Little's Law

Theorem 6.1 (Little’s Law for Open Systems)

For any ergodic open system, we have that

\[ E[N] = \lambda E[T], \]

where

- \( E[N] \) is the expected # jobs in the system
- \( \lambda \) is the avg. arrival rate
- \( E[T] \) is the mean time a job spends in the system

Fig. 6.1 [14]
Fig. 6.2 [H]
Little's Law for Closed Systems (6.3 [H])

**Theorem 6.2 (Little’s Law for Closed Systems)** Given any ergodic closed system,

\[ N = X \cdot E[T], \]

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

\[ E[T] = E[R] + E[I] \]

- \( E[T] \): time in system
- \( E[R] \): response time
- \( E[I] \): think time
6.4 [t7] Proof of Little's Law for Open Systems

**Theorem 6.3 (Little's Law for Open Systems Restated)**  Given any system where 
\[
\frac{N}{T \text{ Time Avg}} = \lambda \cdot \frac{X}{T \text{ Time Avg}}.
\]

\[
\lambda = \lim_{t \to \infty} \frac{A(t)}{t} \quad \text{and} \quad X = \lim_{t \to \infty} \frac{C(t)}{t}.
\]

\[\lambda = X, \text{ if jobs are not dropped, except in special cases.}\]

Ergodicity implies the assumptions of Thm. 9.3.
Graph of arrivals in an open system (Fig. 6.5 [T])
$\sum_{i \in C(t)} T_i \leq A \leq \sum_{i \in A(t)} T_i$

- Completed by time $t$
- Arrived by time $t$

$q = \int_0^t N(s)\,ds$ ("vertical view": sum # jobs in system at any moment in time)

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

or equivalently

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

$\sum_{i \in C(t)} T_i \leq \int_0^t N(s)\,ds \leq \sum_{i \in A(t)} T_i$
Taking limits as $t \to \infty$,

\[
\lim_{t \to \infty} \sum_{i=0}^{T_i} \frac{C(t)}{t} \leq \overline{N \text{ Time Avg}} \leq \lim_{t \to \infty} \sum_{i=0}^{T_i} \frac{Q(t)}{t}
\]

\[
\overline{\text{Time Avg}} \cdot X \leq \overline{N \text{ Time Avg}} \leq \overline{\text{Time Avg}} \cdot d
\]

Avg time completion in system

Avg rate

Avg arrival rate

Avg system rate

Avg viewing arrivals

Since $d = X$, \[\overline{N \text{ Time Avg}} = \overline{\text{Time Avg}}\]
Corollary 6.4 (Little's Law for Time in Queue) Given any system where \( \overline{N_Q^{\text{Time Avg}}} \), \( \overline{T_Q^{\text{Time Avg}}} \), \( \lambda \), and \( X \) exist and where \( \lambda = X \), then

\[
\overline{N_Q^{\text{Time Avg}}} = \lambda \cdot \overline{T_Q^{\text{Time Avg}}},
\]

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

The same kind of "geometric" proof can be carried out, except that now the "rectangles" \( T_Q(i) \) represent time in queue for job \( i \), and they can be broken up as jobs leave one queue and enter a processor.
Corollary 6.5 (Utilization Law) Consider a single device $i$ with average arrival rate $\lambda_i$ jobs/sec and average service rate $\mu_i$ jobs/sec, where $\lambda_i < \mu_i$. Let $\rho_i$ denote the long-run fraction of time that the device is busy. Then

$$\rho_i = \frac{\lambda_i}{\mu_i}.$$
The expected number of jobs in the system is
\[ L_i = 1 \times P_i \text{ system is busy} + 0 \times \left(1 - P_i\right) = P_i. \]

So, applying Little's Law, we have:

\[ L_i = \text{Expected number of jobs in the system} = \]
\[ = (\text{arrival rate in the system}) \times (\text{mean time in the system}) = \]
\[ = \lambda_i \times E\left[\text{service time at device } i\right] = \lambda_i \times \frac{1}{\mu_i}. \]

The Utilization Law is also written

\[ U_i = \lambda_i E\left[ S_i \right] = X_i E\left[ S_i \right]. \]
6.5 [TI] Proof of Little's Law for Closed Systems

Theorem 6.6 (Little's Law for Closed Systems Restated) Given any closed system (either interactive or batch) with multiprogramming level $N$ and given that $\frac{1}{T_{\text{TimeAvg}}}$ and $X$ exist and that $\lambda = X$, then

$$N = X \cdot \frac{1}{T_{\text{TimeAvg}}},$$

$$X = \lim_{t \to \infty} \frac{C(t)}{t},$$

where $C(t)$ is the number of system completions by time $t$.

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

where $Q(t)$ is the number of jobs generated by time $t$. (Note: not the number of arrivals.)
$N = 3$

- $T_3$
- $T_8$
- $T_{11}$
- $T_2$
- $T_6$
- $T_{10}$
- $T_1$
- $T_4$
- $T_5$
- $T_7$
- $T_9$
- $T_{12}$

$t$
6.6 [H] Generalized Little's Law

Little's Law has been generalized to higher moments, e.g., $E[N^2]$, $E[T^2]$, but only under restrictive conditions, such as a system with a single FCFS queue.
6.7 [H] Examples applying Little’s Law

Example 1 (Closed Interactive System)

What is the throughput, $X$, of the system?

$N = X \times E[T] = X \times (E[2] + E[R])$

$\Rightarrow X = \frac{N}{E[2] + E[R]} = \frac{10}{5 + 15} = \frac{1}{2}$ (bb/s)

Response Time Law for Closed Systems:

$E[R] = \frac{N}{X} \times E[2]$

$E[2] = 5$ sec

$E[R] = 15$ sec

Fig. 6.8 [H]
Example 2: A more complex interactive system

- $X_{disk3} = \frac{40 \, \text{requests}}{\text{sec}}$
- $E[S_{disk3}] = 0.0225 \, \text{sec}$
- $E[N_{disk3}] = 40 \, \text{jobs}$

What is the utilization of Disk 3?

- $E_{disk3} = X_{disk3} \cdot E[S_{disk3}] = 40 \cdot 0.0225 = 9.0\%$

Figure 6.9 [H]
What's the mean time spent querying at disk 3?

$T_{disk3}$ is the time spent querying plus serving at disk 3.

$T_{disk3} = T_{Q} + S_{disk3}$, where $T_{Q}$ is the queue at disk 3.

$E[T_{disk3}] = \frac{E[N_{disk3}]}{C_{disk3}} = \frac{4}{40} = 0.1 \text{ sec}$

$E[T_{Q}] = E[T_{disk3}] - E[S_{disk3}] = 0.1 - 0.025 = 0.075 \text{ sec}$

Find the number of requests queued at disk 3.

$E[N_{Q}] = E[N_{disk3}] - E[\text{Number served at disk 3}] = 4 - E_{disk3} = 4 - 0.9 = 3.1$
Alternatively, use Little’s Law on the queue at Disk3:

\[ E[N_{\text{Disk3}}] = E[T_{\text{Disk3}}] \times \text{Disk3} = 0.075 \times 40 = 3.1 \]

What is the system throughput?

\[
\begin{align*}
X &= \frac{N}{E[R] + E[S]} = \frac{10}{E[R] + 5} \\
E[R] &= \frac{E[N_{\text{not-thrashing}}]}{X} = \frac{7.5}{X} \\
\Rightarrow & \quad X = 5 \quad E[R] = 15
\end{align*}
\]
Example 3: A finite buffer

\[
\lambda = 3 \quad \mu = 4
\]

7 jobs in system; E[T] = \frac{1}{\mu - \lambda} \text{ E[N]} = \lambda (1 - P(N > 0)) = \lambda \left(1 - \frac{\mu}{\lambda + \mu} \right)

If \lambda > \mu, so Little's Law does not apply to the finite buffer system.
However, the rate of jobs that get through is \lambda (1 - P(N > 0) jobs in the system).
This is the effective arrival rate. Little's Law applies with the effective arrival rate:

\[ E[N] = \lambda (1 - P(N > 0)) \text{ E}[T] \]
6.8 More operational laws: the forced flow law

\[ X_i = E[V_i] \cdot X \]

- \( X \) is the system throughput
- \( X_i \) is the device throughput
- \( V_i \) is the number of visits to device \( i \) per job.

**Fig. 6.11**
Example of Forced Flow Law

\[ C_a = C_{cpu} \cdot \frac{80}{181} \]
\[ C_b = C_{cpu} \cdot \frac{100}{181} \]
\[ C = C_{cpu} \cdot \frac{1}{181} \]
\[ C_{cpu} = C_a + C_b + C. \quad \text{So,} \]
\[ E[V_a] = E[V_{cpu}] \cdot \frac{80}{181} \]
\[ E[V_b] = E[V_{cpu}] \cdot \frac{100}{181} \]
\[ 1 = E[V_{cpu}] \cdot \frac{1}{181} \]
\[ E[V_{cpu}] = E[V_a] + E[V_b] + 1 \]

\[ \Rightarrow E[V_{cpu}] = 181, \quad E[V_a] = 80, \quad E[V_b] = 100. \]

Fig. 6.12 [H] Calculating the visit ratios
6.9[17] Composing operational laws

Simple Example

\( N=25 \) (25 terminals), 18 sec avg think time \((E[2]=18)\)

20 visits per second on avg. to a specific disk \((E[U_{disk}]=20)\)

30\% utilization of that disk \((E[disk]=0.3)\)

0.025 sec avg. service time per visit to that disk \((E[S_{disk}]=0.025)\)

What is the mean response time \((E[R]=E[T]-E[2])\)?

The Response Time Law for Closed System states:
\[ E[Q] = \frac{N}{X} = E[Z] \Rightarrow N = 25, \ E[Z] = 18. \ X \ ? \]

The Forced Flow Law states:

\[ = \frac{25}{0.6} - 18 = 41.7 - 18 = 23.7 \text{ sec} \]

\[ X_i = E[V_i]. X \Rightarrow X = \frac{X_{\text{disk}}}{E[V_{\text{disk}}]} E[V_{\text{disk}}] = 20. \ X_{\text{disk}} ? \]

The Utilization Law states:

\[ E_i = \frac{X_{\text{disk}}}{E[V_{\text{disk}}]} = \frac{1.2}{20} = 0.06 \text{ interactions/sec} \]

\[ E_i = \frac{\lambda}{\mu}, \text{ or } (p. 101) \]

\[ E_i = X_i E[S_i], \ i.e. X_{\text{disk}} = \frac{E_{\text{disk}}}{E[S_{\text{disk}}]} = \frac{0.2}{0.025} = 1.2 \text{ request/sec} \]

"Working backwards"
Header example
(Lozowska et al.)

$N = 23$

$E[27] = 21 \text{ sec}$

$x = 0.45 \text{ interactions per second}$

$E[N \text{ getting memory}] = 11.65$

$E[V_{\text{CPU}}] = 3 \text{ visits to CPU per interaction}$

$E[S_{\text{CPU}}] = 0.21 \text{ sec}$
What is the average amount of time that elapses between getting a memory partition and completing the interaction?

\[ E[\text{Time in Central Subsystem}] = E[\text{Response Time}] - E[\text{Time to get memory}] \]

By the Response Time Law,

\[ E[\text{Response Time}] = \frac{N}{\lambda} \cdot E[\bar{X}] = \frac{23}{0.45} \cdot 21 \approx 51.11 \text{ sec} \]

By Little's Law for closed systems \((N = \lambda \bar{X})\),

\[ E[\text{Time to get memory}] = \frac{E[\text{Number Getting Memory}]}{\bar{X}} = \frac{11.65}{0.45} \approx 25.88 \text{ sec} \]
\( \text{What is the CPU utilization?} \) By the utilize the law (version of p. 101):

\[
e_{\text{cpu}} = X_{\text{cpu}} \cdot E[S_{\text{cpu}}] = (\text{ Forced Flow Law })
\]

\[
= X \cdot E[V_{\text{cpu}}] \cdot E[S_{\text{cpu}}] = 0.45 \cdot 3 \cdot 0.21 \approx 0.28
\]
Define $D_i$ as the total demand of one job to device $i$:

$$D_i = \sum_{j=1}^{V_i} S_{i,j}^{(s)}$$

where $S_{i,j}^{(s)}$ is the time required by the $j$-th visit of a job to device $i$.

$E[D_i] = E[V_i] \cdot E[S_i]$ if $V_i$ and $S_{i,j}^{(s)}$ are independent.

To compute $E[D_i]$:

$$E[D_i] = \frac{B_i}{C} = \frac{\text{total busy time of device } i}{\text{number of system completions in time } t}.$$
Utilization Law (3rd version, p. 101)

\[ e_i = X_i \cdot E[S_i] = X \cdot E[V_i], \quad E[S_i] = X \cdot E[D_i] \]

\[ \rho_i = X \cdot E[D_i] \]

The Bottleneck Law