M/M/1 queue

Performance measures

Now we propose to analyze the performance measures for the M/M/1 queue.

Mean performance measures

Mean performance measures in the steady state behavior

- We compute mean performance measures (stable system)
- Main performance measures:
  - mean number of customers: N
  - mean sojourn time (in the system): R
  - mean waiting time (in the queue): W
  - mean throughput: X
  - mean utilization rate: U

We suppose that we compute the mean performance measures in the steady-state, so in the case of a stable system.
The main performance measures are:

- the mean number of customers N,
- the mean sojourn time R,
- the mean waiting time in the queue W,
- the mean throughput X,
- and the mean utilization rate U.

Little's Law

Little's Law

$$\underbrace{N}_{\text{Mean number of customers}} = \underbrace{X}_{\text{Mean throughput}} \cdot \underbrace{R}_{\text{Mean sojourn time}}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ unit : customers & unit : customers/s & unit : s \end{array}$$

We give an important formula in the steady state, called Little's Law: it says that the mean number of customers N is equal to the product of the mean throughput X by the mean sojourn time R. The relation is verified in terms of units: the unit for N is customers, for X it is customers per time unit, and for R it is the time unit.

Mean throughput

Mean throughput X

- The M/M/1 has an infinite capacity $\Rightarrow$ no losses of customers
- Mean input throughput = Mean output throughput

$$\boxed{X = \lambda}$$

For the mean throughput X, as the M/M/1 has an infinite capacity, then there are no losses, and so the mean input throughput is equal to the mean output throughput. So X is equal to λ.

Mean number of customers

Mean number of customers N

- N is computed as follows:

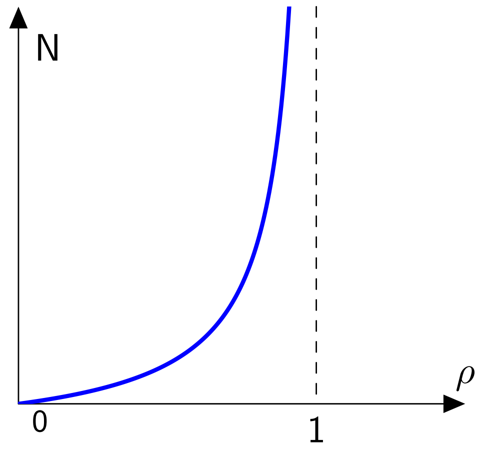
$$\begin{aligned} N &= \textstyle\sum_{i=0}^{\infty} i\pi_i = (1-\rho)\sum_{i=0}^{\infty} i\rho^i \\ &= (1-\rho)\,\rho \underbrace{\sum_{i=1}^{\infty} i\rho^{i-1}}_{\frac{1}{(1-\rho)^2}} \\ &= \frac{\rho\,(1-\rho)}{(1-\rho)^2} \end{aligned}$$

$$\boxed{N = \rho\,/\,(1-\rho)}$$

For the mean number of customers N, then it is computed as the sum over i of i multiplied by π_i. If we replace π_i by its expression, then we obtain $(1-\rho)\sum_{i\geq 0} i\rho^i$.
We put ρ outside the sum, and we obtain the sum $\sum_{i\geq 0} i\rho^{i-1}$.
In the case $\rho < 1$, as the sum is equal to $\frac{1}{(1-\rho)^2}$, we obtain this mathematical expression.
Simplifying, we obtain that N is equal to $\frac{\rho}{1-\rho}$.

Analysis of the mean number of customers

Effect of the load ρ



Here we can see the curve of the number of customers N, according to rho. So when rho increases, the number of customers increases, as the system is more and more loaded. When rho is close to 1, N increases very quickly.

Mean sojourn time

R: mean sojourn time of the system

- R is computed from Little's law

$$R = N \,/\, X = \rho \,/\, [(1-\rho)\lambda] = \frac{1}{\mu}\frac{1}{1-\rho}$$

- mean waiting time, mean service time:

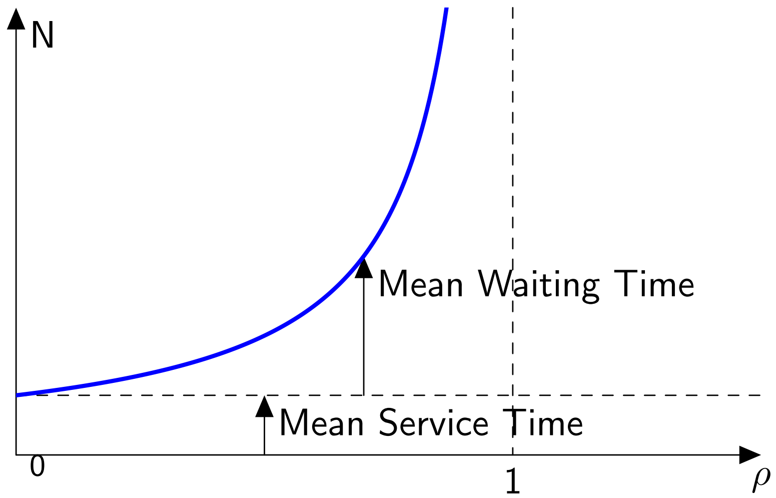
$$\begin{array}{ccccc} R & = & \frac{1}{\mu} & + & \frac{\rho}{\mu(1-\rho)} \\ \uparrow & & \uparrow & & \uparrow \\ \text{Mean sojourn time} & & \text{S} & & \text{W} \\ & & \uparrow & & \uparrow \\ & & \text{Mean time of service} & & \text{Mean waiting time} \end{array}$$

Now, we focus on the mean sojourn time R. We use Little's law. R is equal to N divided by X, we replace N by $\rho/(1-\rho)$ and X by λ, so we obtain this formula, which is equal to this other formula as ρ is equal to λ/μ.

Then we can obtain that the mean sojourn time is equal to the sum of the mean service time S, and the mean waiting time W in the queue.

Analysis of the mean sojourn time

Effect of the load ρ



Here we give the curve of the mean sojourn time according to the load ρ.
We can see that R is always greater than the mean service time S.
It is equal to the sum of the service time S and the waiting time W in the queue.

Mean utilization rate

U: mean utilization rate

- U: the proportion of time that the server is busy (at least one customer)

$$\boxed{U = 1 - \pi_0 = \rho = \lambda/\mu}$$

The mean utilization rate U is defined as the proportion of time that the server is busy, so when there is at least one customer in the system. We deduce that U is equal to $1 - \pi_0$, so it is equal to ρ.