisson distributed with rate λ.
- Service times are Exponentially distributed with parameter μ.

The stochastic behaviour can be modeled as a Continuous Time Markov Chain, whose state space refers to the number of customers in the system.

The chain has an infinite number of states, but it has a stationary distribution, hence preserves ergodicity, just for $\lambda < \mu$. The main indices of the system are the following:

- $\mathbb{E}[L]$, Average number of customers in the system: from the stationary distribution, it can be easily computed.
- $\mathbb{E}[T_s]$, Average serving time: since service times are Exp(μ), it is always equal to $\frac{1}{\mu}$.
- $\mathbb{E}[T]$, Average time spent in the queue: it can be computed with Little's Result: $\mathbb{E}[T] = \frac{\mathbb{E}[L]}{\lambda}$
- $\mathbb{E}[T_w]$, Average waiting time: it can be computed by difference: $\mathbb{E}[T_w] = \mathbb{E}[T] - \mathbb{E}[T_s]$.

Once we considered all the theoretical issues, we proceeded with simulations, by computing:

Average queueing delay

In order to get to a global overview of the results when the load factor ρ, vary in $]0,1[$, we wrote a bash script which executes the provided C program for the simulation for different values of ρ.

We fixed $\mu = 1$ and let λ, hence ρ, vary in $]0,1[$.

A complete simulation consisted in letting λ vary from 0.001 to 0.999 with step 0.001 whose output will report here in plots.

In Figure 2 is reported an overall view of the comparison between theoretical and empirical Times spent in the queue.

The average time spent Serving is supposed to be constant and equal to $\mathbb{E}[T_s] = \frac{1}{\mu} = 1$, whereas the time spent waiting $\mathbb{E}[T_w]$ and the total time $\mathbb{E}[T]$ are obtained from the simulation and compared to the theoretical ones.

As we can see, the theoretical results are very likely until the condition $\rho < 1$ is satisfied.

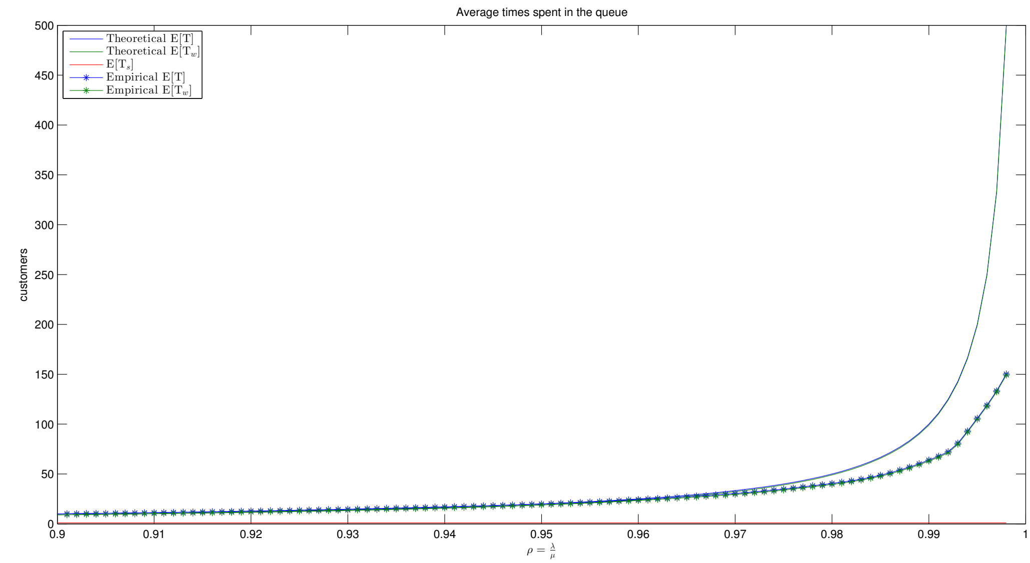


Figure 2: Comparison of theoretical and empirical $\mathbb{E}[T], \mathbb{E}[T_s], \mathbb{E}[T_w]$.

Probability that the server is idle

In order to get such a probability, we computed the ratio between the amount of time in which the server is idle and the total time needed to serve all the customers arrived at the queue during the time duration of a simulation.

In order to do that, we let a counter start when a departure leaves the queue empty, and stopped it when the soon after arrival occurs.
From Figure 3 we can see that the empirical results match exactly the theoretical ones.

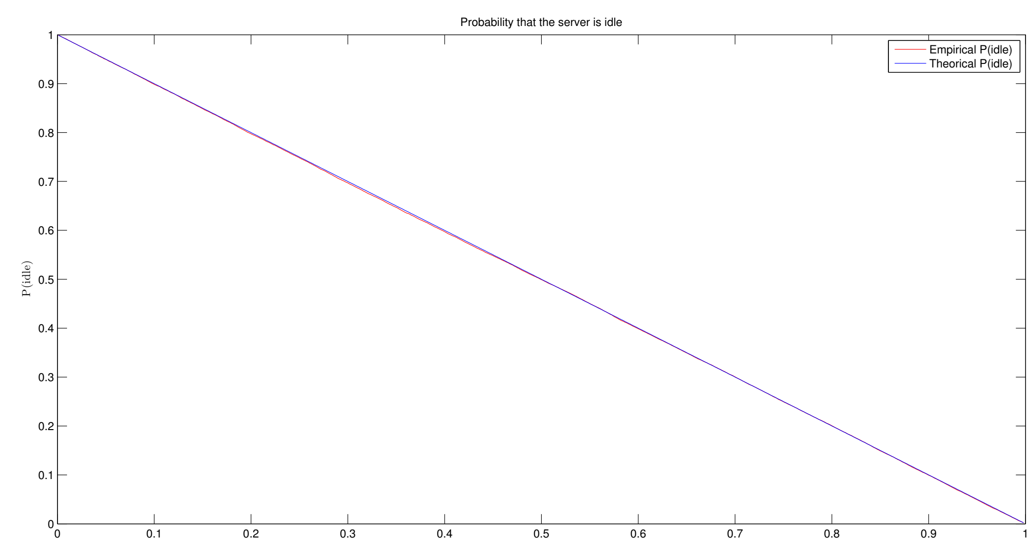


Figure 3: Comparison between the theoretical and the empirical probabilities that the server is idle.

As we expected, the probability that the server is idle, decays as the load factor grows. At fixed average of the service times, the more arrivals we have, the less probable is that the server is idle.

In Figure 4, we have considered both the theoretical and the Empirical average number of customers, computed according to Little's Result: $\mathbb{E}[L] = \lambda \cdot \mathbb{E}[T]$.

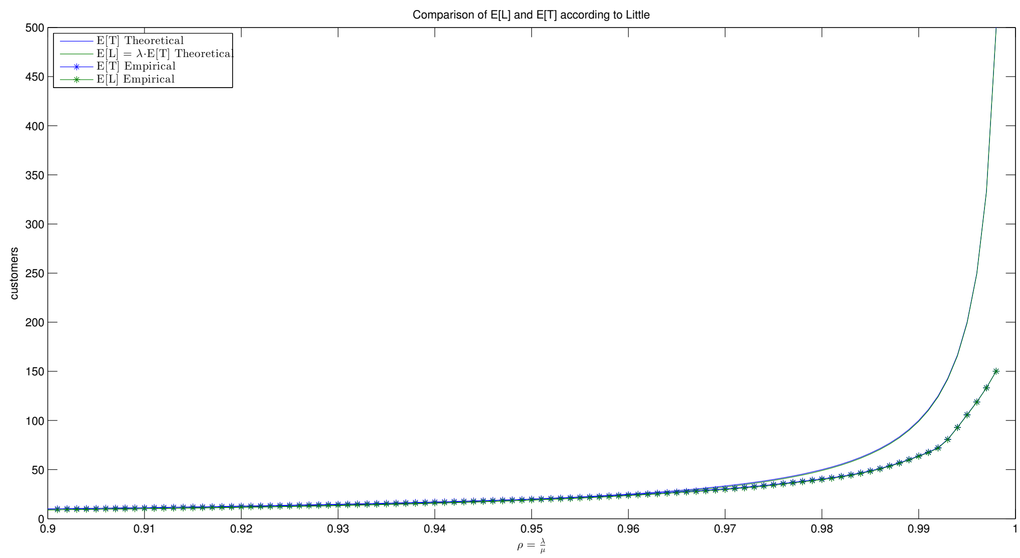


Figure 4: theoretical and Empirical comparison of the number of customers according to Little's result.

Task 2: Effect of the service time distribution

Modify the simulator of the M/M/1 queue and study the performance of an M/H2/1 queue and an M/E2/1 queue.

· Compare the performance of the three considered queues, for the same values of load

· Verify the Pollaczeck-Khinchine formula

M/H$_2$/1 Queue

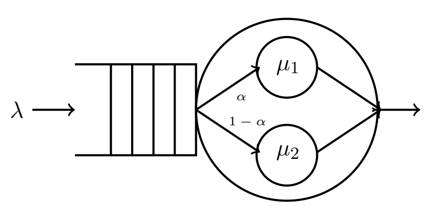


Figure 5: Model of an M/H$_2$/1 Queue

An M/H$_2$/1 queue has a different distribution of the Service times. In order to schedule the departure of customers, we have to simulate an Hyperexponential random variable.

A Hyperexponential(2) random variable T_s has the following mean: $\mathbb{E}[T_s] = \dfrac{\alpha}{\mu_1} + \dfrac{1-\alpha}{\mu_2}$.

Hence the load factor of an M/H$_2$/1 queue is equal to: $\rho = \dfrac{\alpha \cdot \lambda}{\mu_1} + \dfrac{\lambda \cdot (1-\alpha)}{\mu_2}$

M/E_2/1 Queue

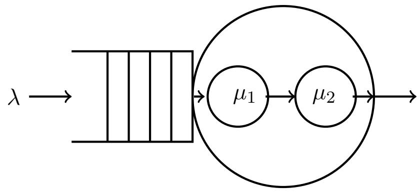


Figure 6: Model of an M/E_2/1 Queue

This time the Service times have a different distribution. By supposing that the service is composed of two phases, it's easy to see that the amount of time spent in service is the sum of the time spent in the two phases. In order to schedule the departure of customers, we have to simulate Erlang distribution made of two contributions (E_2).

An Erlang(2) random variable T_s has the following mean: $\mathbb{E}[T_s] = \dfrac{1}{\mu_1} + \dfrac{1}{\mu_2}$, and load factor: $\rho = \dfrac{\lambda}{\mu_1} + \dfrac{\lambda}{\mu_2} = \dfrac{\lambda}{\mu}$.

In order to evaluate the performances of the 3 queues, we compared the results for:

· M/M/1 Queue with Exponential service times, with parameter $\mu = 1$.

· M/H_2/1 Queue with Hyperexponential(2) service times, with $\mu_1 = 1, \mu_2 = 1, \alpha = \frac{1}{2}$.

· M/E_2/1 Queue with Erlang(2) service times: with $\mu_1 = 2, \mu_2 = 2$.

By letting hence ρ, vary in $]0, 1[$, we wanted to evaluate the performances of the three queueing systems, for the same values of load.

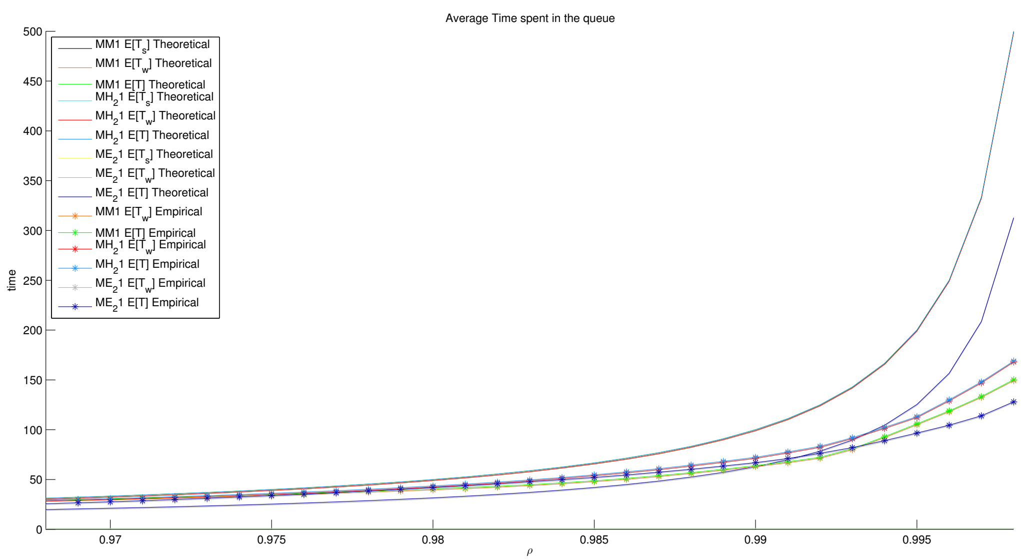


Figure 7: Comparison of theoretical and empirical $\mathbb{E}[T_s], \mathbb{E}[T_w], \mathbb{E}[T]$ in the three queueing systems.

In Figure 7 and Figure 8 we reported the behaviour of the three queues, both empirical and theoretical data.

From the comparison of the results, we could see that in terms of average number of users and average time spent in the queue, the M/E_2/1 has better performances than the M/M/1, which still has better performances than the M/H_2/1.

From the Pollanczeck-Khinchine, we shall consider:

$$\mathbb{E}[L] = \rho + \frac{\rho^2(1+c_s^2)}{2(1-\rho)}, \qquad \mathbb{E}[T] = \frac{\mathbb{E}[L]}{\lambda} = \mu + \frac{\rho(1+c_s^2)}{2\mu(1-\rho)} \quad \text{where } c_s = \frac{\sigma_s}{\mathbb{E}[S]}$$

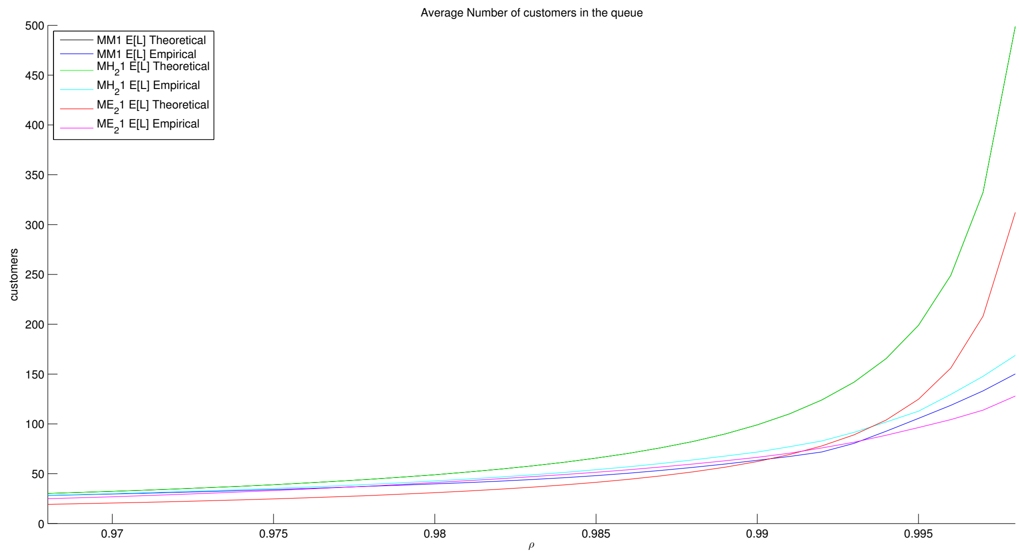
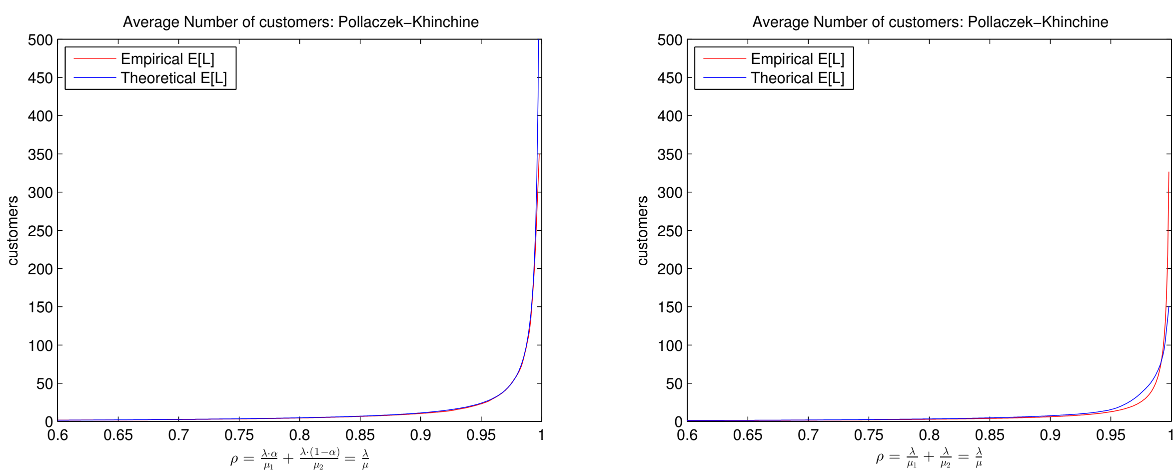


Figure 8: Comparison of theoretical and empirical $\mathbb{E}[L]$ in the three queueing systems.

$$c_s^2 = \frac{\sigma^2}{\mathbb{E}[S]^2} = \frac{\frac{2\alpha}{\mu_1^2} + \frac{2(1-\alpha)}{\mu_2^2} - \left(\frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2}\right)^2}{\left(\frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2}\right)^2} = 2\frac{\alpha\mu_2^2 + (1-\alpha)\mu_1^2}{(\alpha\mu_2 + (1-\alpha)\mu_1)^2} - 1 \qquad \text{for M/H}_2\text{/1},$$

$$c_s^2 = \frac{\sigma^2}{\mathbb{E}[S]^2} = \frac{\frac{1}{\mu_1^2} + \frac{1}{\mu_2^2}}{\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right)^2} = \frac{\mu_1^2 + \mu_2^2}{(\mu_1 + \mu_2)^2} = 1 - 2\frac{\mu_1\mu_2}{(\mu_1 + \mu_2)^2} \qquad \text{for M/E}_2\text{/1}.$$



(a) M/H$_2$/1 queue: $\mathbb{E}[L], \alpha = \frac{1}{3}, \mu_1 = \frac{2}{3}, \mu_2 = \frac{4}{3}$.

(b) M/E$_2$/1 queue: $\mathbb{E}[L], \alpha = \frac{1}{3}, \mu_1 = \frac{2}{3}, \mu_2 = \frac{4}{3}$.

Figure 9: Comparison between theoretical and empirical $\mathbb{E}[L]$ in both the queues.

As we can see from Figure 9a and Figure 9b, the Pollaczek-Khinchine formula is an almost accurate way to compute the average number of users in the queue.

Task 3: The M/G/k queue

Modify the simulator of the single server queue to study the performance of an M/G/k queue. Plot the performance of the queue versus load computing also the average number of busy servers.

M/G/k Queue

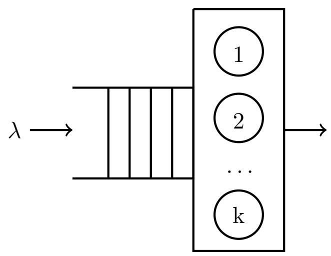


Figure 10: Model of an M/G/k Queue

We modified the provided simulator, so that the number of servers was k, and when a new arrival occurs, the customer is served if at least one of the servers is idle.
The service times may be distributed in a general way:

$$\begin{array}{lll} \text{Markovian:} & T_s \sim \text{Exp}(\mu), & \mathbb{E}[T_s] = \dfrac{1}{\mu} \\ \text{Geometric:} & T_s \sim \text{Geo}(p) & \mathbb{E}[T_s] = \dfrac{1}{p} \\ \text{Deterministic:} & T_s = t_s, & \mathbb{E}[T_s] = t_s \end{array}$$

In order to get equal load factors, we fixed $k = 5$, and: $\mathbb{E}[T_s] = \frac{1}{\mu} = \frac{1}{p} = t_s = 2$, so we set: $\mu = \dfrac{1}{2}, p = \dfrac{1}{2}, t_s = 2$.

This time $\rho = \dfrac{\lambda}{k\mu}$, so we let λ vary in $]0, \frac{k}{2}[$, hence $\rho = \lambda \cdot \dfrac{\mathbb{E}[T_s]}{k} = \lambda \dfrac{2}{k}$ vary in $]0, 1[$.

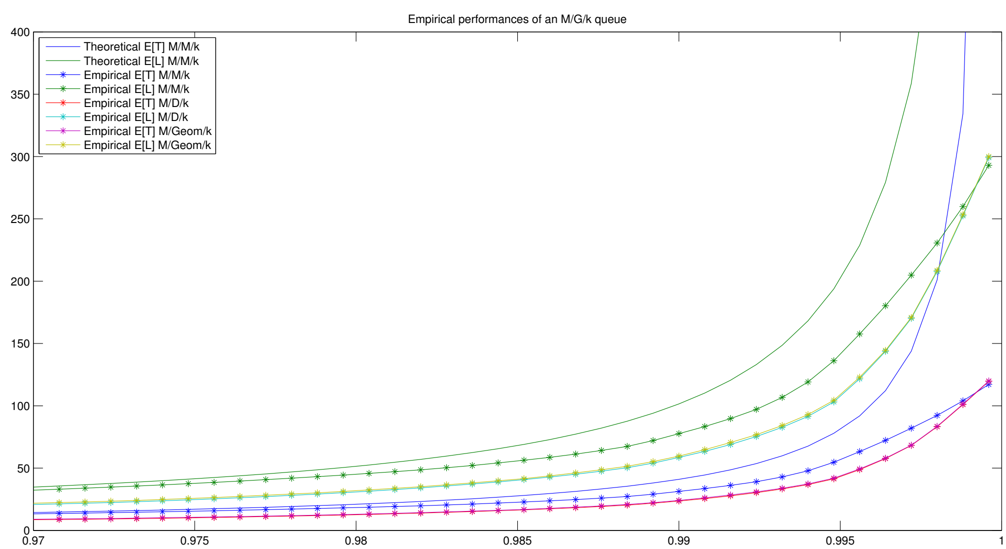


Figure 11: Comparison of theoretical and empirical $\mathbb{E}[L]$ considering three M/G/5 queuing systems.

For the computation of the average number of busy serves in the case of exponential service times, we considered the stationary probability of the CTMC associated to the M/M/k queuing system. By denoting with π_m the probability of having m customers in the queue, we get to:

$$\pi_m = \begin{cases} \pi_0 \left(\dfrac{\lambda}{\mu}\right)^m \dfrac{1}{m!} & \text{for } m \leq k \\ \pi_0 \left(\dfrac{\lambda}{\mu}\right)^m \dfrac{1}{k!\ k^{(m-k)}} & \text{for } m \geq k \end{cases} \quad \text{where } \pi_0 = \dfrac{1}{\displaystyle\sum_{m=0}^{k-1} \dfrac{(k\rho)^m}{m!} + \dfrac{(k\rho)^k}{k!} \dfrac{1}{(1-\rho)}}$$

Letting M be the number of busy servers, we may compute the average number of busy servers $\mathbb{E}[\mathrm{M}]$:

$$\mathbb{E}[\mathrm{M}] = \sum_{m=0}^{k-1} m\pi_m + k\sum_{m=k}^{\infty}\pi_m = \sum_{m=0}^{k-1}\frac{m\pi_0(k\rho)^m}{m!} + \frac{k\pi_0(k\rho)^k}{k!(1-\rho)} = \cdots = k\rho = \frac{\lambda}{\mu}$$

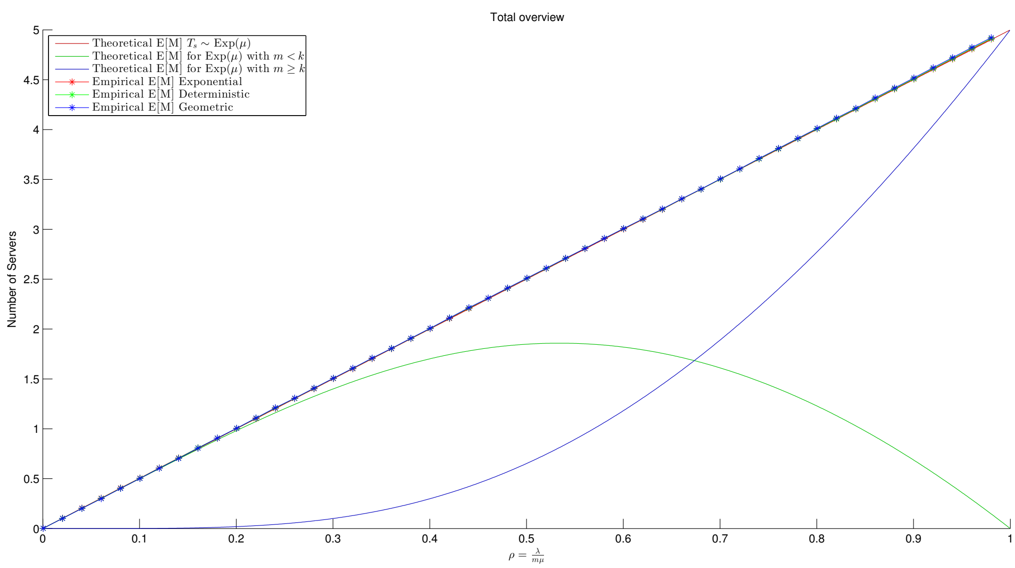


Figure 12: Comparison of the theoretical $\mathbb{E}[M]$ for M/M/5, with empirical $\mathbb{E}[M]$ of M/G/5 queues.

As we can see from the plot, even for Deterministic and Geometric cases: $\mathbb{E}[\mathrm{M}] = \lambda\mathbb{E}[T_s]$.

Task 4: Finite queueing line

Modify the simulator of the M/G/k queue to study the performance of an M/G/k/B queue.
Plot the performance of the queue versus load computing also the loss probability.
In the special case B=0 show the insensitivity of the results with respect to the service time distribution.

M/G/k/B Queue

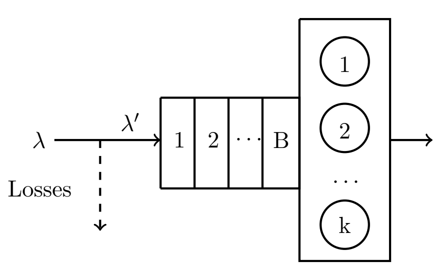


Figure 13: Model of an M/G/k/B Queue

We have the same configuration we had above, but a limited size of the buffer to B.
First of all, we computed the Loss Probability.
By considering the queue without losses, we computed the probability of having $(k + B)$ customers in it.

$$\mathrm{P(loss)} = \pi_{(k+B)} = \frac{k^k\pi_0}{k!}\rho^{(k+B)}, \quad \rho = \frac{\lambda}{k\mu}$$

$$\pi_0 = \frac{1}{\displaystyle\sum_{m=0}^{k-1}\frac{1}{m!}(k\rho)^m + \frac{k^k}{k!}\sum_{m=k}^{k+B}\rho^m}$$

Once we had the loss probability, we could consider the rate of arrivals (without losses)

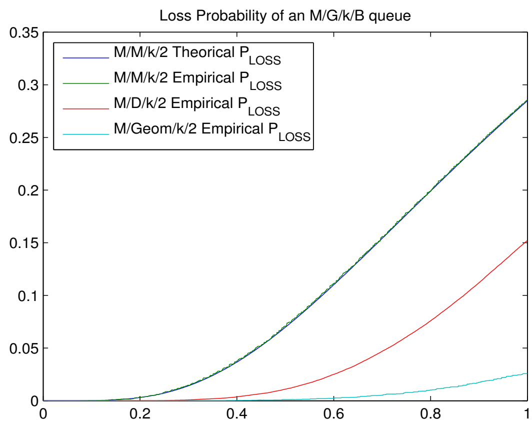
$$\lambda' = \lambda \cdot (1 - \mathrm{P(loss)})$$

. With the new arrival rate, we computed the average time spent waiting:

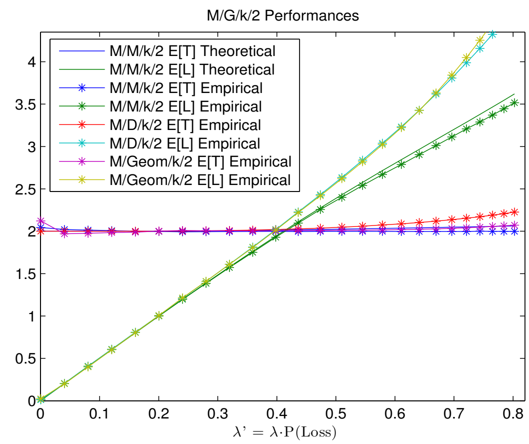
$$\mathbb{E}[T_w] = \sum_{m=k}^{k+B} \pi_m \frac{(m-k+1)}{k\mu} = \sum_{m=k}^{k+B} \pi_k \left(\frac{\lambda'}{\mu}\right)^{(m-k)} \frac{(m-k+1)}{k\mu} = \frac{\pi_k}{k\mu} \sum_{j=1}^{B+1} j \left(\frac{\lambda'}{\mu}\right)^{(j-1)}$$

$$\mathbb{E}[T] = \mathbb{E}[T_w] + \mathbb{E}[T_s], \qquad \mathbb{E}[L] = \lambda' \mathbb{E}[T]$$

.



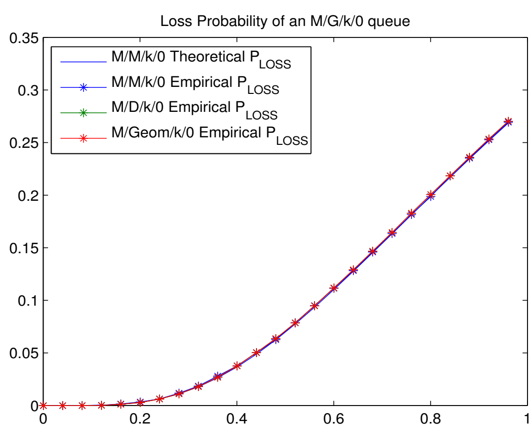
(a) Loss Probability.



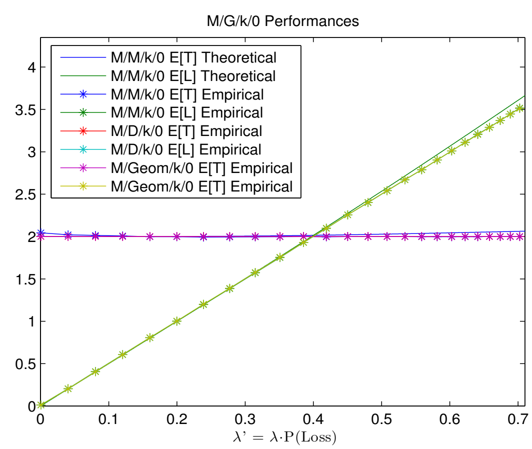
(b) Performances in terms of $\mathbb{E}[T]$ and $\mathbb{E}[L]$.

Figure 14: Loss Probability and Performances of M/M/5/2, M/D/5/2 and M/Geom/5/2 queues.

In the case B=0 we have that, according to Erlang-B formula: $\mathrm{P(loss)} = \frac{1}{k!}(k\rho)^k \left[\sum_{m=0}^{k} \frac{(k\rho)^m}{m!}\right]^{-1}$.



(a) Loss Probability.



(b) Performances in terms of $\mathbb{E}[T]$ and $\mathbb{E}[L]$.

Figure 15: Loss Probability and Performances of M/M/5/0, M/D/5/0 and M/Geom/5/0 queues.

Laboratory 2

Task 1: Two M/M/1 queues

Consider two M/M/1 queues in sequence. Implement the simulator to derive the steady state behavior of the system. Verify the product form solution.

In order to implement the queueing network we modified the simulator, scheduling an arrival in the second queue at every departure from the first queue. The record of one costumer coming from exogenous arrivals is created when it arrives in the first queue, it's moved from one queue to another and it's freed only when it leaves the second queue, leaving the network.
We measured the time in which the network was in a specific (k_1, k_2) condition and we divided by the total time to get the steady state probabilities π_{k_1,k_2}. Finally we compared these empirical results with the theoretical product form solution for a network of two M/M/1 queues:

$$\pi_N = \pi_{k_1,k_2} = \pi^{(1)}_{k_1} \cdot \pi^{(2)}_{k_2} = (1-\rho_1)\rho_1^{k_1} \cdot (1-\rho_2)\rho_2^{k_2}$$

In order to verify the product form solution we used $\mu_1 = 1$ $\mu_2 = 1$ and let λ vary in $]0,1[$ so that the two M/M/1 had the same $\rho = \rho_1 = \rho_2$, varying in $]0,1[$.

The product form solution reduces to:

$$\pi_N = \pi_{k_1,k_2} = (1-\rho)^2 \rho^{k_1+k_2}$$

The computation of the probability density functions was implemented in the simulator, and in the following figures (Figure 16, Figure 17) the empirical results are compared to the theoretical ones.

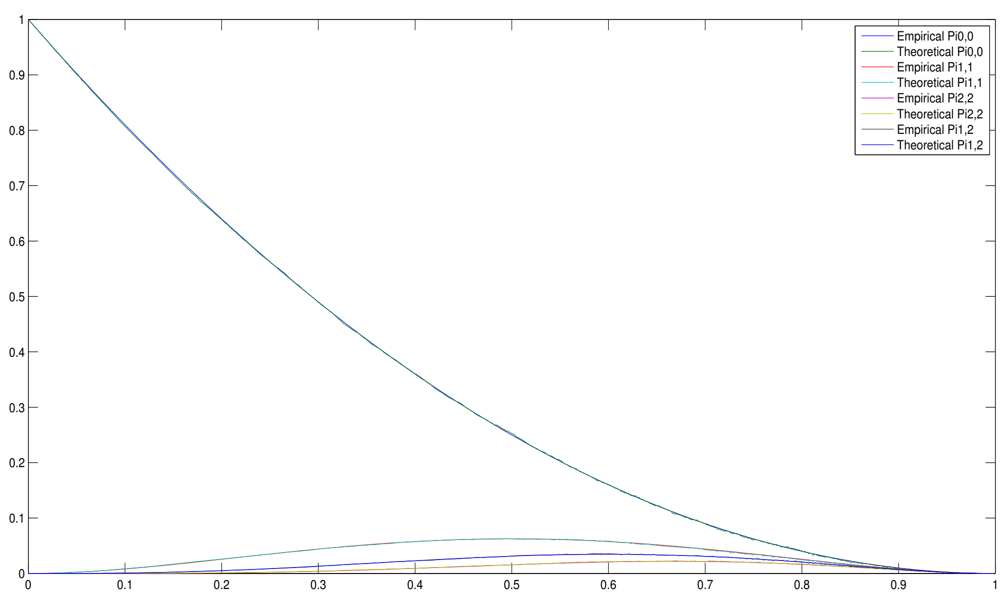
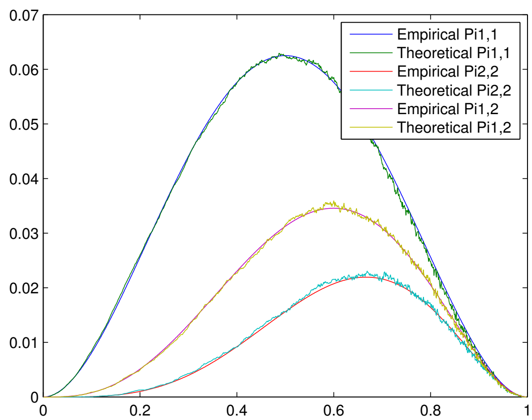
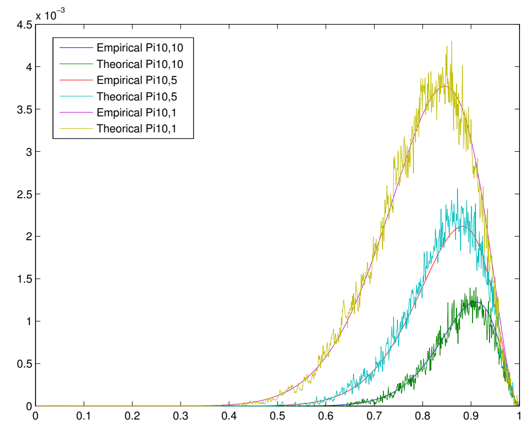


Figure 16: Steady state probabilities for $\pi_{0,0}, \pi_{1,1}, \pi_{2,2}$.



(a) Steady state probabilities for $\pi_{1,1}, \pi_{2,2}, \pi_{1,2}$.



(b) Steady state probabilities for $\pi_{0,0}, \pi_{1,1}, \pi_{2,2}$.

Figure 17: Steady state probabilities π_{k_1,k_2}: theoretical and empirical.

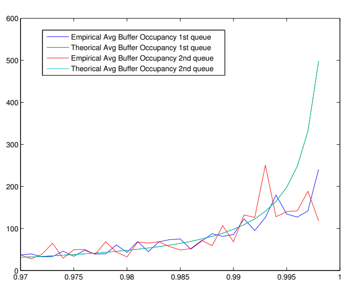
Task 2: Effect of the service time distribution and first moment

- Consider the case of two queues with different service rate. Evaluate and compare the average buffer occupancy of the two queues when $\mu_1 > \mu_2$ and when $\mu_1 \leq \mu_2$.
- Consider the case of deterministic service times with $\mu_1 \leq \mu_2$, evaluate and compare the average buffer occupancy and the buffer occupancy distribution for the two queues.

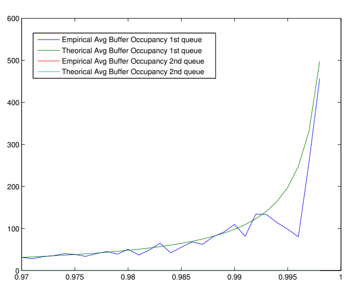
We computed the average buffer occupancies adding counters in the code able to track the behavior of the buffer occupancies over time.
We compared the results of the simulator with the following theoretical formula for the buffer occupancy of an M/M/1 queue:

$$\mathbb{E}[\text{buffer occupancy}] = \mathbb{E}[L_w] = \frac{\rho^2}{(1-\rho)}$$

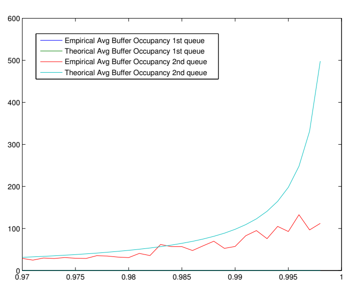
In Figure 18 are reported the plots of the comparison between the ideal M/M/1 and the actual queueing network.



(a) Average buffer occupancy when $\mu_1 = 1$ and $\mu_2 = 1$.



(b) Average buffer occupancy when $\mu_1 = 1$ and $\mu_2 = 10$.



(c) Average buffer occupancy when $\mu_1 = 10$ and $\mu_2 = 1$.

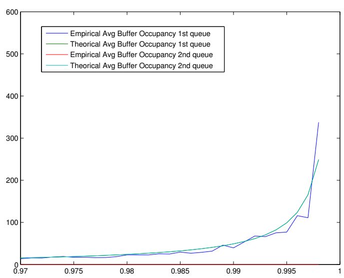
Figure 18: Average buffer occupancy for different values of μ_1, μ_2.

Then we considered the case of deterministic service times:

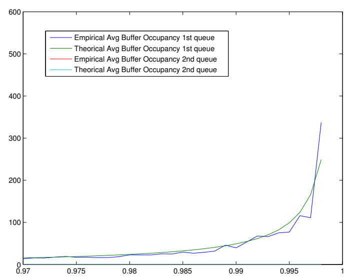
In the case where $\mu_1 = 1$, $\mu_2 = 1$, the average buffer occupancy of the second queue when the servers have the same constant rate $T_s = t_s$ is always equal to 0, because the out-coming flow of costumers from the first queue is no longer Poisson. The two queues are synchronized, this leads indeed to a 0 occupancy buffer in the second queue.

In the case where $\mu_1 = 1$, $\mu_2 = 10$ we have a similar behavior, the first queue is way too slow with respect to the second one, so it's very likely that the latter is going to be free when a new customer arrives from the former.

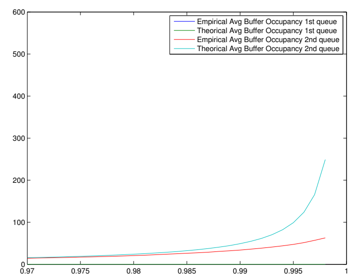
In the case where $\mu_1 = 10$, $\mu_2 = 1$ the second queue is slower to serve its customers, its buffer occupancy is going to be higher than the first one.



(a) Average buffer occupancy when $\mu_1 = 1$ and $\mu_2 = 1$.



(b) Average buffer occupancy when $\mu_1 = 1$ and $\mu_2 = 10$.



(c) Average buffer occupancy when $\mu_1 = 10$ and $\mu_2 = 1$.

Finally we tried to determine the behavior of the buffer occupancy distribution for the two queues. In order to do it we implemented in the simulator the histogram of the buffer occupancy at different values of ρ. The result is an exponential-like curve:

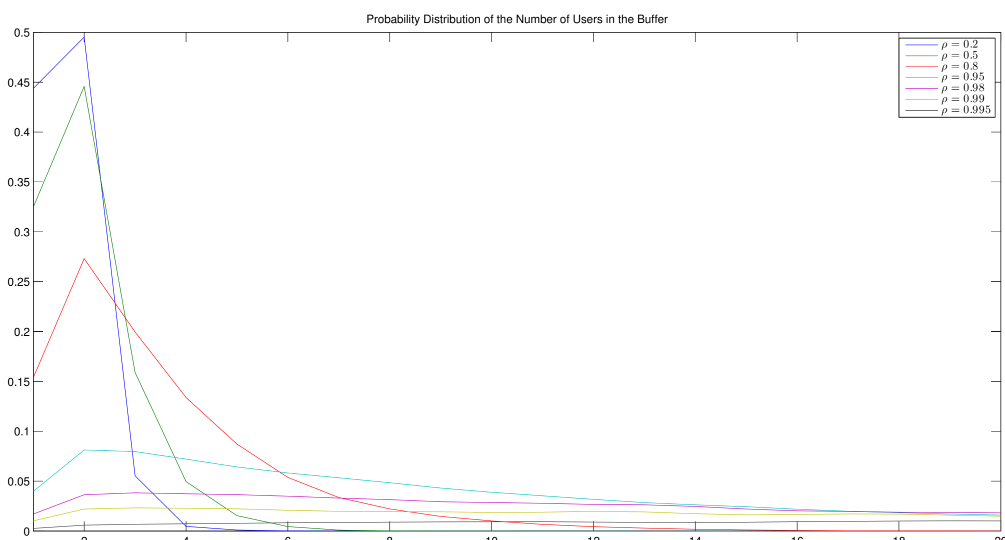
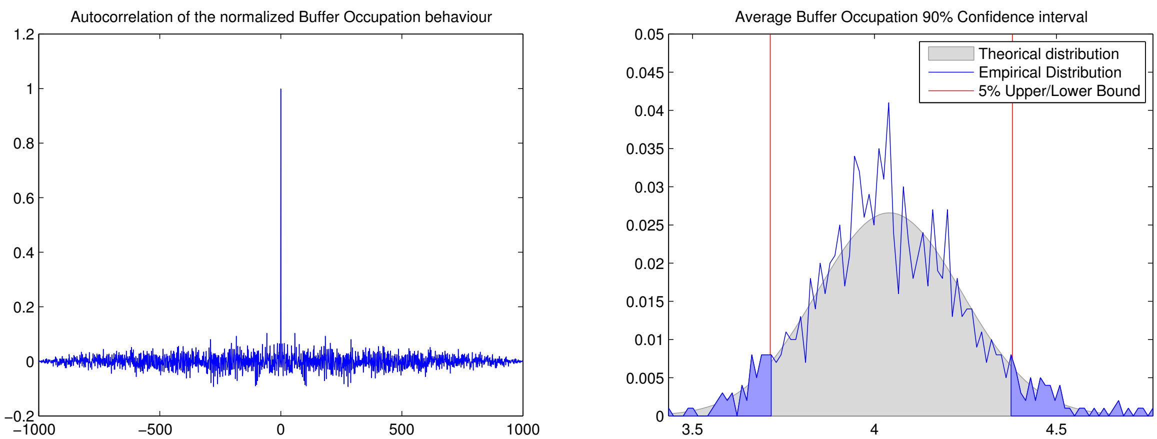


Figure 20: Probability distribution of the number of users in the buffer.

We calculated the Confidence Interval for the average buffer occupancy of the first queue, in the case of two M/D/1 queues in series, for a given value of $\rho = 0.9$.
First we had to solve the problem of independence of seeds for the random values generation. We used a different seed for any simulation, starting from last seed of the previous simulation.
Doing this we could generate independent simulations because the different seeds have a big spacing in the same period. We can observe the independence of our samples from the autocorrelation shown in Figure 21a.
After collecting a big number of independent samples for a given ρ we calculated the average $\hat{X}$, the variance σ_X^2 and, according to a certain level of confidence (90%), the confidence interval.
The results of our analysis are shown in Figure 21b where the *pdf* of our samples is compared to the $N(\hat{X},\sigma_X^2)$.



(a) Autocorrelation of the sequence used for the simulation.

(b) Comparison of the *pdf* of our seeds and the $N(\hat{X},\sigma_X^2)$

In the end, in Figure 22 we represented the sequence of samples with their mean, standard deviation and confidence interval.

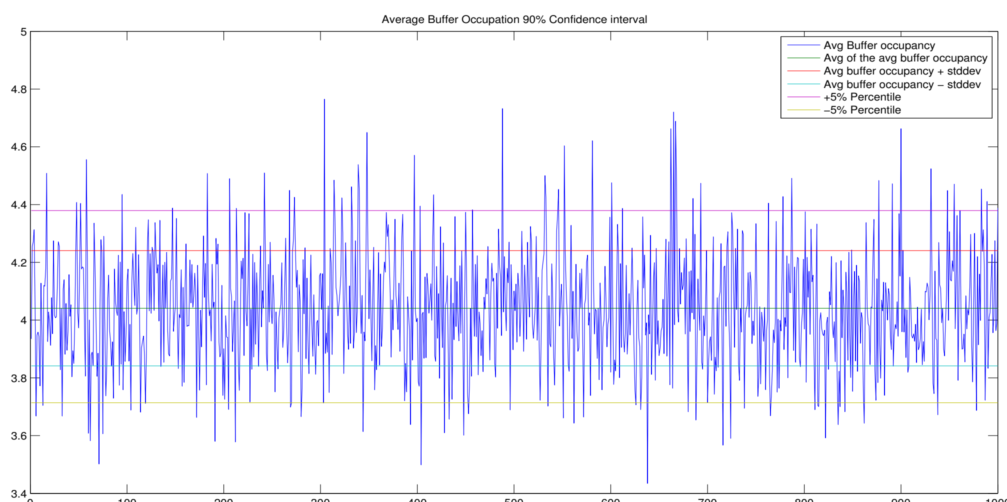


Figure 22: Sequence of samples with their mean, standard deviation and confidence interval.