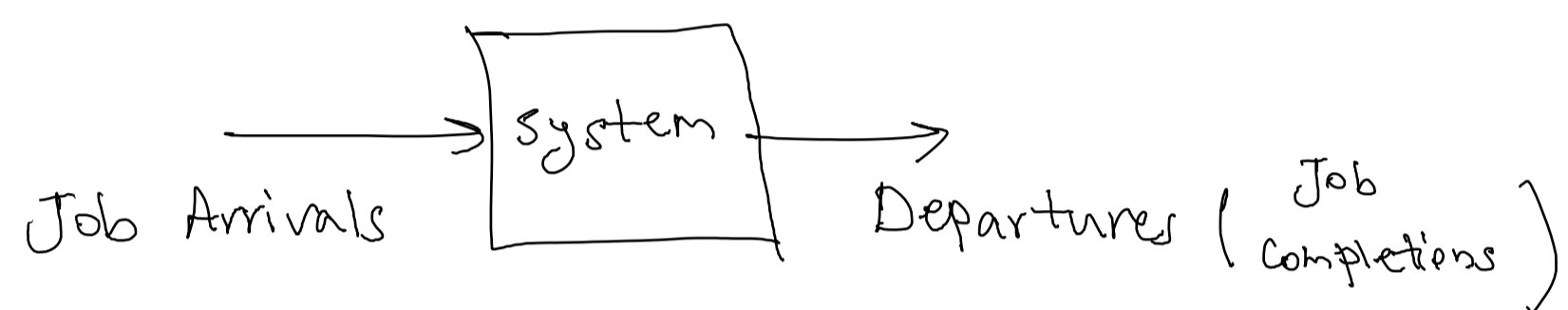


Fundamentals of performance evaluation

1- **Operational laws**: laws that do not require any assumptions about distribution of service times or inter-arrival times.



$A(t)$: # of arrivals in $[0, t)$

$D(t)$: # of departures in $[0, t)$

$B(t)$: total time duration that system is busy in $[0, t)$

$N(t)$: # of jobs in the system at time t (queue length / backlog)

Average quantities:

arrival rate $\lambda = \lim_{t \rightarrow \infty} \frac{A(t)}{t}$

throughput $r = \lim_{t \rightarrow \infty} \frac{D(t)}{t}$

utilization $\rho = \lim_{t \rightarrow \infty} \frac{B(t)}{t}$

mean service time $s = \lim_{t \rightarrow \infty} \frac{B(t)}{D(t)}$

average backlog: $N = \lim_t \frac{1}{t} \int_0^t N(s) ds$

* Utilization Law: $\rho = r s$

proof: $\frac{B(t)}{t} = \frac{D(t)}{t} \times \frac{B(t)}{D(t)}$ take the limit from both sides

* Job flow balance: For a well-behaved & stable system

$$\lambda = \nu$$

well-behaved: system does not create/destroy jobs

stable: backlog remains bounded $N < \infty$

proof: $N(t) = A(t) - D(t) + \underbrace{N(0)}_{\text{initial queue } < \infty}$

$\frac{1}{t} \int_0^t N(s) ds < \infty \implies N(t)$ is (essentially) bounded

$$\implies \lim_t \frac{N(t)}{t} = 0 = \lim_t \left(\frac{A(t)}{t} - \frac{D(t)}{t} + \frac{N(0)}{t} \right) = \lambda - \nu$$

Example: Consider a network gateway at which packets arrive at a rate of 125 packets/sec and the gateway takes an average of 6 milliseconds to forward them

$$\text{Throughput} = \text{Exit rate} = \text{Arrival rate} = 125 \text{ Pkt/sec} = \nu$$

$$\text{service time } S = 0.006 \text{ sec}$$

$$\text{Utilization } \rho = \nu S = 125 \times 0.006 = 0.75 = 75\%$$

"no assumptions on arrival or service process"

* Little's Law: average backlog = arrival rate \times average delay

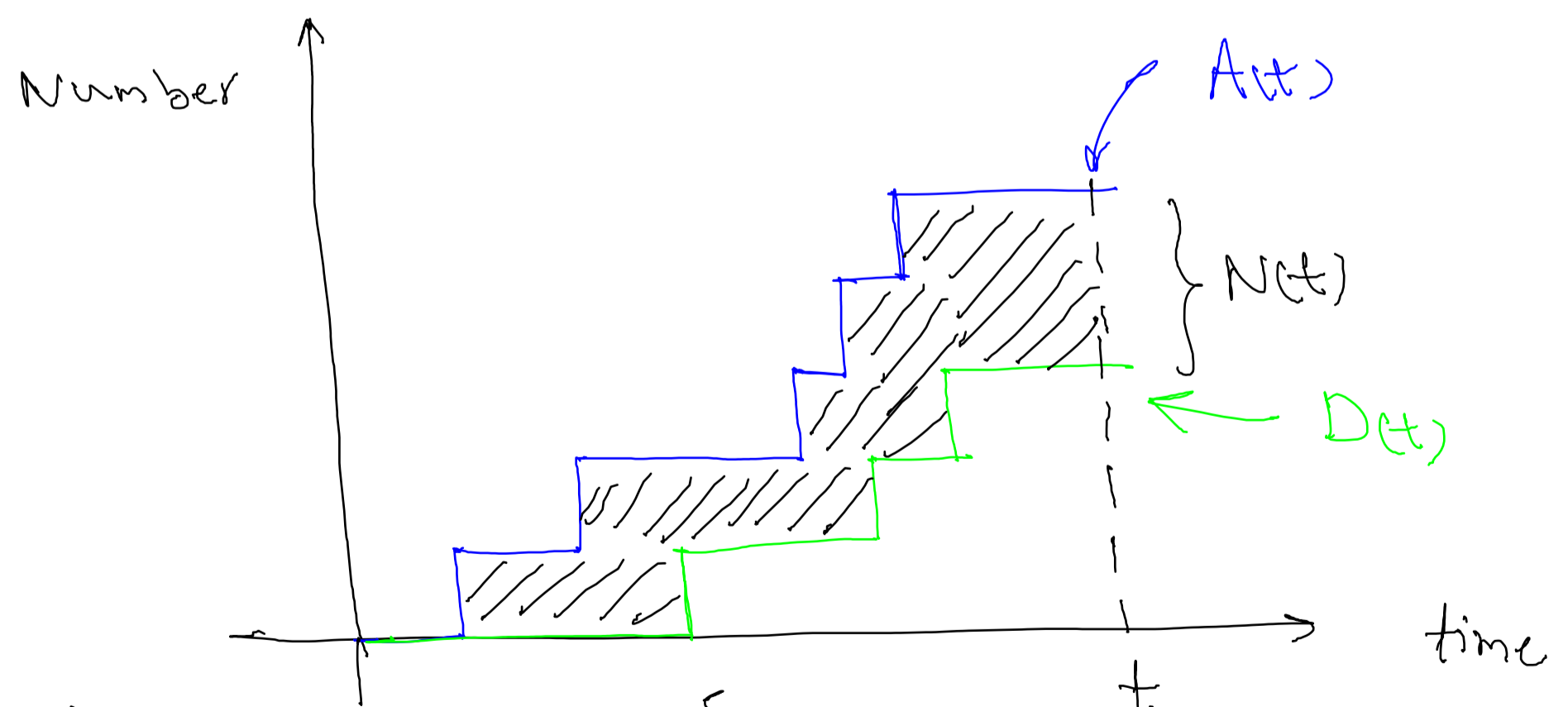
Formal statement: Let $T(i)$ be the time spent by i -th customer

in the system, average delay $T = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n T(i)$,

assume the limits T, N, λ exist, then

$$N = \lambda T$$

proof:



$$\text{Shaded area} = \int_0^t N(s) ds = \lambda_t \quad (\text{integral with vertical strips})$$

$$\sum_{i=1}^{D(t)} T(i) \leq \lambda_t \leq \sum_{i=1}^{A(t)} T(i) \quad (\text{integration with horizontal strips})$$

$$\text{But } \frac{1}{t} \gamma_t = \frac{1}{t} \int_0^t N(s) ds \longrightarrow N \quad \text{as } t \rightarrow \infty$$

$$\frac{1}{t} \sum_{i=1}^{A(t)} T(i) = \frac{A(t)}{t} \cdot \frac{1}{A(t)} \sum_{i=1}^{A(t)} T(i) \longrightarrow \lambda \cdot T \quad \text{as } t \rightarrow \infty$$

$$\frac{1}{t} \sum_{i=1}^{D(t)} T(i) = \frac{D(t)}{t} \cdot \frac{1}{D(t)} \sum_{i=1}^{D(t)} T(i) \longrightarrow r T \quad \text{as } t \rightarrow \infty$$

(because $A(t) \rightarrow \infty$ as $t \rightarrow \infty$ and $N < \infty \Rightarrow D(t) \rightarrow \infty$)

by flow balance ($N < \infty$) $\lambda = r$

$$\Rightarrow \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t N(s) ds = N = \lambda T$$

Note: The order of customer service is not essential for the validity of Little's law.

Example: Consider a transmission link, N_Q is the average # of pkts waiting in queue (not under transmission), λ is the arrival rate, average transmission time of a pkt is s . Then:

$$\text{average waiting time to start getting service: } W_Q = \frac{N_Q}{\lambda}$$

$$\text{average # of pkts under transmissio: } P = \lambda s$$

(applying Little's law only to pkts under transmission which is at most 1 pkt)

average waiting time to leave the link: $W = \frac{N}{\lambda} = \frac{Nq + e}{\lambda}$
 $= W_q + S$

"We can apply Little's law to smaller components of a system"