Derivation of Little's Law

M. Veeraraghavan, Feb. 10, 2004

1. Proof for Little's law using one sample function

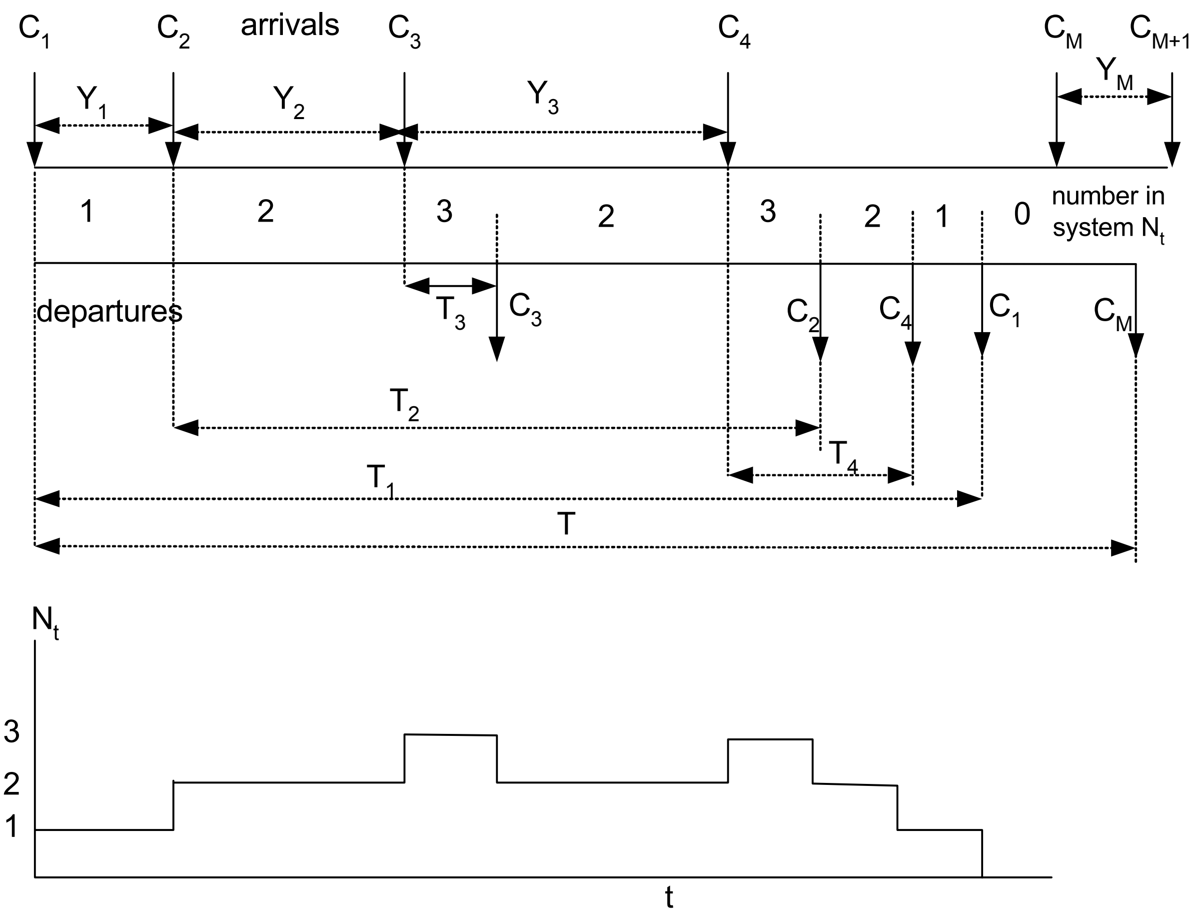

Figure 1:Illustration of call arrivals, departures and the number of calls in the system

T and M are random variables with both $M, T \to \infty$, where M is the number of arrivals in time $(0, T)$. The area under the curve in the N_t plot shown in Fig. 1, which is the product of N_t and time t is given by:

$$\int_0^T N_t dt = \sum_{i=1}^{M} T_i \text{ as seen from Fig. 1.} \tag{1}$$

The **time average of customer delay** up to time T spent by different arrivals in the system is given by (where M is the number of arrivals in time $(0, T)$):

$$\lim_{M \to \infty} \left(\frac{1}{M} \sum_{i=1}^{M} T_i \right) = \bar{T} \tag{2}$$

The **time average arrival rate** is $\lambda_T = M/T$ where M is the number of customers who have arrived in the interval $[0, T]$. The steady-state arrival rate (if it exists) is given by:

$$\lim_{M, T \to \infty} \left(\frac{M}{T} \right) = \bar{\lambda} \tag{3}$$

Time average number of number of customers in the system:

$$\lim_{T \to \infty} \left(\frac{1}{T} \int_0^T N_t dt \right) = \bar{N} \tag{4}$$

Dividing both sides of (1) by T, we get

$$\frac{1}{T} \int_0^T N_t dt = \frac{M}{T} \times \frac{1}{M} \sum_{i=1}^{M} T_i \tag{5}$$

Taking the limit of both sides of (5) as $T \to \infty$, which means both M and T also approach ∞, we get:

$$\bar{N} = \bar{\lambda} \bar{T} \tag{6}$$

This is a general result and can be applied even if customers are not served in the order thet arrive. The assumption that the system is initially empty is not required.

2. Application of Little's law to ensemble averages - multiple sample functions

The above graphical proof is for a single sample function. We can now replace time averages by ensemble averages. To clarify the ensemble average,

$$\overline{N(t)} = \sum_{n=0}^{\infty} n p_n(t) \qquad (7)$$

where $p_n(t)$ is the probability of n customers in the system at time t. In other words, if we take many sample functions and find the probability that at a given instant past the start of each sample funtion, say at time t, what is the probability that there are n customers (across the sample functions), this is $p_n(t)$. By taking the average of the number of customers in the system at time t across these multiple sample points gives us $\overline{N(t)}$, the ensemble average.

Reference [2] simply states that almost every system of interest to us is **ergodic** in the sense that the time average, $N = \lim_{t \to \infty} N_t$ of a sample function is, with probability 1, equal to the steady-state (ensemble) average $\bar{N} = \lim_{t \to \infty} \overline{N(t)}$, that is

$$N = \lim_{t \to \infty} N_t = \lim_{t \to \infty} \overline{N(t)} = \bar{N} \text{ - ON PAGES 156-157.} \qquad (8)$$

A similar result holds for the time average of customer delay

$$T = \lim_{M \to \infty} \left(\frac{1}{M} \sum_{i=1}^{M} T_i \right) = \lim_{M \to \infty} \overline{T_M} = \bar{T} \qquad (9)$$

where $\bar{T}$ is the steady-state time average of delay experienced by a customer, $\overline{T_M}$ is the ensemble average of delays experienced by the first M customers who arrive and depart.

Thus, assuming time average = ensemble average (i.e., ergodic process), we can write

$$N = \lambda T, \qquad (10)$$

where N and T are stochastic (ensemble) steady state averages and λ is the steady state arrival rate.

3. Applicability of Little's law

Little's law applies for any queueing system that reaches steady-state. The system can have more than 1 queue. It is very general.

4. Relation between mean waiting time and mean number in queue vs. mean response time and mean number in system

Note also that

$$N = N_Q + N_S \text{ (number in queue + number in service)} \tag{11}$$

$$E[N] = E[N_Q] + E[N_S] \tag{12}$$

If server is idle (which occurs with probability $1 - \rho$), then there are 0 in service; else there is 1 job in service:

$$E[N_S] = \rho = \lambda E[X] \tag{13}$$

$$E[T] = \frac{E[N]}{\lambda} = E[W] + E[X] \text{ (by Little's Law)} \tag{14}$$

Substituting for $E[N]$ in (12) using (14) and for $E[N_S]$ in (12) using (13)

$$\lambda(E[W] + E[X]) = E[N_Q] + \lambda E[X] \text{ or equivalently} \tag{15}$$

$$E[N_Q] = \lambda E[W] \tag{16}$$

Difference between average number in queue and average number in system is not 1 but ρ because server has to be busy. Note that both (14) and (16) are true, i.e., relation relation between average response time and average number in system, and between average waiting time and average number in queue.

5. Examples

See [2], pages 157-162.

References

[1] R. Boorstyn, Class notes.
[2] D. Bertsekas and R. Gallager, "Data Networks," Prentice Hall, Second Edition, 1992.