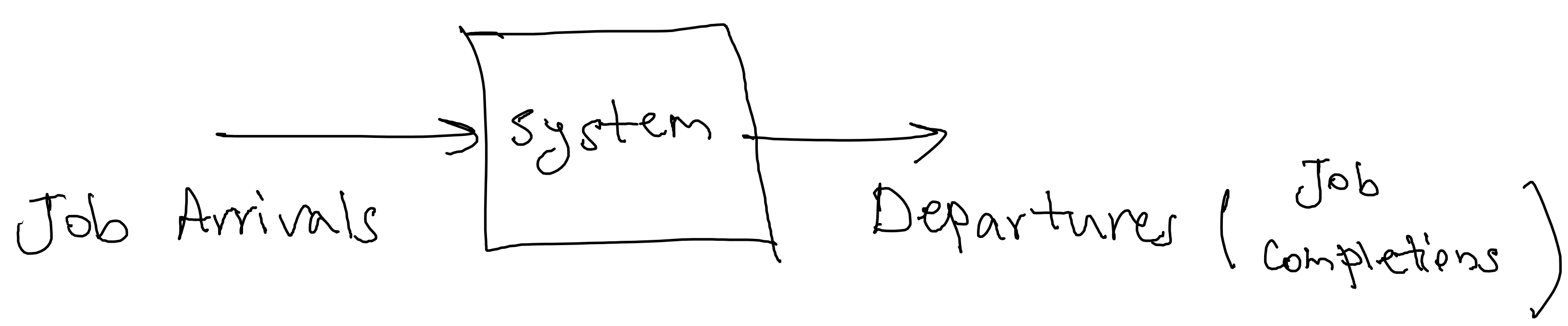
Lecturer : Javad Ghaderi

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 $\lambda = \lim_{t \to \infty} \frac{A(t)}{t}$

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

Utilization $\rho = \lim_{t \to \infty} \frac{B(t)}{t}$

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

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

* Utilization Law : $\boxed{\rho = r s}$

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

$$\boxed{\lambda = r}$$

<u>well-behaved:</u> system does not create/destroy jobs

<u>stable:</u> 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

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

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

Throughput = Exit rate = Arrival rate = 125 Pkt/sec $= r$

Service time $S = 0.006$ sec

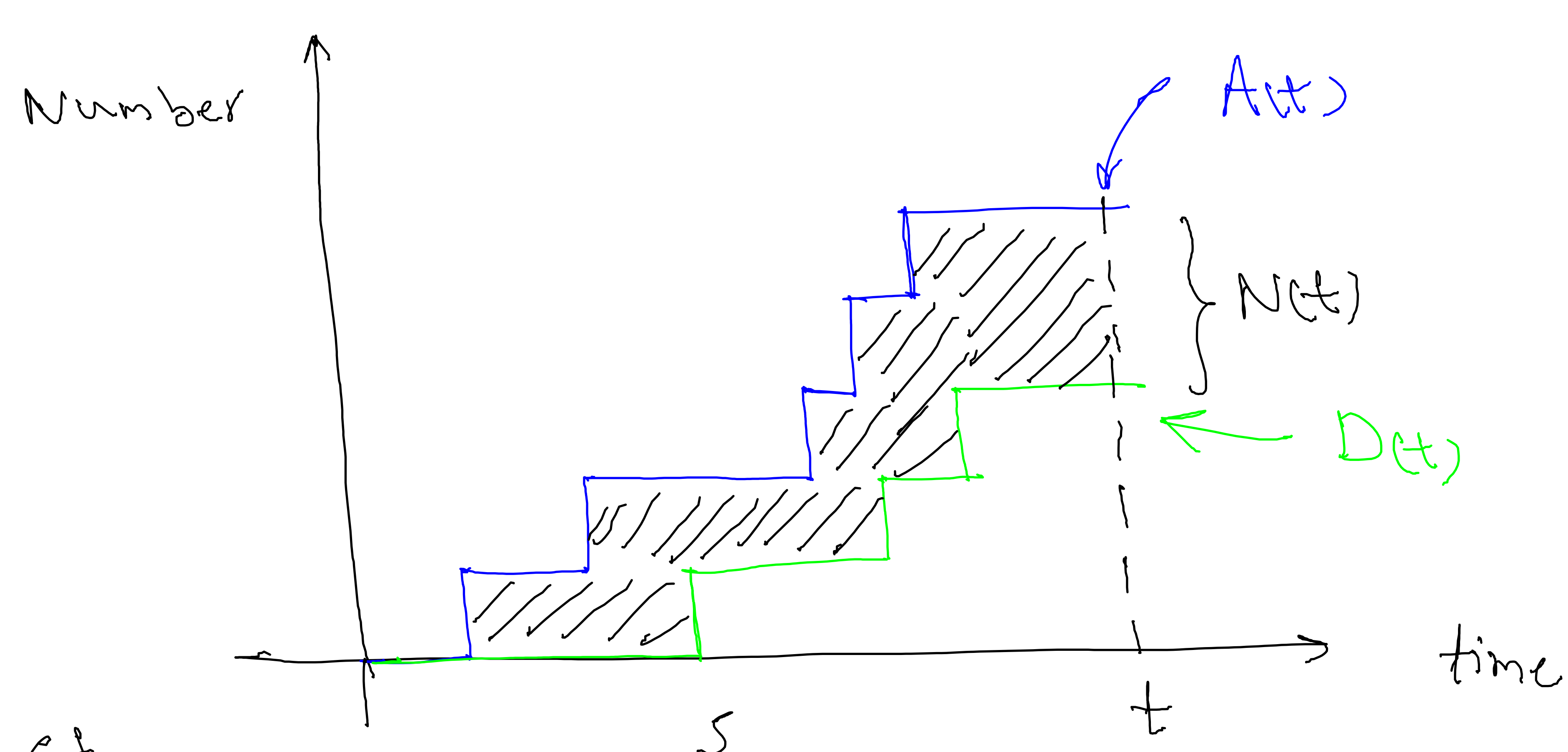
Utilization $\rho = rS = 125 \times 0.006 = 0.75 = 75\%$

"no assumptions on arrival or service process"

* Little's Law: $\boxed{\text{average backlog} = \text{arrival rate} \times \text{average delay}}$

Formal statement: Let $T(i)$ be the time spent by i-th customer in the system, average delay $T = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} T(i)$, assume the limits T, N, λ exist, then $\boxed{N = \lambda T}$.

proof:



Shaded area $= \int_0^t N(s)\, ds = \gamma_t$ (integral with vertical strips)

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

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

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

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

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

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

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

Note: The order of customer service is not essensial for the validity of little's Law.

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

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

average # of pkts under transmissio: $\rho = \lambda s$

(applying Littl's law only to pkts under transmission which is at most 1 pkt) average waiting time to leave the link: $W = \dfrac{N}{\lambda} = \dfrac{N_Q + \rho}{\lambda}$

$$= W_Q + S$$

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