Network modeling: theory and simulation Report for Laboratory 1 and 2

Rico Mendrzik, Samuele Racca, Samuele Maci
Politecnico di Torino

rico.mendrzik@tuhh.de, samuele.racca@studenti.polito.it, samuele.maci@studenti.polito.it

Laboratory 1

***Abstract*—In this report the results of the theory that has previously been derived in the above-mentioned lecture and the results of the simulation approach will be compared according to the exercise sheets provided. Starting with a basic M/M/1 queue, we will show the behavior of the queue for various different offered traffics. Furthermore we will show the influence of different service time distributions on the performance of the queue. The mentioned service time distributions are the Markovian, Erlang and Hyperexponential, respectively. Afterwards we will generalize the approach by introducing general service time distributions and multiple servers. The performance will be analyzed for the special case of a deterministic service time distribution. At last, the effect of a finite waiting line on the performance of the queue will be depicted.**

***Index Terms*—Queueing theory, single server, multiple server, General service time, finite waiting line, performance evaluation**

I. Introduction

Simulation is a powerful technique to evaluate the performance of a system. Especially in the field of queueing theory it is a vital means to handle systems, analyze their performance and verify results that have been derived analytically.

Unfortunately, many problems cannot be solved analytically or the complexity for the calculation of the solution grows rapidly and makes the analytic approach unattractive to perform. In these cases simulation can help to achieve the desired goal.

Furthermore, simulation can be used to cross-validate analytically obtained results. In case of a correct implementation of the simulation and assuming that the analytic results are true, both results must coincide - neglecting errors that result from numerical issues and the effect of a finite simulation time. Despite that, if the analytic result contains error(s), the results will in general be different from the simulated result since it is highly unlikely to make the same mistake in both the analytic and simulation approach, respectively.

These examples emphasize that simulation is a mighty tool that can be used for different purposes.

Eventually, some very important issues regarding simulation difficulties must be mentioned. The simulation is a sample of a stochastic process. In order to reliably calculation the quantities of interest (e.g. average queueing delay) the simulation must be performed sufficiently long. This has several reasons. One reason is to minimize the effect of the transients. Another reason is to deal with the stochastic variations. The number of events should be bigger by some orders of magnitude than the average number of customers in the system.

II. Task I

We are going to compare the theoretic results for the M/M/1 queue with the results generated by our simulation. The quantities of interest are the average queueing delay, the idle probability and the average number of customers in the system in order to verify Little's result by simulation.

A. Average Queueing Delay

For the first task we have simulated the behavior of a M/M/1 queue depending on the offered traffic $\rho = \frac{\lambda}{m\mu}$ [1]. λ represents the arrival rate, μ the service rate and m the number of servers. ρ ranges theoretically from [0,∞] whereas all

$$\rho \geq 1 \tag{1}$$

result in an overload of the system and the queue will never reach a steady state. Thus, we have only considered values for $\rho \in (0,1)$. The obtained results will be compared to the analytic solution. The two measures of interest are:

1) the average queueing delay
2) the probability that the server is idle.

The average queueing delay can be easily calculated as the sum of the average waiting time and average service time:

$$E[T] = E[T_w] + E[T_s] \tag{2}$$

The average service time is simply given by:

$$E[T_s] = \frac{1}{\mu} \tag{3}$$

[1] By using this definition ρ represents the offered traffic for each server

whereas the average service can be calculated using the theorem of total probability:

$$E[T_w] = \sum_{k=0}^{\infty} \{E[T_s^R] + (k-1)E[T_s]\}\pi_k^a \qquad (4)$$

Applying the PASTA theorem and the Markovian assumption to this equation it is straightforward to show that the equation yields:

$$E[T_w] = \sum_{k=0}^{\infty} k\frac{1}{\mu}\pi_k \qquad (5)$$

Using the result for the state probability of the M/M/1 we can obtain:

$$E[T_w] = \frac{1}{\mu}\frac{\rho}{1-\rho} \qquad (6)$$

Combining the two results from 3 and 6 the average queueing delay is:

$$E[T] = \frac{1}{\mu}\frac{\rho}{1-\rho} + \frac{1}{\mu} \qquad (7)$$

Plotting the analytic results for a M/M/1 queue against the offered traffic: The x-axis shows the offered

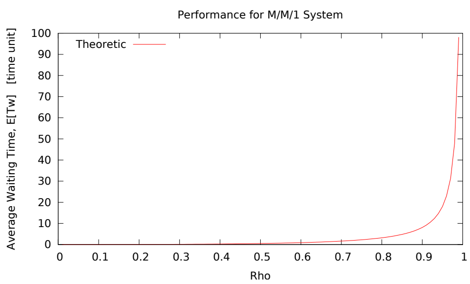

Fig. 1. Average waiting time of an M/M/1 - analytic

traffic, which is obviously dimension less and the y-axis shows the average waiting time in time units. The simulated results for the same case yield:

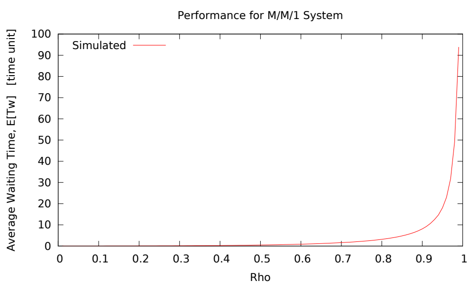

Fig. 2. Average Waiting time of an M/M/1 - simulated

Comparing the results from Fig.1 with the simulated results Fig.2 we should take a closer look at the offered traffic that is further away from the convergence radius. In Fig. 3 we compare the error between the simulated and analytically derived results:

$$\epsilon = \left|\frac{E[T_{ana}] - E[T_{sim}]}{E[T_{ana}]}\right| \qquad (8)$$

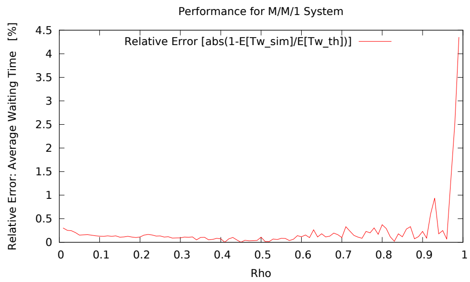

Fig. 3. Relative error between simulated and analytic results

The error is almost constant at roughly 0.5% for all values of ρ yielding a relative error in the order of 10^{-4}. Still, from Fig. 3 it can be obtained that for bigger traffic loads the error is significantly higher - up to almost 5%.

Anyhow, the simulation accuracy is satisfactory.

B. Idle Probability

In this subsection we investigate the probability of the server to be idle depending on the offered traffic. Once again, a comparison between the results obtained by simulation and the analytic results is performed.

First, we consider the analytic probability. This quantity can easily be obtained from the state definition. It is given by the probability of zero customers in the system that is equal to the probability to be in state zero. Thus, we obtain for the M/M/1 queue

$$P\{'zero \quad customers'\} = \pi_0 = 1 - \rho \qquad (9)$$

The idle probability shows a linear dependency on the offered traffic. Fig.4 shows idle probability for, both, the simulation and the analytic results.

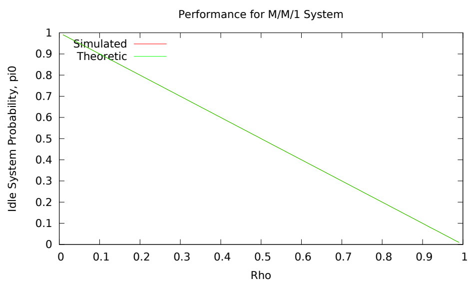

Fig. 4. Idle probability versus offered traffic - simulated & analytic

The curves are almost perfectly overlapping and a difference is not noticeable. Thus, we take a look at the relative error. Fig.5 shows the relative error depending on the offered traffic.

From Fig. 5 we obtain that the simulation is very accurate. The error is almost for all traffic loads below 0.1% until the system diverges.

C. Little's Result

In this subsection we will show the verification of Little's Results based on the results of the simulation.

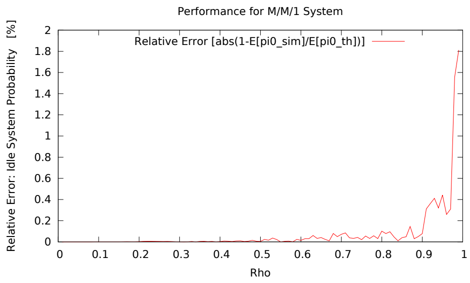

Fig. 5. Relative error of the idle probability

Little's Result states the following:

The Average number of customers in a queue (or portion of a queue, or groups of queues) equals the average arrival rate times the average time spent in the queue (or portion of a queue, or groups of queues) .

$$E[L] = E[T] \cdot \bar{\lambda} \tag{10}$$

Manipulating the Eq. 10 we can obtain the following relationship between the offered traffic $\rho = \frac{\lambda}{\mu}$

$$\frac{E[L]}{E[T]\mu} = \frac{\lambda}{\mu} = \rho \tag{11}$$

From Fig. 6 we can obtain that Little's result holds for our simulation. The theoretically derived result

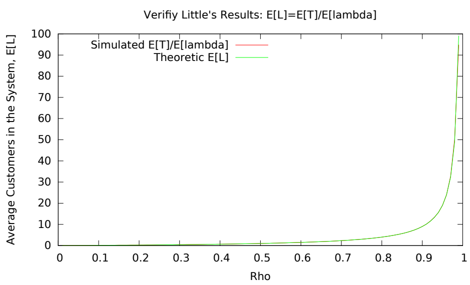

Fig. 6. Average number of customers in the system. Calculation based on average arrival rate and average wating time

matches the observation of our simulation until a certain given simulation accuracy.

III. Task II

In this section we investigate the influence of the service time distribution on the performance of the queue. In particular, the Erlang-2 (E-2) and the Hyperexponential-2 (H-2) distribution are analyzed. The Erlang-2 is a single server with two internal service states. These two services are in series. Despite the Erlang distribution the internal services of the Hyperexponential distribution are connected in parallel. Finally, we are going to verify the result of the Pollaczek-Khinchine formula.

A. Average Queueing Delay

In order to compare the Markovian, E-2 and H-2 in a fair manner the average service time must coincide. To achieve this let us observe the average service time of each particular distribution one by one analytically.

1) Markovian: the average service time is μ by definition. This quantity will be the reference to which we will match the other two.
2) Erlang-n: The overall service time is the sum of all individual service times. Thus,

$$S = \sum_{k=1}^{n} S_i \tag{12}$$

Applying the average we obtain:

$$E[S] = E\left[\sum_{k=1}^{n} S_i\right] = \sum_{k=1}^{n} E[S_i] = \sum_{k=1}^{n} \frac{1}{\mu_i} \tag{13}$$

Hence, we can easily derive the required service rates of the individual stations for the E-2 - $\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_2}$.
3) Hyperexponential-n: In case of the Hyperexponential, we choose with probability α_i branch i which has a service rate μ_i. The average service time is given by

$$\begin{cases} E[S] = \sum_{k=1}^{n} P\{T_{s_k}|S=k\}P\{S=k\} = \sum_{k=1}^{n} \frac{1}{\mu_k}\alpha_k \\ \sum_{i=1}^{n} \alpha_i = 1 \end{cases} \tag{14}$$

Eq. 14 must match the average service rate of the Markovian. In case the of the H-2 this equation simplifies to

$$\alpha_1 \frac{1}{\mu_1} + (1-\alpha)\frac{1}{\mu_2} = \frac{1}{\mu} \tag{15}$$

It is straightforward to see that the choice of the average service rate of the individual stations strongly depends on the choice of the α_i. Fixing the α_i we can derive the values for the service rates.

Once that the average service time is matched the comparison of the systems can begin. Similar as in II we will compare the effect of the offered traffic on the average waiting time and the idle probability.

We start the average waiting time for a customer. Fig. 7 shows the average waiting time versus the offered traffic.

We would expect the average waiting time to be largest for the Hyperexponential and to be smallest for the Erlang. Thus, the Markovian is intermediate. These assumptions are well justified considering the coefficient of variation for the different distributions. In general it is defined as

$$C_s^2 = \frac{Var[S]}{E[S]^2} \tag{16}$$

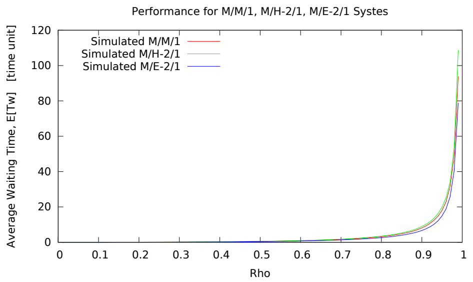

Fig. 7. Average waiting time versus offered traffic for different service time distributions

In case of the Markovian service time distribution $C_s^2 = 1$. For the E-n distribution the coefficient of variation yields

$$C_s^2 = \frac{\sum_{i=1}^{n} \frac{1}{\mu_i^2}}{\left(\sum_{i=1}^{n} \frac{1}{\mu_i}\right)^2} < 1 \tag{17}$$

In the case of $\mu_1 = \mu_2 = 2\mu$ coefficient of variation for the E-2 results in $C_s^2 = \frac{1}{2}$.

This calculation can be repeated for the H-2 distribution. It is straightforward to perform but, unfortunately, the result is not as demonstrative as the result for the E-2. Still, it can be shown, that

$$C_s^2 \geq 1 \tag{18}$$

Taking a closer look at Fig. 7 we can verify that the assumptions from above are also valid for smaller values of ρ.

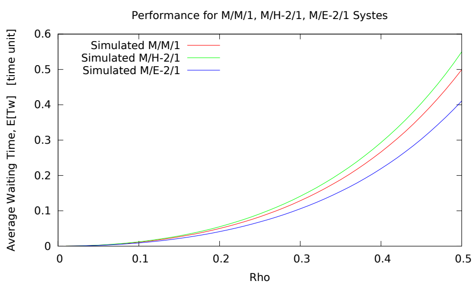

Fig. 8. Average waiting time versus offered traffic for different service time distributions

B. *The Pollaczek-Khinchine Formula*

Reaching this point we conclude the effect of the service time distribution by comparing the simulated results with the analytic results from the Pollaczeck-Khinchine formula. This formula provides a means to evaluate the average number of customers in the system concerning the coefficient of variation. It is given by

$$E[L] = \rho + \frac{\rho^2(1 + C_s^2)}{2(1-\rho)} \tag{19}$$

Note that this formula is only valid for the case of an M/G/1 queue. Multiple servers scenarios are not supported. Applying Little's result to Eq. 19 one shows that the average waiting time yields:

$$E[T] = \frac{1}{\mu} + \frac{\rho\frac{1}{\mu}(1 + C_s^2)}{2(1-\rho)} \tag{20}$$

Eq. 20 allows us to calculate the average waiting time analytically. Fig. 9 depicts the analytic and simulated results versus the offered traffic in a given excerpt. The error is reasonable small which indicates a good fit between the simulation and the analytic evaluation.

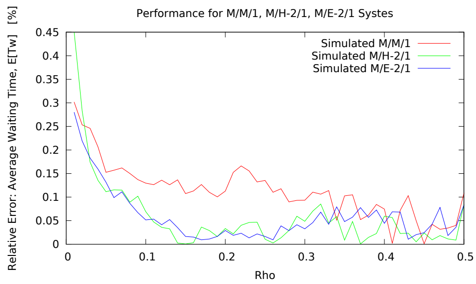

Fig. 9. Error between analytic and simulated results versus offered traffic for different service time distributions

IV. Task III

In this section we are going to analyze the influence of multiple servers. The usual notation M/M/m is used where lower case m denotes the numbers of servers. As performance metrics we choose, similar to the previous tasks, the average waiting time and the probability of the system to be idle. As it's mandatory we also investigate the average number of busy servers.

First we are comparing the results of the M/M/2 queue since this is the biggest step in the generalization of a single server queue to a multiple servers queue. Once that we cleared this hurdle we are going to change the service time distribution. Doing this we will investigate the special case of a deterministic service time distribution.

A. *Average Queueing Delay*

For the M/M/2 queue one can analytically derive the following dependency for the average queueing delay of a customer and the average number of customers in the system, respectively.

$$E[T] = \frac{1}{\mu} + \frac{\lambda^2}{\mu(4\mu^2 - \lambda^2)} \tag{21}$$

Using Little's result it is straightforward to show that

$$E[T] = \frac{\lambda}{\mu} + \frac{\lambda^3}{\mu(4\mu^2 - \lambda^2)} \tag{22}$$

Fig. 10 shows both the performance of the analytic and simulated solutions for the M/M/2 queue versus the offered traffic. A coarse look already provides that both curves coincide almost perfectly. The error between these curves is shown in Fig. 11

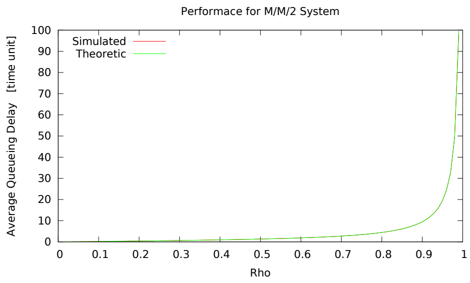

Fig. 10. Average queueing delay of the M/M/2 queue versus offered traffic - simulated and theoretic

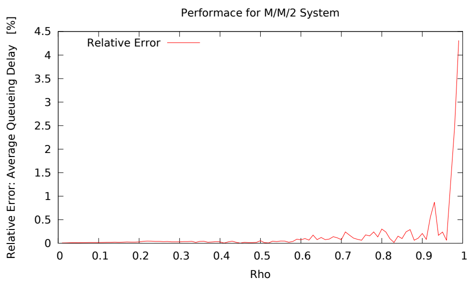

Fig. 11. Error between analytic and simulated results versus offered traffic for the M/M/2 queue

B. Deterministic Service time

Now that we have verified our simulation in case of multiple servers we change the service time distribution from Markovian to Deterministic. As a reference we compare the Deterministic service time to the Markovian, E-2 and H-2, respectively. Prior simulation we would expect the average queueing delay to be the least of all curves for all values of ρ. This is well justified by the fact that the statistical variation is the least of all distributions - the coefficient of variation is, obviously, zero. In Fig.12 we can observe that the previous assumption was actually true.

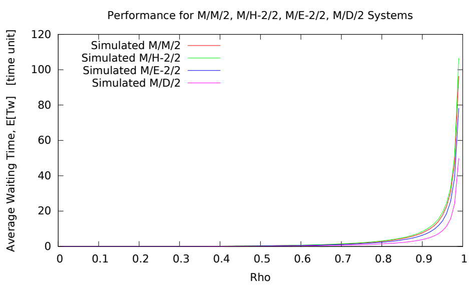

Fig. 12. Average queueing delay versus the offered traffic - service time distributions: M,E-2,H-2,D

Fig. 13 provides a closer look at the graphs to achieve more insight.

C. Average Number Of Busy Servers

The last quantity of interest for this task is the average number of busy servers. It can be calculated analytically if a steady state distribution has been

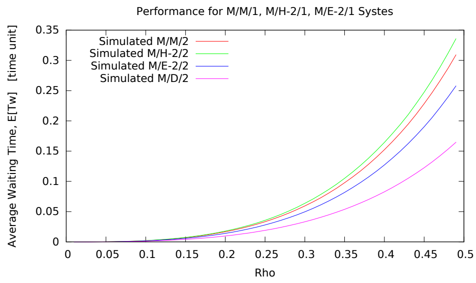

Fig. 13. Average queueing delay versus the offered traffic - service time distributions: M,E-2,H-2,D - scaled

found. Let's consider the case of an M/G/m queue where Q represents the average number of busy servers and k the total number of customers in the system

From TABLE I we can conclude that the average

π_k	Q
π_0	0
π_1	1
π_2	2
$\vdots$	$\vdots$
π_m	m
π_{m+1}	m
$\vdots$	$\vdots$

TABLE I
NUMBER OF BUSY SERVERS DEPENDING ON THE STATE π_k

number of busy servers yields:

$$E[Q] = \sum_{k=1}^{m} k\pi_k + m \sum_{k=m+1}^{\infty} \pi_k \tag{23}$$

To get rid of the infinite sum we make use of the complementary probability

$$E[Q] = \sum_{k=1}^{m} k\pi_k + m \left(1 - \sum_{k=0}^{m} \pi_k\right) \tag{24}$$

In the case of a general service time distribution the solution for E[Q] is not trivial to calculate analytically. Thus, we use the results of the simulation to obtain a solution. Anyhow we can state some boundaries for result prior simulation. The average number of busy servers must lay in the range of

$$0 < Q < m \tag{25}$$

if we consider the queue in equilibrium - $0 < \rho < 1$. If ρ would be equal to the total number of servers, m, we would operate the queue in overload. From Fig. 14 we can obtain the evolution of the average number of busy servers for increasing offered traffic. Certainly, the number converges towards 2 the closer we get to the convergence radius of the queue.

V. Task IV

In the final task of Lab.1 the effect of a finite waiting line will be investigated and evaluated. Starting with the analysis of average queueing delay we will discuss the special case of no waiting line with respect to the service time distribution.

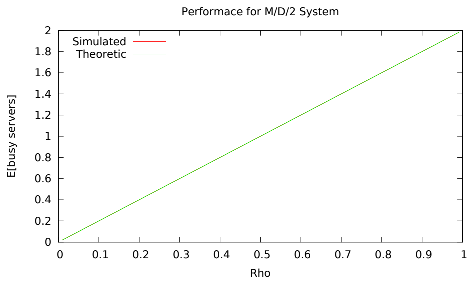

Fig. 14. Average number of busy servers versus the offered traffic

A. Average Queueing Delay

The respective queue that is going to be investigated is the M/G/m/B. Certainly, the choice of the parameters will affect the results of the simulation significantly. Similar to the investigation in Section IV we will choose the number of servers to be m=2. Again, it is the step from one to two which comprises the most difficult part. The capacity of the waiting line is a crucial quantity for the performance of the queue. Despite the case of B=0 - which is investigated in Subsection V-B - we will analyze the average queueing delay and loss probability for two further cases:

1) $B = 1$
2) $B = 2$

Of course we can forecast some features of the behavior in advance. In general, we expect the loss probability of the queue with the bigger waiting line to be less than for the queue with the smaller waiting line. Furthermore, the loss probability for small loads should be significantly low in both cases.

Fig. 15 depicts the loss probability versus the offered traffic. Both of the above mentioned aspects can be seen in Fig. 15.

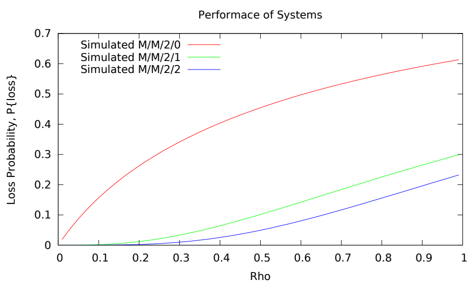

Fig. 15. Loss probability of M/M/2 queues with waiting line size equal 1 and 2, respectively

B. Special Case: Zero Waiting Line

The case where the server does not have a waiting line is a nice example of a queue which is insensitive to the service time distribution. This can also be shown analytically. We will not provide a full prove for the M/G/m/0 queue but we will focus our argumentation on some basic examples: the M/M/1/0 and M/E-2/1/0, respectively. Starting with the analysis of the M/M/1/0 we can directly obtain the state equations

$$\begin{cases} \lambda\pi_0 = \mu\pi_1 \\ \pi_0 + \pi_1 = 1 \end{cases} \tag{26}$$

The solution for the two states is obviously

$$\begin{cases} \pi_0 = \frac{\mu}{\mu+\lambda} \\ \pi_1 = \frac{\lambda}{\mu+\lambda} \end{cases} \tag{27}$$

Applying the PASTA theorem the loss probability yields:

$$P('loss') = \pi_1^{(a)} = \pi_1 = \frac{\lambda}{\mu+\lambda} \tag{28}$$

To compare this result with the other queues we will analyze the M/E-n/1/0 queue in the following step. Here the story is a little bit more involved. Asssuming that $\mu_i = \frac{n}{\mu}$ we can derive the following state equations:

$$\begin{cases} \lambda\pi_{0,0} = n\mu\pi_{1,n} \\ n\mu\pi_{1,1} = \lambda\pi_{0,0} \\ n\mu\pi_{1,k} = n\mu\pi_{1,k-1} \\ \pi_{0,0} + \sum_{k=1}^{n} \pi_{1,k} = 1 \end{cases} \tag{29}$$

Manipulating the equations properly we can achieve the following result:

$$\begin{cases} \pi_{1,n} = \frac{\lambda}{n\mu}\pi_{0,0} \\ \pi_{1,k} = \pi_{1,k-i} \qquad i < n \\ \Rightarrow \pi_{0,0} + n\pi_{1,n} = 1 \end{cases} \tag{30}$$

Thus, the state probabilities are:

$$\begin{cases} \pi_{0,0} = \frac{\mu}{\mu+\lambda} \\ \pi_{0,k} = \frac{\lambda}{n(\mu+\lambda)} \qquad 1 \le k \le n \end{cases} \tag{31}$$

Reaching this point the loss probability yields

$$P('loss') = \sum_{k=1}^{n} \pi_{0,k}^{(a)} = \sum_{k=1}^{n} \pi_{0,k} = \frac{\lambda}{\mu+\lambda} \tag{32}$$

The general proof for the M/G/m/0 queue is straightforward but exceeds the scope of this task by far. It can be provided exploiting the concept of the Renewal Process. Thus, from this point we will assume that the insensitivity holds for all variants of the mentioned queue.

This can of course be shown by simulation. Fig. 16 shows the above mentioned insensitivity. Concerning Eq. 28 we would expect the loss probability to converge to $\frac{1}{2}$ for ρ growing to one. That, indeed, is what can be read off the graph.

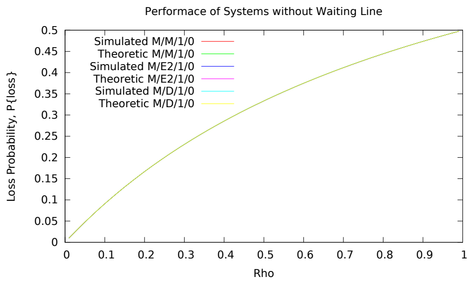

Fig. 16. Loss probability versus offered traffic - special case: zero waiting line

Laboratory 2

***Abstract*—In this second report we again make use of simulation to compare theoretical results obtained in class with simulated one. Here we focus on a more general and slightly complex system, the network of queues. We will show the behavior of a system made of two M/M/1 queues in series firstly with same service rate and in the second part varying it for each queue. At the beginning we will prove the product form solution and then we will focus our discussion on the average buffer occupancy when the service rates change and change with respect to each other.**

Since the results we obtain are realizations of a random variable, we also provide confidence intervals for our results, to make them more meaningful.

***Index Terms*—Queuing theory, confidence interval, network of queues, performance evaluation, Burke's theorem**

VI. Introduction

We again make use of simulation to evaluate performances of a system. As discussed in Section I, simulation is a powerful tool and can be used in many different situations, one of that is to validate theoretical results.

In this laboratory we introduce the concept of confidence interval. In the past to get more general results, we forced the simulation to last for a huge amount of time. When we make a simulation we get a result, but this number is given by a single realization of a stochastic process, however long, but it is only a special case of a single realization. We now want to study how much the result we get is likely to appear in real situations. To do so we run several times the same simulation[2] and we get different results. All those results have similar values, but not always the same. They all fall in the same interval, the one which is called *confidence interval*. The confidence interval helps the user to understand the precision of a measure and the error related to it. A measure without confidence interval is not as meaningful as a measure that has it because we do not know if it is reasonable or not. In the end the confidence interval is defined by the mean value of the interval and delimited by the mean value plus and minus a specific value.

[2]with a shorter time with respect to the tests we did for Sections from II to V

Before stating our simulation we define our required *confidence level*, so the percentage of goodness of our result. This is a very important parameter because lowering this parameter will make our computation simpler because we need less precision, but the confidence interval would be enlarged leading to a greater variability on the error on our final result. This short discussion is already able to show the importance of the variance associated to the mean value of our result.

In this report our main purpose is to analyze a simple network made of two queues and compare the simulated results to the theoretical ones. Nowadays networks have plenty of queues where packets remain for a random amount of time depending on the different arrival and service rates of network devices, so this is a good starting point to study more complex systems and related situations.

VII. Task I

In the first task it is required to consider two M/M/1 queues, derive the steady state of the system and verify the product form solution.

Our simple network is composed by two queues in series with only one server each that serves customers at rate μ_1 for the first queue and μ_2 for the second. The arrival rate in the first queue is, as usual, λ. Furthermore we keep the constraint of the previous laboratory $\rho < 1$. Rates in the system, both arrival and departure, are considered to be Poisson in order to use a simpler mathematical model that in the end allows us to use the Burke's theorem and get the product form solution. We call product for solution the relation:

$$\pi_{k_1,k_2} = \pi^{(1)}_{k_1}\pi^{(2)}_{k_2} = (1-\rho_1)\rho_1^{k_1} \cdot (1-\rho_2)\rho_2^{k_2} \quad (33)$$

This formula expresses the probability of having exactly k_1 customers waiting in queue 1 and k_2 customers waiting in the other queue. Since the the two queues are independent, the value of π_{k_1,k_2} is given by the product of the probabilities of the two queues taken independently. In the end, we know the arrival and service rate of a queue and so we can use these information to quantify the probability of having exactly k_i customers in each one. We know from Burke's theorem that the departure process from an M/M/1 queue in the steady state is a Poisson process with the same parameter of arrivals, in our case λ. This observation, maybe trivial, is of great importance because allows us to consider the arrival rate to queue 2 equal to λ and further to evaluate independently the behaviors of the two queues.

In our simulation to validate the product form solution we plot the average time spent in the system against its load. First of all we begin from the simplest case possible where $\mu_1 = \mu_2$ and so ρ, the load of the whole system is equal to both ρ_1 and ρ_2. Being the two queues independent, the average time spent in the system is given by the sum of the times in the two queues separately (result obtained starting from Eq. 7):

$$E[T] = \sum_{i=1}^{2} E[T_i] = \sum_{i=1}^{2} \frac{1}{\mu_i}\frac{\rho_i}{1-\rho_i} + \frac{1}{\mu_i} \qquad (34)$$

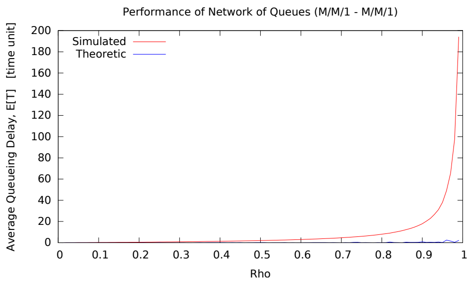

Fig. 17. Average time spent in a network of two M/M/1 queues

In Figure 17 we compare the theoretical sojourn time in the system obtained with the product form solution with the one we get with the simulation. We can easily see that we get a simulated value very closed to the theoretical one obtained with the product form. After this test, we can say that product form solution holds in this specific system.

For what concerns the accuracy of our results the confidence interval is given by the formula:

$$I = \left[\hat{x} - z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \hat{x} + z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right] \qquad (35)$$

The value $\hat{x}$ is the mean of the expected values of the different observations of our simulation. *n* is the number of observations of our simulation and *s* is the standard deviation. According to the number of observations per instance taken into account, we can use one between two different distributions to approximate our results: *Student's t* distribution if we take less than 30 observations per instance or the *normal* distribution if we take more than 30 ones. A higher number of observations over several independent simulations allows us to collect independent samples, so we decide to take 40 observations for each instance and then to use the formulas for the normal distribution and the central limit theorem.
Another issue related to the simulation's duration is the initial transient. When we start generating values and averaging them, we get a mean value which is different from the one we get in the steady state. There are many ways to solve this problem. The solution we adopt in our program, which is the simplest but also most widely used, is to take a very long lasting simulation so that the initial transient has a negligible effect on the total mean value. In this way our simulation lasts for a longer time and we get a grater amount of data, but we are sure to get more significant values.
In Eq. 35 there is the term z which is a variable we use to evaluate the error in the simulation. z is proportional to the probability that the value x_i we extract from the average of the observations of the simulation i falls inside the confidence interval.
z depends on the value $1-\alpha/2$. This is the confidence level of our measure, so how much we are sure that our measure follows inside this interval. We have to fix this value before running the simulation to evaluate the minimum number of observations to have the final result with the precision we want. Geometrically we obrain that

$$P\left\{z \in \left[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha \qquad (36)$$

The smaller the value of z, the higher the precision of the measure. In MATLAB© we can find the value of z simply using the `erfcinv`, the inverse function of the erfc:

$$z = \sqrt{2}\ \text{erfcinv}(1-\alpha) \qquad (37)$$

In Eq. 35 we see that the confidence interval also depends on the ratio between the standard deviation s and the square root of number of observations for each instance.

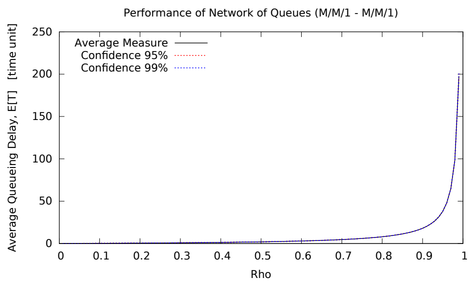

Fig. 18. Confidence level of the estimate average time in the system

In Figure 18 we represent our simulation with two confidence levels we defined: 95% and 99%. In this two cases we get confidence intervals which are very closed to the mean value got from the average values of each simulation. This is straightforward to the fact that we want confidence level of 95% and 99% on a simulation time lasting 10^6 time units and we get 40 observations for each instance.

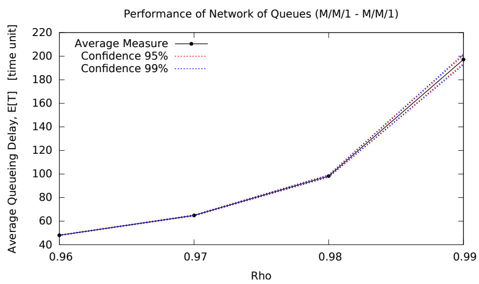

Fig. 19. Zoom of confidence intervals between load 0.96 and 0.99

In Fig. 19 we take a closer look to the confidence level for high loads, where it is easier to find larger confidence intervals. Also in this situation we see thin confidence intervals, meaning very precise values. Here we see that the confidence interval of 99% is wider than the other and this is what we expect since a higher level of confidence corresponds to a wider confidence interval.

In the end we can say that our measurements are very precise because as shown in Figure 20 we always get small percentage differences meaning our confidence interval is very thin and all the estimated values are very likely. We can easily see that in the case of the confidence level of the 99% we always have a gap smaller than the 2%, meaning very negligible. In the case of high loads we also see a greater dissimilarity because we are reaching the instability of the system and it is even more difficult have good values in this case. With a longer simulation we can reduce this effect, also if it will always be greater with respect to the previous part of the simulation because we will always have a situation near the overload.

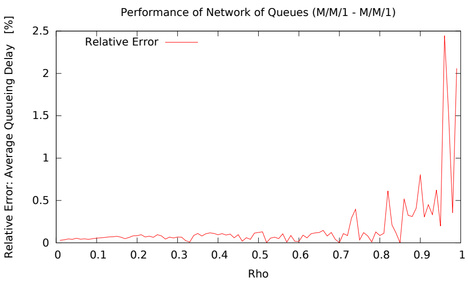

Fig. 20. Error on the average time spent in the system

VIII. Task II

In this last section we analyze the network of queues of the previous section with different service rates and we compare the different numbers of customers in the waiting lines of the two queues. Furthermore this evaluation is done also in the case of a network with two M/D/1 queues. In both cases we plot the average number of customers in the waiting line on the y-axis and the ratio between ρ_2 and ρ_1 on the x-axis. The request of the laboratory is to show the different behaviors of the system in the case where $\mu_1 \ll \mu_2$ and then $\mu_1 \gg \mu_2$. Since all the graphs of this report have been plotted in terms of ρ, we will show the graph $\rho_1 \gg \rho_2$ for $\mu_1 \ll \mu_2$ and $\rho_1 \ll \rho_2$ for $\mu_1 \gg \mu_2$.

In Fig. 21 we keep constant the arrival rate and the service rate of queue 2 while changing the service rate of queue 1. On a logarithmic scale it is easier to see that for small values of μ_1 the customers in the system tent to wait in the first waiting queue. Increasing μ_1 we see that the number of customers in waiting line 1 decreases until we reach the point studied in the Section VII and then decreases rapidly reaching the opposite case, when the majority of customers waits in the second waiting line.

When the service rate of the first queue is very fast, this means that as soon as there is an arrival in queue 1, there is immediately a very fast service that brings the customer to the second queue. Increasing μ_1 only in very few cases customers have to wait in the first waiting line. Then we see that if the first service rate is very high, so high that does't stop customers, the number of customers in the system tent to be equal to the theoretical average number of customers in the waiting line of a queue.

In Fig. 22 we plot the same test for a system made of two M/D/1. It is immediately evident that performances are very similar for the two systems.

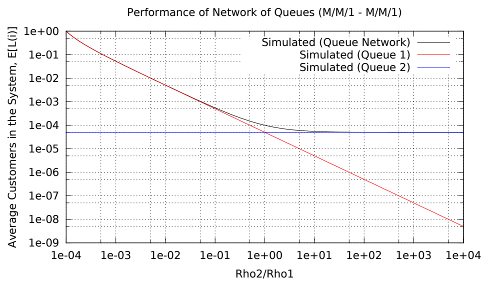

Fig. 21. Average number of customers waiting in the queues with Markovian services

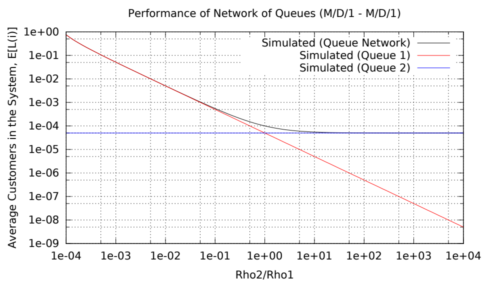

Fig. 22. Average number of customers waiting in the queues with Deterministic services

In Fig. 23 and 24 we see, again on a logarithmic scale, the percentage of customers in queue 1 and in queue 2 with increasing ρ_2/ρ_1. Again, as expected from Fig. 21 and 22, the percentage of the customers waiting in the first line decreases with the increasing of μ_1 and since probabilities must sum to one, the probability that customers waiting in the second line increases. As in the previous graphs, we see that when the two service rates are equal, it is obvious to have the same probability of being in queue 1 or in queue 2.

In the end in Fig. 25 and 26 we can see the average number of customers in the waiting lines for a system of two M/M/1 queues and of two M/D/1 queues. As usual, on the x axis there is the ratio ρ_1/ρ_2.

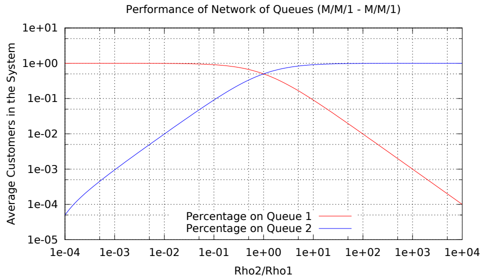

Fig. 23. Percentage of customers in the two queues for a network of two M/M/1 queues

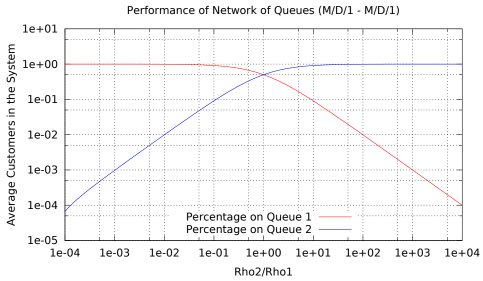

Fig. 24. Percentage of customers in the two queues for a network of two M/D/1 queues

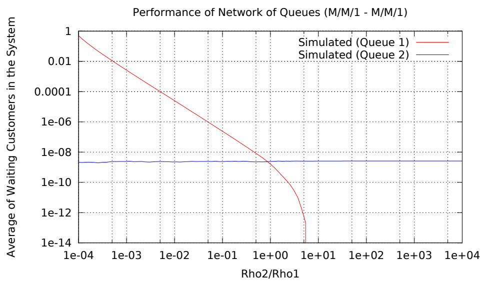

Fig. 25. Average value of customers waiting in the waiting lines for a system of two M/M/1 queues

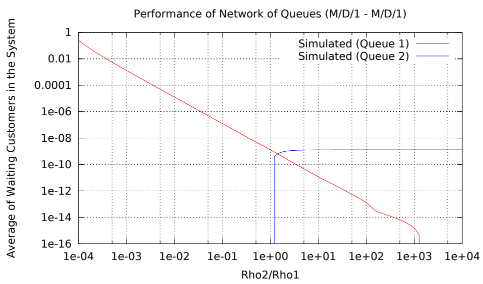

Fig. 26. Average value of customers waiting in the waiting lines for a system of two M/D/1 queues