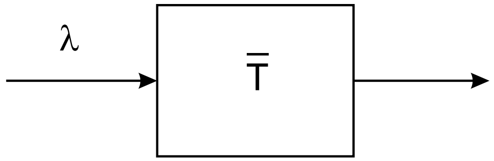
Little's result

The result

Little's result or Little's theorem is a very simple (but fundamental) relation between the arrival rate of customers, average number of customers in the system and the mean sojourn time of customers in the system.



Consider any system as a "black box" and define

$$\begin{cases} \lambda &= \text{average arrival rate (average number of customers arriving per time unit)} \\ \bar{T} &= \text{average sojourn time (time spent) in the system} \\ \bar{N} &= \text{average number of customers in the system} \end{cases}$$

Little's result says that $\boxed{\bar{N} = \lambda \bar{T}}$

The result is very useful because of its generality

- Nothing is assumed about the system
  - any part of the system can be considered as a black box
- The arrival process can be anything
  - in particular, one need not assume it to be a Poisson process
  - the process, however, has to be stationary

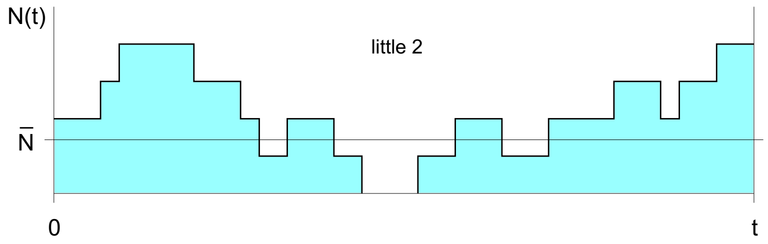
The use of Little's result

In queueing theory it is often simpler to derive a result concerning the average number of customers in the system. Then Little's can be used to calculate the corresponding time in system (sojourn time), $\bar{N} \rightarrow \bar{T}$.

Sometimes the result is used in the converse direction, i.e. starting from the average sojourn time one deduces the average number in system, $\bar{T} \rightarrow \bar{N}$.

Justification of Little's result

The number in system N_t forms a stochastic process. Its average value over a long period of time t can be calculated by dividing the shaded area in the figure by the length of the period t.



- On average, each customer contributes to the area by the amount $\bar{T}$.
- The average number of customers arriving in interval t is λt.
- The area is thus $(\lambda t)\bar{T} \quad \Rightarrow \quad \bar{N} = \lambda \bar{T}$.

Intuitive reasoning: Assume that each customer is charged a fee of $(T/\text{min})\cdot$euro, i.e. one euro per each minute stayed in the system. The average income rate of the system is thus $\lambda \bar{T}$ euro/min , but it is also $\bar{N}$ euro/min since each customer in the system generates an income rate 1 euro/min.

Application: traffic intensity (load)

Consider some elements in a network, e.g. trunks, ports or any logical units such that each customer in the system reserves one such element. Let

$$\begin{cases} \lambda & = \text{ the arrival rate of customers into the system} \\ \bar{T} & = \text{ average holding time of an element} \end{cases}$$

The quantity $\boxed{a = \lambda \bar{T}}$ is called the (offered) traffic intensity or load of the traffic.

- Traffic intensity is a pure number, but in order to emphasize the context one often denotes as its "unit" erlang or erl (as a tribute to the Danish pioneer of traffic theory, A.K. Erlang).

By Little's result, the traffic intensity is the same as the average number of simultaneously reserved elements.

Example. In the trunk group from a PBX (private branch exchange) to the central office there are on the average 150 calls per hour. The mean holding time of a call is on 3 min. The traffic intensity is

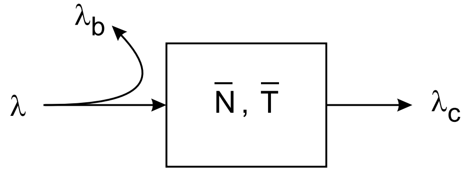
$$a = 150 \text{ h}^{-1} \times 3 \text{ min} = 7.5 \text{ erl}$$

In other words, in the trunk group there are on the average 7.5 calls in progress simultaneously.

Offered, blocked and carried traffic

All real systems have a finite capacity and some of the customers requesting service have to be refused (blocked). Therefore, we have to distinguish the following concepts

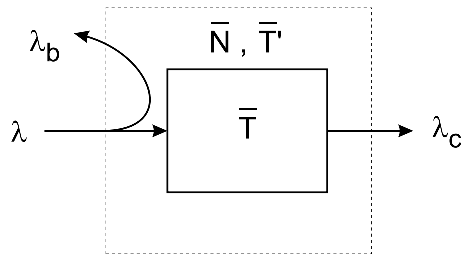
$$\begin{cases} \lambda & = \text{offered traffic stream (arrival rate)} \\ \lambda_b & = \text{blocked traffic stream} \\ \lambda_c & = \text{carried traffic stream} \end{cases}$$



Little's result applies to the stream of customers admitted into the system

$$\boxed{\bar{N} = \lambda_c \bar{T}} \qquad \text{traffic intensity of the carried traffic}$$

Little's result can also be applied by extending the boundary of the system as shown in the figure. Then all the customers enter the system (arrival rate λ), but part of the customers receive a harsh service (immediate kick out).



The average time in system is now

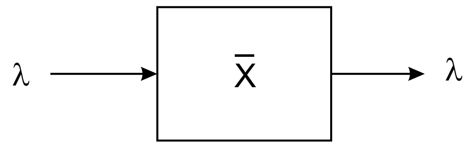
$$\bar{T}' = \frac{\lambda_b}{\lambda} \cdot 0 + \frac{\lambda_c}{\lambda}\bar{T} = \frac{\lambda_c}{\lambda}\bar{T}$$

Little's result gives

$$\bar{N} = \lambda \bar{T}' = \lambda_c \bar{T}$$

which is the same result as previously.

Traffic intensity of the carried traffic in a singel server system

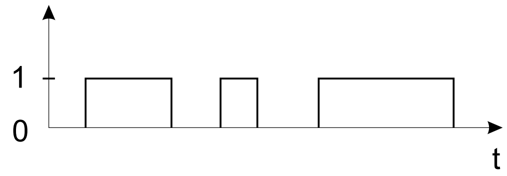


Consider a single server system, i.e. a system which can accommodate only one customer at a time.

$$\begin{cases} \lambda & = \text{the carried traffic stream passing through the server (served customers)} \\ \bar{X} & = \text{average service time} \end{cases}$$

By Little's result we have $\bar{N} = \lambda \bar{X}$.

On the other hand, the average occupancy can be calculated directly



$$\bar{N} = p_0 \cdot 0 + p_1 \cdot 1 = p_1$$

$$\begin{cases} p_0 & = \text{probability that the server is free} \\ p_1 & = \text{probability that the server is occupied; the usage of the server, } \rho \end{cases}$$

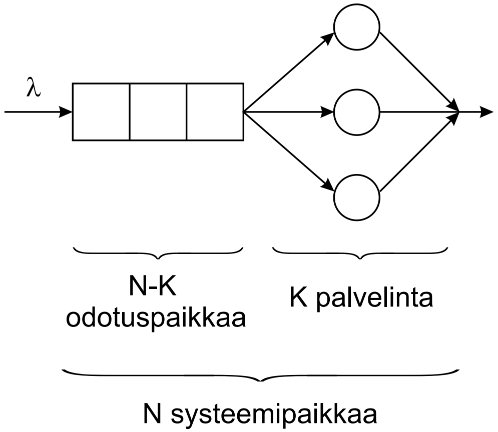
$$\boxed{a = \lambda \bar{X} = \rho}$$

In a single server system, the carried traffic intensity is the same as the usage of the server (proportion of time the server is being used).

Little's result: example

In the system there are

$$\begin{cases} N & \text{system places in total} \\ K & \text{servers, mean service time } \bar{X} \\ N-K & \text{waiting places} \end{cases}$$



Assume that the arrival rate is so high that the system si always saturated (all places occupied). Denote by λ the rate of the carried customer stream.

Question: What is the mean sojourn time in system $\bar{T}$?

We apply Little's result twice: first to the set of servers and then to the whole system.

1. Set of servers

Mean time spent in a server $= \bar{X}$

$$K = \lambda \bar{X} \quad \Rightarrow \quad \lambda = \frac{K}{\bar{X}}$$

2. The whole system

$$N = \lambda \bar{T} \quad \Rightarrow \quad \bar{T} = \frac{N}{\lambda} = \frac{N}{K} \cdot \bar{X}$$

$$\boxed{\bar{T} = \frac{N}{K} \cdot \bar{X}}$$