Processor Scheduling in Parallel and Distributed Systems

Part II – Basic Modeling Techniques

Larry Dowdy
Department of Computer Science
Vanderbilt University
Nashville, TN 37235
dowdy@vuse.vanderbilt.edu

June 16, 1997

Outline

Basic example and techniques (25 minutes)

Fishing application (15 minutes)

Intermediate example and techniques (5 minutes)

Summary of concepts (1 minute)

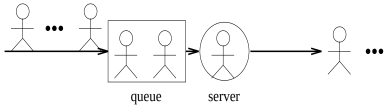
Goal: To "solve" the system

$\Longrightarrow$

utilization,
throughput,
mean queue length,
response time,
•••

Basic Example

"the system"



Workload Characterization

- mean inter-arrival times ($\frac{1}{\lambda}$)
- mean service time ($\frac{1}{\mu}$)

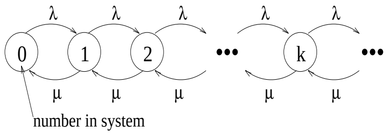
Assumptions:

- exponential distributions
- single class
- work conserving

... and a few more ...

Basic Solution Technique

Step 1: Build state space diagram



Step 2: Construct (global) flow balance equations

Flow in = Flow out

$$\mu P_1 = \lambda P_0$$

$$\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1$$

$$\vdots$$

$$\lambda P_{k-1} + \mu P_{k+1} = \lambda P_k + \mu P_k$$

Step 3: Solve 'em

$$P_k = \frac{\lambda}{\mu} P_{k-1} = \frac{\lambda}{\mu}\frac{\lambda}{\mu} P_{k-2} = \cdots = \left(\frac{\lambda}{\mu}\right)^k P_0$$

$$P_0 + P_1 + P_2 + \cdots = 1$$

$$\rightarrow \; P_0 + \left(\frac{\lambda}{\mu}\right) P_0 + \left(\frac{\lambda}{\mu}\right)^2 P_0 + \cdots = 1$$

$$\rightarrow \; P_0 = \frac{1}{1 + \left(\frac{\lambda}{\mu}\right) + \left(\frac{\lambda}{\mu}\right)^2 + \cdots} = 1 - \frac{\lambda}{\mu}$$

$$\rightarrow \; P_k = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right)$$

Step 4: Get the performance measures

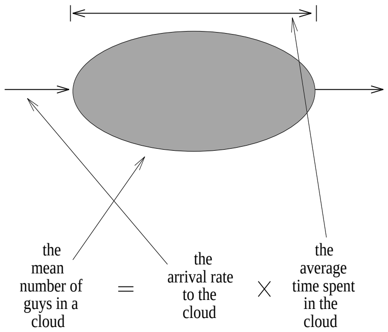
Utilization $= P_1 + P_2 + \cdots = 1 - P_0 = \frac{\lambda}{\mu}$

Throughput $= 0P_0 + \mu P_1 + \mu P_2 + \cdots$
$= \mu(P_1 + P_2 + \cdots)$
$= \mu *$ utilization
$= \lambda$

Queue Length $= 0P_0 + 1P_1 + 2P_2 + \cdots$
$= \sum_{k=1}^{\infty} k \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right)$
$\vdots$
$= \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}}$

Response Time $= \overbrace{\frac{1}{\mu} * \underbrace{\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}}}_{\text{\# I see at arrival time}}}^{\text{time to finish those ahead of me}} + \underbrace{\frac{1}{\mu}}_{\text{my service time}} = \cdots = \frac{1}{\mu - \lambda}$

Alternately, Little's result says

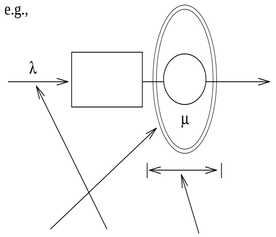


Therefore, QLEN = TPUT * RT

$$\Longrightarrow \; \text{RT} = \frac{\text{QLEN}}{\text{TPUT}} = \frac{\frac{\lambda/\mu}{1 - \lambda/\mu}}{\lambda} = \frac{1}{\mu - \lambda}$$

Little's result is very general

e.g.,



UTIL = TPUT * mean service time

$$= \lambda * \frac{1}{\mu}$$

Play "what-if" prediction games:

Homework assignment #1:

What if the system is upgraded by a factor of 2?

UTIL?

TPUT?

QLEN?

RT?