Little's Law, Operational Laws

University of Tehran
Prepared by: Dr ahmad Khonsari
The University of Tehran
Excerpt from : **Performance Modeling and Design of Computer Systems**
Queueing Theory in Action
Mor Harchol-Balter

Little’s Law, Operational Laws

- **Theorem 6.1 (Little’s Law for Open Systems)** *For any ergodic open system we have that*
- $\mathbf{E}[N] = \lambda \mathbf{E}[T]$

where $\mathbf{E}[N]$ *is the expected number of jobs in the system,* λ *is the average arrival rate into the system, and* $\mathbf{E}[T]$ *is the mean time jobs spend in the system.*

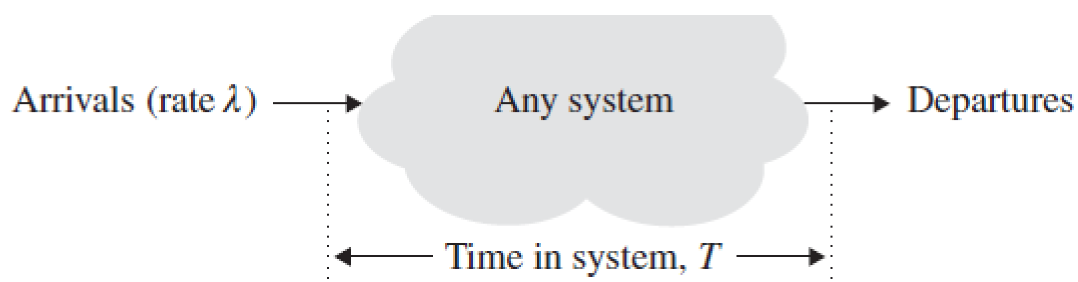


Figure 1. Setup for Little’s Law

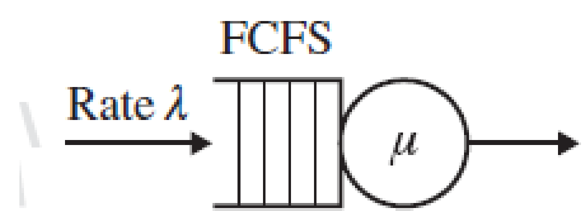
Little’s Law, Operational Laws

- Note: that Little’s Law makes no assumptions about the arrival process, the service time distributions at the servers, the network topology, the service order, or anything!
- In studying Markov chains, we see many techniques for computing $\mathbf{E}[N]$.
- Applying Little’s Law will then immediately yield $\mathbf{E}[T]$.

Little’s Law, Operational Laws

- Consider a single FCFS queue, shown in Fig 2.
- A customer arrives and sees **E**[N] jobs in the system.
- The expected time for each customer to complete is $1/\lambda$ (not $1/\mu$), since the average rate of completions is λ. Hence the expected time until the customer leaves is **E**[T] ≈ $(1/\lambda)$.**E**[N]
- **Fig2.** Little’s Law applied

to a single server



Little’s Law, Operational Laws

- **Theorem 2 (Little’s Law for Closed Systems)** *Given any ergodic closed system,*

$N = X \cdot \mathbf{E}[T]$,

where N is a constant equal to the multiprogramming level, X is the throughput (i.e., the rate of completions for the system), and $\mathbf{E}[T]$ *is the mean time jobs spend in the system.*

Little's Law, Operational Laws

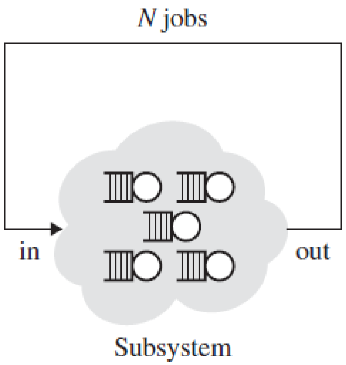
- Fig3 shows a batch system and an interactive (terminal-driven) system.
- Note that for the interactive system (right), the time in system, T, is the time to go from "out" to "out," whereas response time, R, is the time from "in" to "out."
- Specifically, for a *closed interactive system*, we define

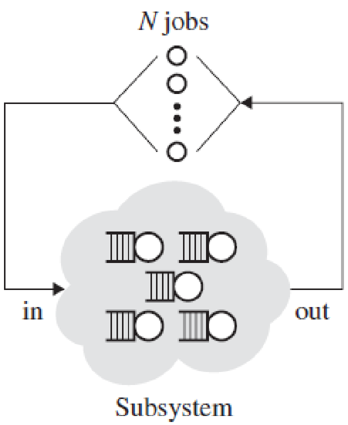
$\mathbf{E}[T] = \mathbf{E}[R] + \mathbf{E}[Z]$

where $\mathbf{E}[Z]$ is the average think time,

$\mathbf{E}[T]$ is the average time in system,

$\mathbf{E}[R]$ is the average response time.

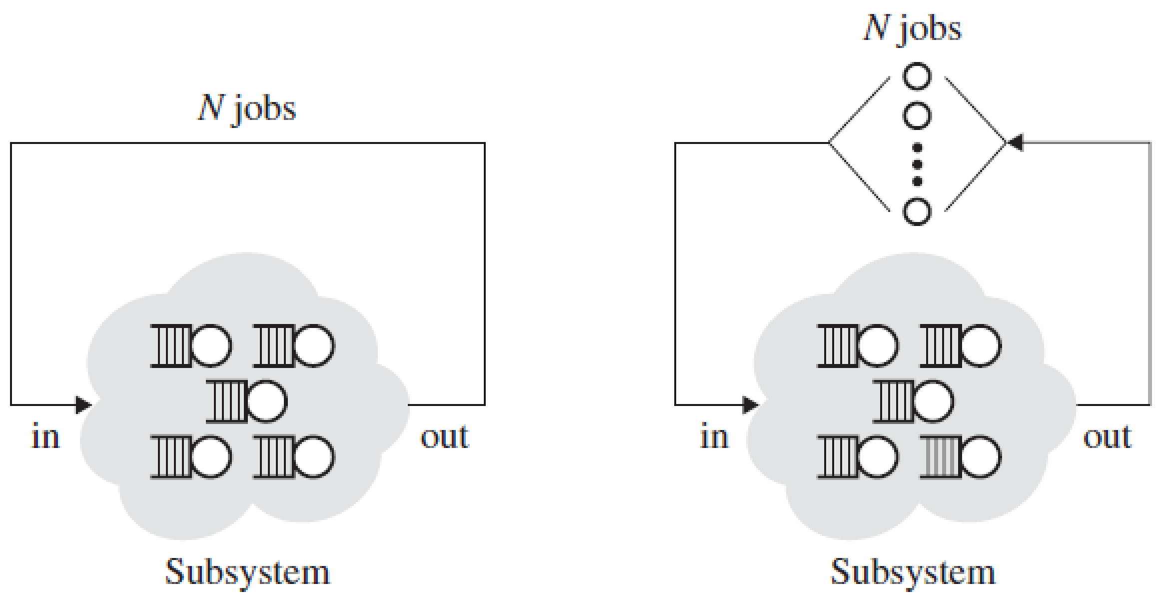




Little's Law, Operational Laws

- Note: for open systems and closed batch systems, we refer to **E**[T] as mean response time,
- whereas for closed interactive systems **E**[T] represents the mean time in system and **E**[R] is the mean response time, since response time does not include thinking.
- for an *open system*, throughput and mean response time are uncorrelated.
- By contrast, Little's Law tells us that, for a *closed* system, X and **E**[T] are inversely related, as are X and **E**[R].
- ***Thus in a closed system, improving response time results in improved throughput and vice versa.***

Little’s Law, Operational Laws



- **Fig3.** Closed systems: A batch system (left) and an interactive system (right).

Proof of Little's Law for Open Systems

Let

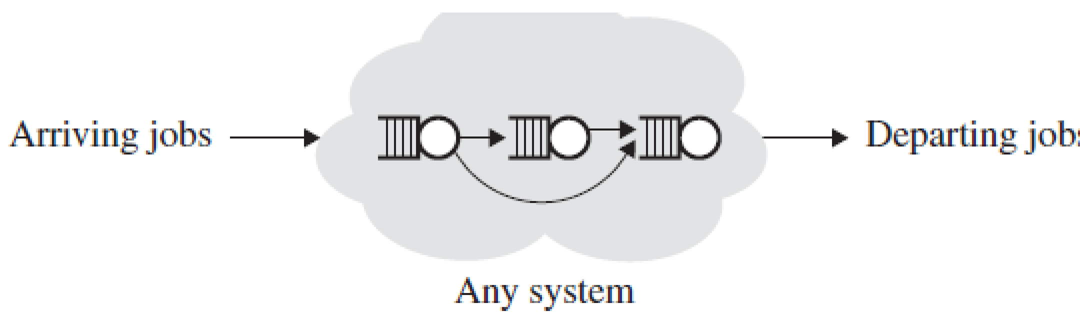
$A(t)$ = the number of arrivals by time t

$C(t)$ = the number of system completions (departures) by time t.

- Little's Law is actually stated as a relationship between time averages

Let $\lambda = \lim_{t\to\infty} \frac{A(t)}{t}$ and $X = \lim_{t\to\infty} \frac{C(t)}{t}$

- it is typically the case that $\lambda = X$ (one could have $\lambda > X$ if some arrivals get dropped, or if some jobs get stuck and never complete for some reason).



- **Figure 4.** Open system.

Proof of Little's Law for Open Systems

- **Theorem 3 (Little's Law for Open Systems Restated)** *Given any system where* $\overline{N}^{\text{Time Avg}}$*,* $\overline{T}^{\text{Time Avg}}$*,* λ*, and* X *exist and where* $\lambda = X$*, then*

$$\overline{N}^{\text{Time Avg}} = \lambda . \overline{T}^{\text{Time Avg}}$$

- It is an equality between *time averages*, not *ensemble averages*.(i.e. the time-average number in system for sample path ω is equal to λ times the time-average time in system for that sample path.)

Proof of Little’s Law for Open Systems

- For *ergodic*, the time average converges to the ensemble average with prob. 1; i.e., on almost every sample path, the time average on that sample path will be equal to the ensemble average over all paths.
- Thus, assuming ergodicity, we can apply Little’s Law in an ensemble-average sense, which we will do.

Proof of Little’s Law for Open Systems

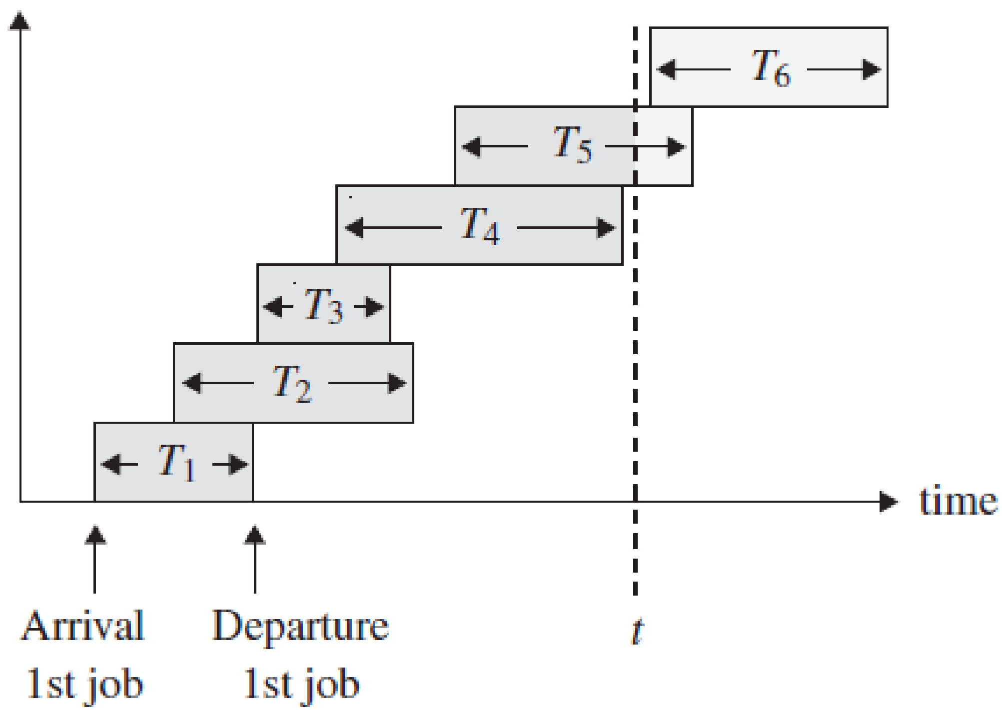
- The requirements in Theorem3 are all subsumed (induced, concluded) by the assumption that the system is *ergodic*, (in ergodic systems the above limits all exist)
- Also the average arrival rate and completion rate are equal, since the system empties infinitely often.
- Furthermore, in ergodic systems the time average is equal to the ensemble (or “true”) average.
- Thus it is sufficient to require that the system is ergodic for Little’s Law, as stated in Theorem1, to hold.

Proof of Little’s Law for Open Systems

- ***Proof*** **(*Theorem3*)**
- T_i :the time that the *i*th arrival to the system spends in the system.
- for any time *t*, consider the area, *A*, contained within all the rectangles in Figure 5, up to time *t* (this includes most of the rectangle labeled *T*5).
- 2 views of this area, *A*,
- by summing *horizontally*
- equivalently, by summing *vertically.*

Proof of Little's Law for Open Systems

- **Figure 5.** Graph of arrivals in an open system.
- the area under the curve= *A*,



Proof of Little's Law for Open Systems

the area under the curve= A,

$A(t)$ = the number of arrivals by time t

$C(t)$ = the number of system completions (departures) by time t.

- horizontal view: summing up the T_i's as follows:
- $\sum_{i \in C(t)} T_i \leq A \leq \sum_{i \in A(t)} T_i$

$\sum_{i \in C(t)} T_i$: sum of the time in system of those jobs that have completed by time t,

$\sum_{i \in A(t)} T_i$: sum of the time in system of those jobs that have arrived by time t.

Proof of Little's Law for Open Systems

- vertical view of A: adds up the number of jobs in system at any moment in time, $N(s)$, where s ranges from 0 to t. =

$$A = \int_0^t N(s)\mathrm{ds}$$

Combining these two views, we have

$$\sum_{i \in C(t)} T_i \leq \int_0^t N(s)\mathrm{ds} \leq \sum_{i \in A(t)} T_i$$

Proof of Little's Law for Open Systems

- Dividing by t throughout, we get

$$\frac{\sum_{i \in C(t)} T_i}{t} \le \frac{\int_0^t N(s)\mathrm{ds}}{t} \le \frac{\sum_{i \in A(t)} T_i}{t}$$

- or, equivalently

$$\frac{\sum_{i \in C(t)} T_i}{C(t)} \frac{C(t)}{t} \le \frac{\int_0^t N(s)\mathrm{ds}}{t} \le \frac{\sum_{i \in A(t)} T_i}{A(t)} \frac{A(t)}{t}$$

Proof of Little's Law for Open Systems

- Taking limits as $t \to \infty$,

$$\lim_{t\to\infty} \frac{\sum_{i\in C(t)} T_i}{C(t)} \lim_{t\to\infty} \frac{C(t)}{t} \leq \overline{N}^{\text{TimeAvg}}$$

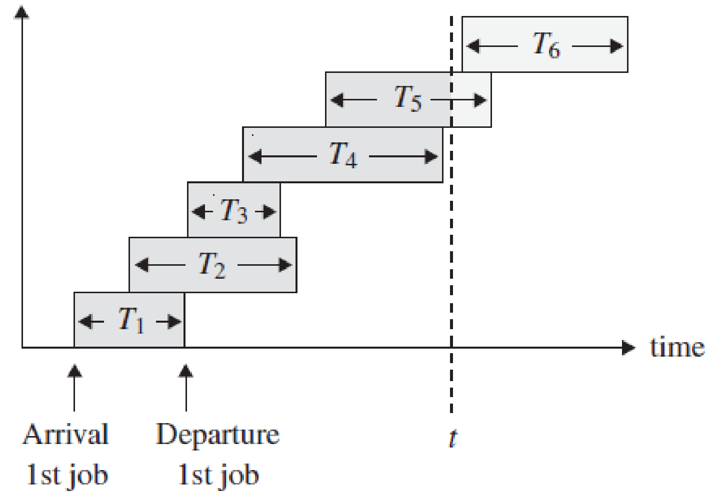
$$\leq \lim_{t\to\infty} \frac{\sum_{i\in A(t)} T_i}{A(t)} \lim_{t\to\infty} \frac{A(t)}{t}$$

$\rightarrow \overline{T}^{\text{Time Avg}} . X \leq \overline{N}^{\text{Time Avg}} \leq \overline{T}^{\text{Time Avg}} . \lambda$

we are given that X and λ are equal. Therefore

$\overline{N}^{\text{Time Avg}} = \lambda . \overline{T}^{\text{Time Avg}}$ ▪

Little’s Law for Open Systems



- **Q:** Does this argument depend on service order?

A: No. Observe that the second arrival departs *after* the third arrival departs.

- **Q:** Does this argument depend on number of servers?

A: No, this argument holds for any system.

- You may apply the little’s law to any parts of the system including the **server** and **queue** itself.

Little's Law for Open Systems-Queue

- **Corollary 4 (Little's Law for Time in Queue)** *Given any system where $\overline{N}_Q^{\text{Time Avg}}$, $\overline{T}_Q^{\text{Time Avg}}$, λ, and X exist and where $\lambda = X$, then*

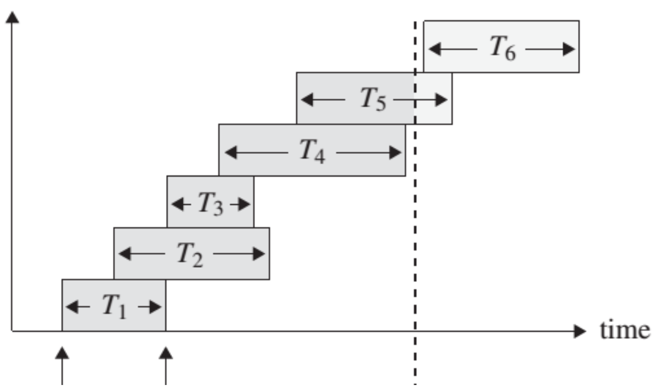
$\overline{N}_Q^{\text{Time Avg}} = \lambda . \overline{T}_Q^{\text{Time Avg}}$

where N_Q represents the number of jobs in queue in the system and T_Q represents the time jobs spend in queues.

- $proof$: Same proof as for Theorem 3, except that now instead of drawing T_i, we draw $T_Q(i)$, i.e. the time of the ith arrival in queues (wasted time).

- Note that $T_Q(i)$ may not be a solid rectangle.

It may be made up of **several rectangles** because the ith job might be in queue for a while, then in service, then waiting in some other queue, then in service, again, etc.



Little's Law for Open Systems- Utilization Law

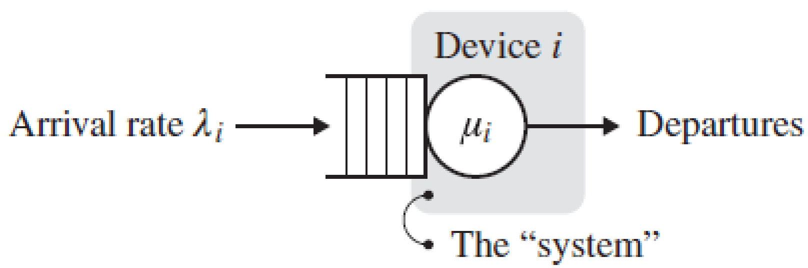
- **Corollary 5 (Utilization Law)** *Consider a single device i with average arrival rate λ_i jobs/sec and average service rate μ_i jobs/sec, where $\lambda_i < \mu_i$.*
- *Let ρ_i denote the long-run fraction of time that the device is busy.*
- *Then* $\rho_i == \dfrac{\text{Average service time required by a job}}{\text{Average time between arrivals}}$

$$= \frac{1/\mu_i}{1/\lambda_i} = \frac{\lambda_i}{\mu_i}$$

- We refer to ρ_i as the "device utilization" or "device load."
- Observe that, given ergodicity, ρ_i represents both the long-run fraction of time (time average) that device *i* is busy and also the limiting probability (ensemble average) that device *i* is busy.

Little's Law for Open Systems- Utilization Law

- ***Proof:*** Let the "system" consist of the server, as shown in the shaded box of Figure 6. Now the number of jobs in the "system" is always just 0 or 1.



- **Figure 6.** Using Little's Law to prove the Utilization Law

Little's Law for Open Systems-server

- **Question:** What is the expected number of jobs in the system (server) as we have defined it?
- **Answer:** The number of jobs in the system is 1 when the device is busy (this happens with probability ρ_i) and is 0 when the device is idle (with probability $1 - \rho_i$).
- Hence the ***expected* number** of jobs in the **server=** $((1 \text{x} \rho_i + 0 \text{x}(1 - \rho_i))) = \boldsymbol{\rho_i}$.
- So, applying Little's Law, we have

ρ_i = Expected number jobs in service facility for device *i*

= (Arrival rate into service facility) · (Mean time in service facility)

$= \lambda_i \cdot \mathbf{E}[\text{Service time at device } i] = \lambda_i \cdot \dfrac{1}{\mu_i}$ ▪

Little's Law for Open Systems-server

- We often express the Utilization Law as

$\rho_i = \lambda_i \mathbf{E}[S_i] = X_i \mathbf{E}[S_i]$

- where ρ_i, λ_i, X_i, and $\mathbf{E}[S_i]$ are the load, average arrival rate, average throughput, and average service requirement at device i, respectively.
- **Question:** Suppose we are only interested in "red" jobs, where "red" denotes some type of jobs. Can we apply Little's Law to just "red" jobs? Prove it.
- **Answer:** Yes.
- **E**[Number of red jobs in system] = $\lambda_{\text{red}} \cdot$ **E**[Time spent in system by red jobs]
- The proof is exactly the same as before, but only the T_i's corresponding to the red jobs are included in Figure 5.