به نام خدا

Performance Evaluation of Computer Systems

Dr. Ali Movaghar

Fall 2022

Performance Modeling and Design of Computer Systems

6- LITTLE'S LAW AND OTHER OPERATIONAL LAWS

Setup for Little's Law for an Open System

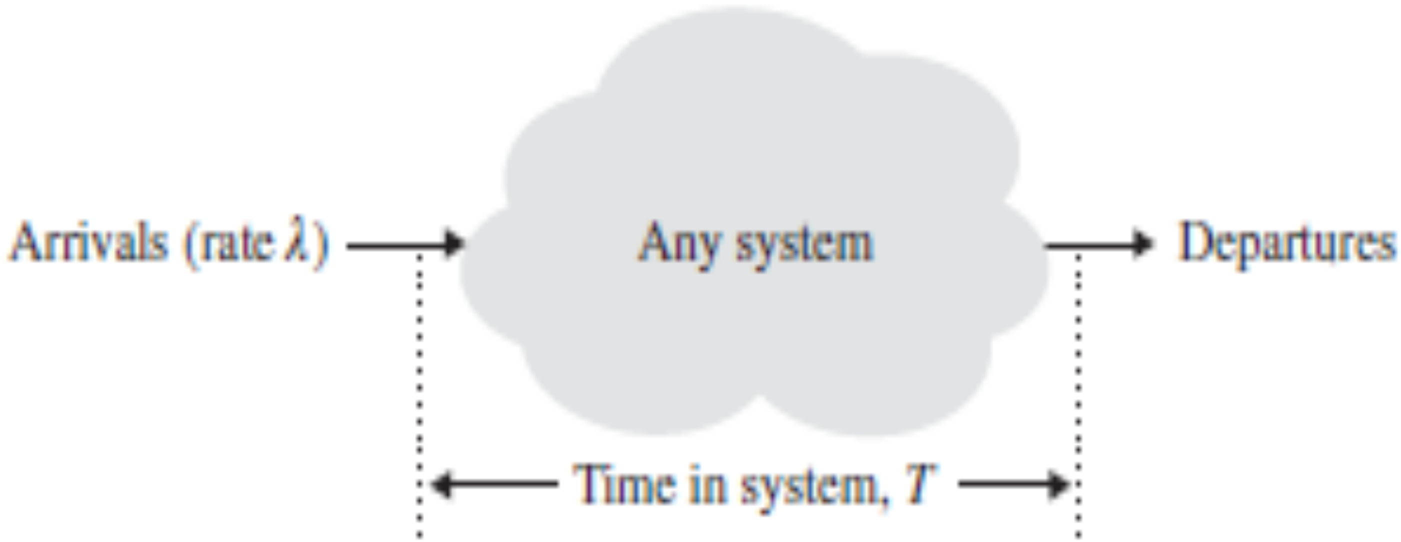


Figure 6.1. Setup for Little's Law.

Theorem 6.1: Little's Law for Open Systems

- For any ergodic open system

$$\mathbf{E}[N] = \lambda \mathbf{E}[T]$$

  - E[N] is the expected number of jobs in the system
  - λ is the average arrival rate into the system
  - E[T] is the mean time jobs spend in the system

Theorem 6.2: Little's Law for Closed Systems

- Given any ergodic closed system

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

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

Theorem 6.3: Little's Law for Open Systems Restated

- Given any system where
  - $\overline{N}^{Time\ Avg}, \overline{T}^{Time\ Avg}, \lambda = X$ and
  - $\lambda = \lim_{t\to\infty} \frac{A(t)}{t}$ and $X = \lim_{t\to\infty} \frac{C(t)}{t}$,
  - then

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

Graph of arrivals in an open system

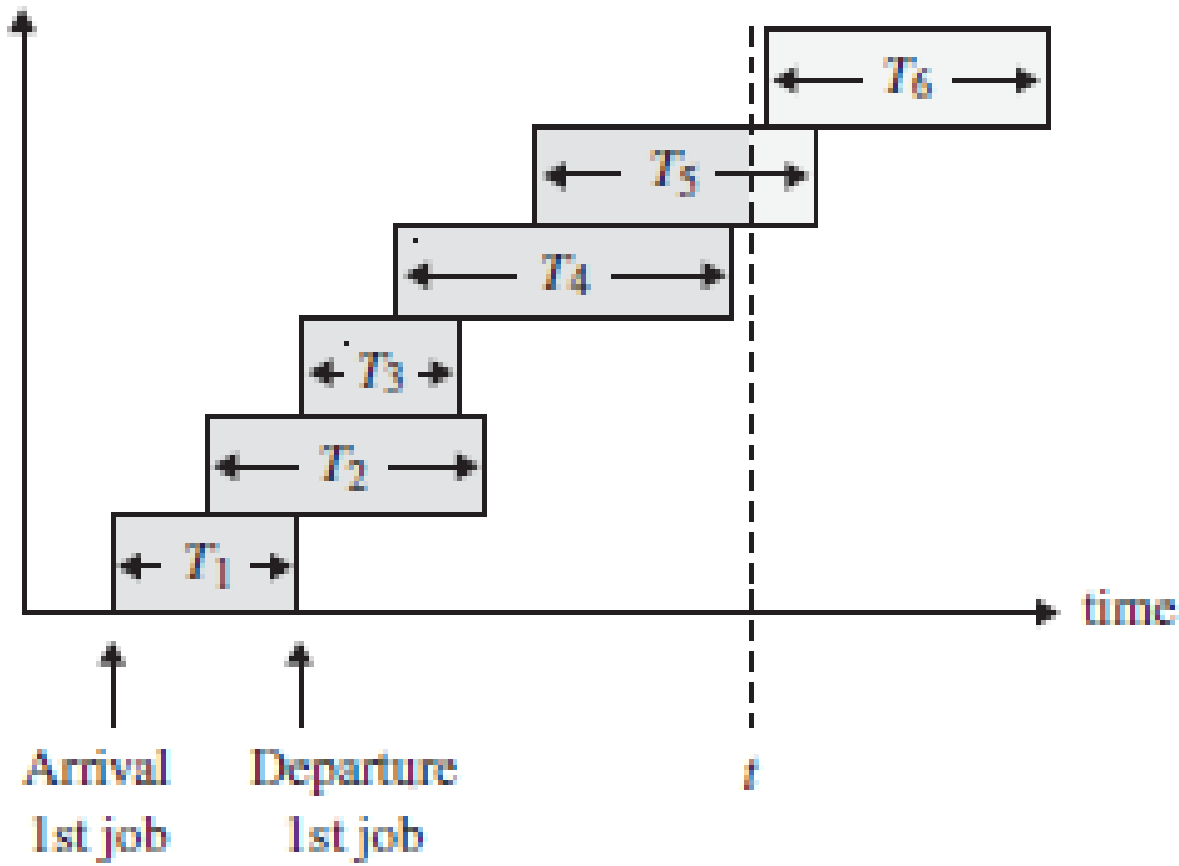


Figure 6.5. Graph of arrivals in an open system.

Proof:

The vertical view of $\mathcal{A}$ adds up the number of jobs in system at any moment in time, $N(s)$, where s ranges from 0 to t. Thus,

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

Combining these two views, we have

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

Dividing by t throughout, we get

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

or, equivalently,

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

Taking limits as $t \to \infty$,

$$\lim_{t \to \infty} \frac{\sum_{i \in C(t)} T_i}{C(t)} \cdot \lim_{t \to \infty} \frac{C(t)}{t} \le \overline{N}^{\text{Time Avg}} \le \lim_{t \to \infty} \frac{\sum_{i \in A(t)} T_i}{A(t)} \cdot \lim_{t \to \infty} \frac{A(t)}{t}$$

$$\Rightarrow \overline{T}^{\text{Time Avg}} \cdot X \le \overline{N}^{\text{Time Avg}} \le \overline{T}^{\text{Time Avg}} \cdot \lambda.$$

Yet we are given that X and λ are equal. Therefore,

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

Corollary 6.4: Little's Law for Time in Queue

- Given any system where
  - $\overline{N}_Q^{\,Time\ Avg}, \overline{T}_Q^{\,Time\ Avg}, \lambda = X$ and
  - Then $\overline{N}_Q^{\text{Time Avg}} = \lambda \cdot \overline{T}_Q^{\text{Time Avg}}$,
- N_Q is the number of jobs in queue and T_Q represents the time jobs spend in queues

Corollary 6.5: 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

$$\rho_i = \frac{\lambda_i}{\mu_i}.$$

- We refer to ρ_i as the “device utilization” or “device load.”

Utilization Law

- We often express the Utilization Law as

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

- where ρ_i , λ_i , X_i , and $E[S_i]$ are the load, average arrival rate, average throughput, and average service requirement at device i, respectively

Theorem 6.6: Little's Law for close Systems Restated

- Given any closed system (either interactive or batch) with multiprogramming level N and given that
  - $\overline{T}^{Time\ Avg}$, $\lambda = X$ and
  - $\lambda = \lim_{t \to \infty} \frac{A(t)}{t}$ and $X = \lim_{t \to \infty} \frac{C(t)}{t}$,
  - then $N = X \cdot \overline{T}^{\text{Time Avg}}.$

Batch and Interactive Systems

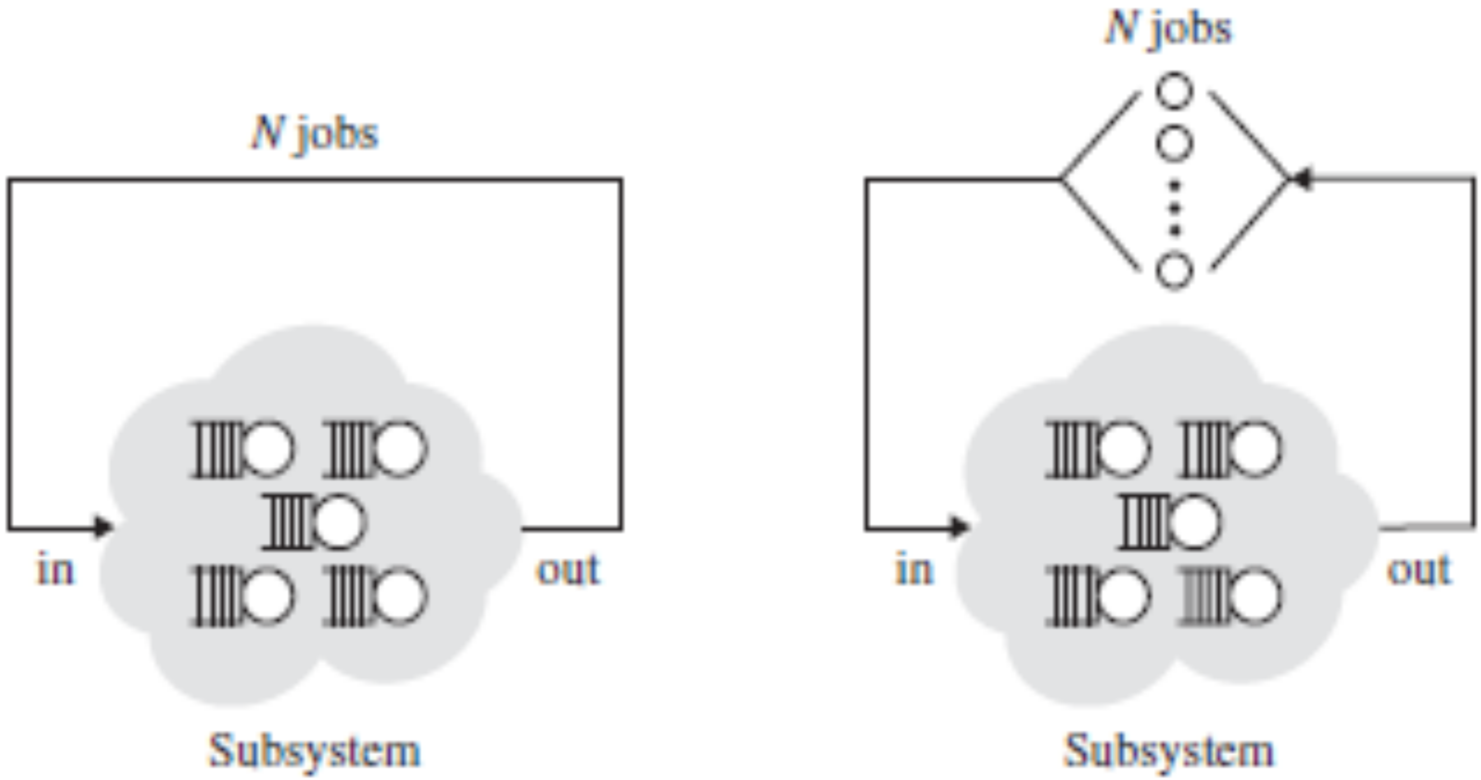


Figure 6.3. Closed systems: A batch system (left) and an interactive system (right).

Closed Interactive Systems

- For closed interactive systems, the time in the system, T, is the time to go from "out" to "out," whereas response time, R, is the time to go from "in" to "out."
- More specifically, we define **E**[T] = **E**[R] + **E**[Z], where **E**[Z] is the average think time, **E**[T] is the average time in the system, and **E**[R] is the average response time.

Graph of job system times in closed system with N = 3

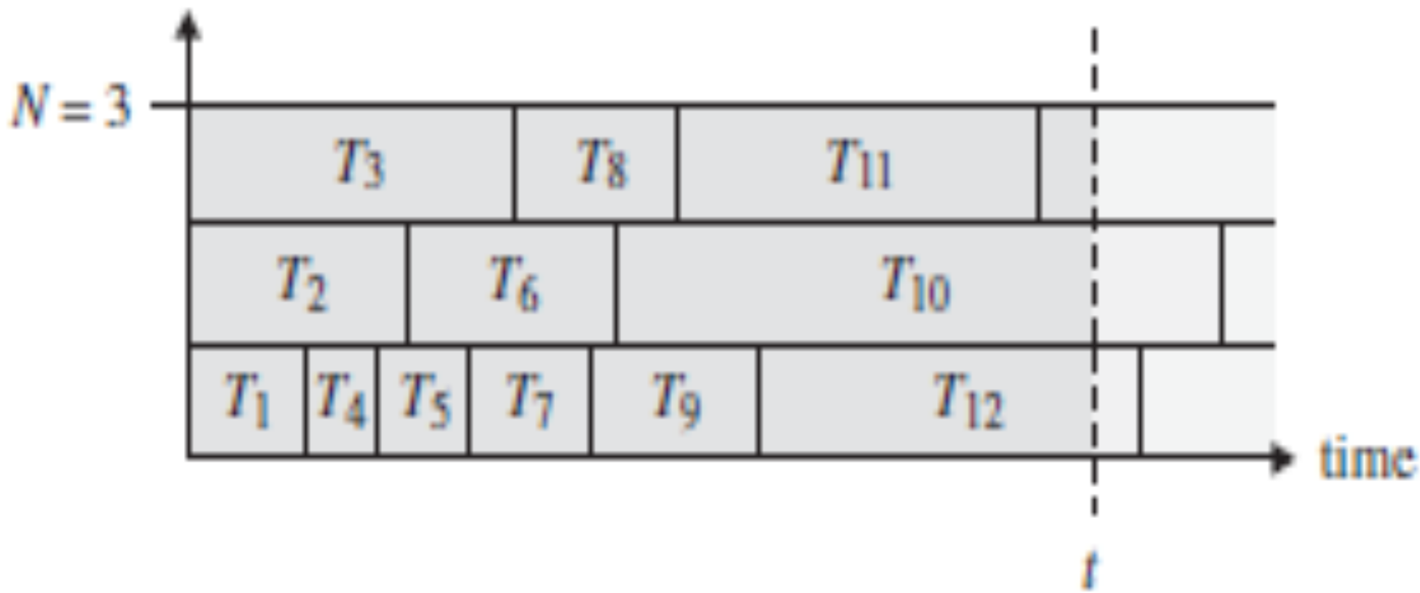


Figure 6.7. Graph of job system times in closed system with $N = 3$.

Proof:

Proof Figure 6.7 shows the time in system for arrivals. Observe that a new job cannot arrive until one departs. Thus there are always N jobs in the system. At any time t, the area made up by all rectangles up to time t is Nt, which can be bounded above and below as follows:

$$\sum_{i \in C(t)} T_i \leq N \cdot t \leq \sum_{i \in A(t)} T_i$$

$$\Rightarrow \frac{\sum_{i \in C(t)} T_i}{t} \leq N \leq \frac{\sum_{i \in A(t)} T_i}{t}$$

$$\Rightarrow \frac{\sum_{i \in C(t)} T_i}{C(t)} \cdot \frac{C(t)}{t} \leq N \leq \frac{\sum_{i \in A(t)} T_i}{A(t)} \cdot \frac{A(t)}{t}$$

Taking limits as $t \to \infty$,

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

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

But X and λ are equal. Therefore

$$N = X \cdot \overline{T}^{\text{Time Avg}}. \qquad \blacksquare$$

Properties of Little’s Law

- Little’s Law is distribution independent. This means that it depends only on mean quantities.

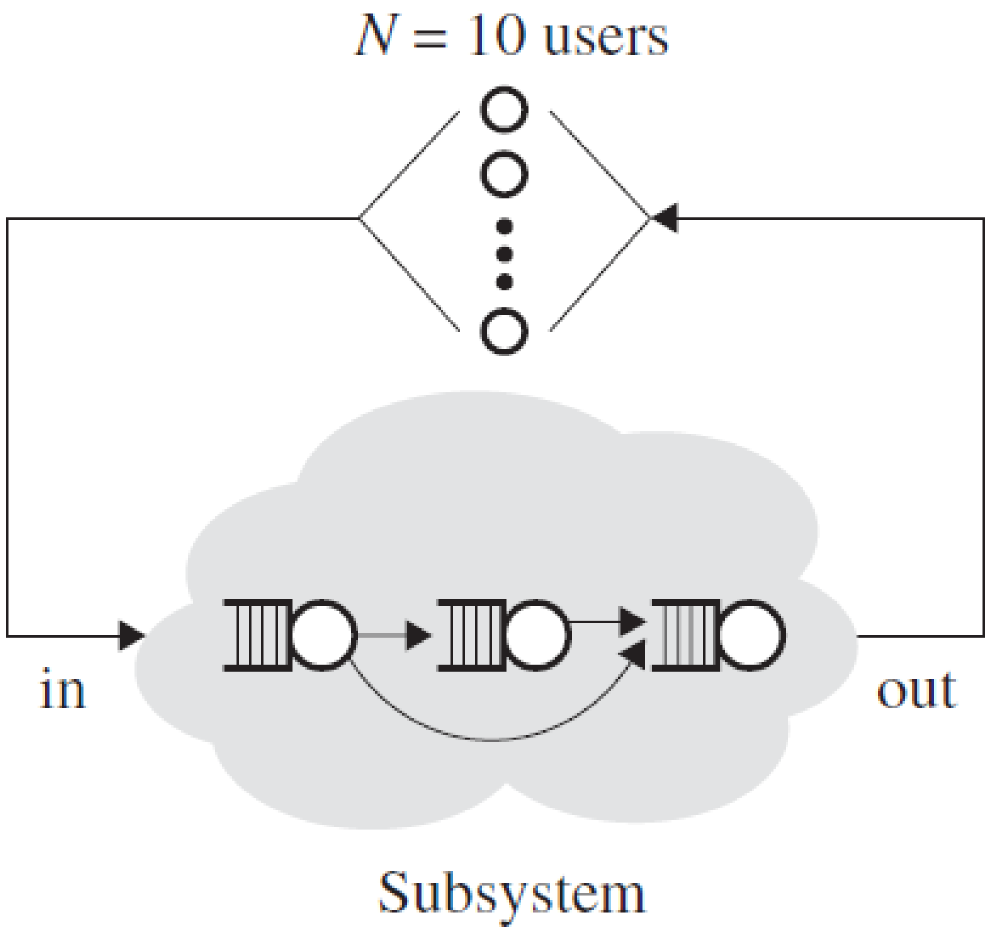
- Little’s Law applies to any system or piece of a system.

Example

- We have an interactive system with $N = 10$ users, the expected think time is $E[Z] = 5$ seconds and the expected response time is $E[R] = 15$ seconds.

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

Example



Answer

- Using Little's Law for closed systems, we have

$$N = X \cdot \mathbf{E}[T] = X(\mathbf{E}[Z] + \mathbf{E}[R])$$
$$\Rightarrow X = \frac{N}{\mathbf{E}[R] + \mathbf{E}[Z]} = \frac{10}{5 + 15} = 0.5 \text{ jobs/sec.}$$

- The application of Little's Law to closed systems is often referred to as the **Response Time Law** for Closed Systems**:**

$$\boxed{\mathbf{E}[R] = \frac{N}{X} - \mathbf{E}[Z]}$$

More Operational Laws: The Forced Flow Law

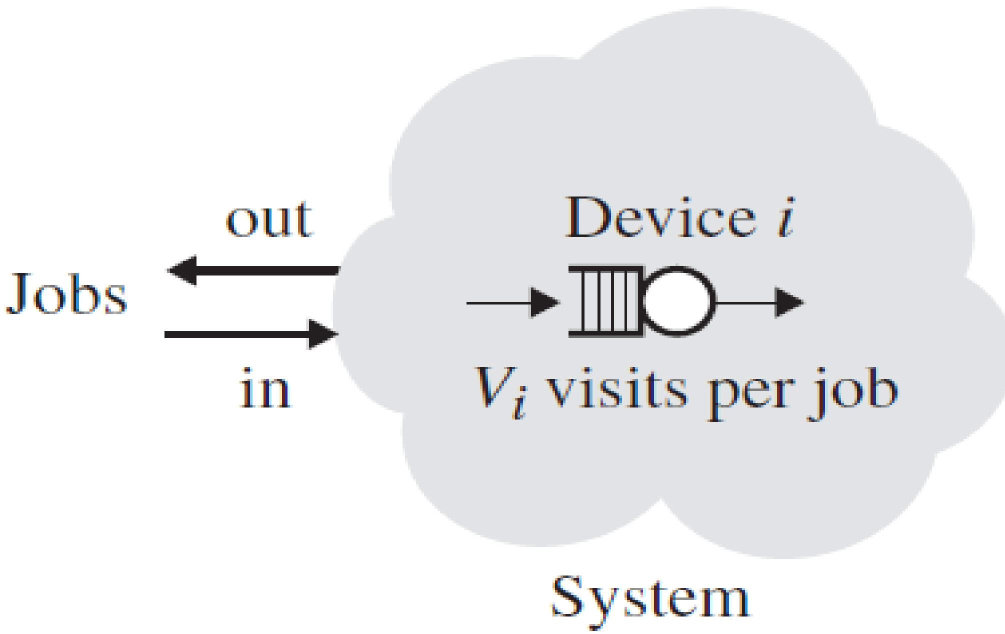
- The **Forced Flow Law** relates system throughput to the throughput of an individual device as follows:

$$\boxed{X_i = \mathbf{E}[V_i] \cdot X}$$

- X denotes the system throughput
- X_i denotes the throughput at device i
- V_i denotes the number of visits to device i per job

The Forced Flow Law

- V_i is often referred to as the **visit ratio** for device i



Proof

[2] A more formal argument would go like this: Consider Figure 6.11. Suppose we observe the system for some large observation period t. Let $C(t)$ denote the number of system completions during time t and let $C_i(t)$ denote the number of completions at device i during time t. Let $V_i^{(j)}$ be the number of visits that the jth job entering the system makes to device i. Then,

$$C_i(t) \approx \sum_{j \in C(t)} V_i^{(j)}$$

$$\frac{C_i(t)}{t} \approx \frac{\sum_{j \in C(t)} V_i^{(j)}}{t}$$

$$\frac{C_i(t)}{t} \approx \frac{\sum_{j \in C(t)} V_i^{(j)}}{C(t)} \cdot \frac{C(t)}{t}$$

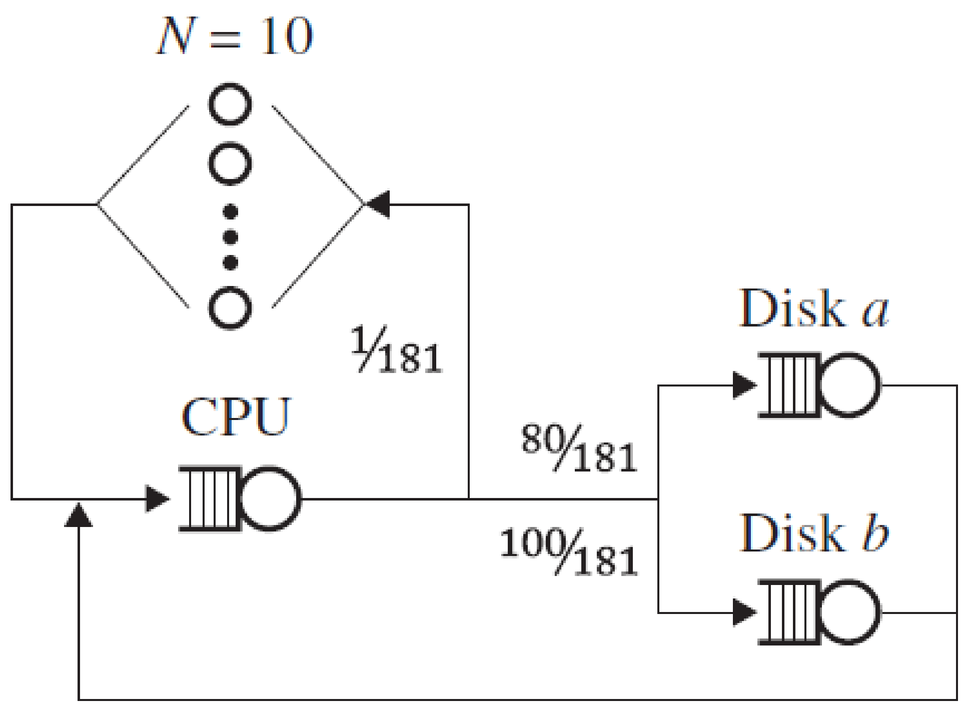
$$\lim_{t \to \infty} \frac{C_i(t)}{t} \approx \lim_{t \to \infty} \frac{\sum_{j \in C(t)} V_i^{(j)}}{C(t)} \cdot \lim_{t \to \infty} \frac{C(t)}{t}$$

$$X_i = \mathbf{E}[V_i] \cdot X.$$

Note that approximation signs are used here because $C(t)$ actually provides a lower bound on the sum, whereas $A(t)$ would provide an upper bound. To be precise, we should use both the upper and lower bounds, but this becomes irrelevant once we take the limit as $t \to \infty$.

Example of Forced Flow Law

- What are the visit ratios? That is, what are E[V_a], E[V_b], and E[V_{cpu}]?



Example of Forced Flow Law

- Looking at the figure, we see

  $C_a = C_{cpu} \cdot 80/181$

  $C_b = C_{cpu} \cdot 100/181$

  $C = C_{cpu} \cdot 1/181$

  $C_{cpu} = C_a + C_b + C.$

- Dividing through by C (the number of system completions) yields the visit ratios. So we get

  $E[V_a] = E[V_{cpu}] \cdot 80/181$

  $E[V_b] = E[V_{cpu}] \cdot 100/181$

  $1 = E[V_{cpu}] \cdot 1/181$

  $E[V_{cpu}] = E[V_a] + E[V_b] + 1.$

Example of Forced Flow Law

- Solving this system of simultaneous equations yields

$$\mathbf{E}\left[V_{\text{cpu}}\right] = 181$$
$$\mathbf{E}\left[V_a\right] = 80$$
$$\mathbf{E}\left[V_b\right] = 100.$$

Example of Forced Flow Law

- See more example in the book.

Bottleneck Law

- Define D_i to be the total service demand on device i for all visits of a single job (i.e., a single interaction). That is,

$$D_i = \sum_{j=1}^{V_i} S_i^{(j)},$$

- where $S_i^{(j)}$ is the service time required by the j th visit of the job to server i

Bottleneck Law

- We immediately see that

$$\mathbf{E}[D_i] = \mathbf{E}[V_i] \cdot \mathbf{E}[S_i]$$

- **Question:** How would you determine $E[D_i]$ in practice?

Bottleneck Law

- **Answer:** Consider a long observation period. Observe that

$$\mathbf{E}[D_i] = \frac{B_i}{C},$$

- B_i is the busy time at device i for the duration of our observation period and C is the number of system completions during this observation period

Bottleneck Law

- The importance of E[D_i] lies in the following law, which we call the **Bottleneck Law**

$$\boxed{\rho_i = X \cdot \mathbf{E}\left[D_i\right]}$$

Proof of the Bottleneck Law

- X is the jobs/sec arriving into the whole system
- Each of these outside arrivals contributes $\mathbf{E}[D_i]$ seconds of work for device i
- So device i is busy for $X \cdot \mathbf{E}[D_i]$ seconds out of every second
- Thus $X \cdot \mathbf{E}[D_i]$ is the utilization of device i