Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.826 – Principles of Computer Systems

Handout 17 September 29, 1993

Basics of Performance Analysis

This handout describes the basic concepts and equations used in operational analysis of queueing networks models.

1 Basic Parameters

The following basic parameters characterize the observable behavior seen by clients of a server (or subserver):

T	length of the observation interval
A	number of arrivals during the observation interval
B	total time the server is busy during the observation interval
C	number of completions during the observation interval

2 Derived Parameters

The following parameters can be derived from the basic parameters:

λ	arrival rate ($= A/T$)
X	completion rate ($= C/T$)
U	utilization ($= B/T$)
S	service time per completed job ($= B/C$)

3 Additional Parameters

The following additional parameters are also useful:

R	response time
n_i	queue length at subserver i (a function of time)
N	average number of jobs in system
V_i	visit ratios (average number of visits to subserver i by each job)

For closed systems, we also need:

M	total number of clients
Z	think time

4 Laws and Formulas

Here is a summary of the basic laws and formulas:

$U_i = X_i S_i$	Utilization Law
$n_i = X_i R_i$	Little's Law
$X_i = V_i X_0$	Forced Flow Law
$R = \sum V_i R_i$	General Response Time Law
$R = M/X_0 - Z$	Interactive Response Time Formula (assumes flow balance)

5 Bottleneck Analysis

The best possible response time $R_{min} = \sum V_i S_i$.

The throughput X_0 is bounded above by two curves: N/R_{min} and $1/(V_b S_b)$ (where subserver b is the bottleneck). The saturation point N^* occurs where the curves intersect; i.e., $N^* = R_{min}/(V_b S_b)$.

By Little's Law, $R_0 = N/X_0$; since $X_0 \leq 1/(V_b S_b)$, it follows that $R_0 \geq N V_b S_b$. Therefore, R_0 is bounded below by the curves R_{min} and $N V_b S_b$.

In a closed system, $R_0 = M/X_0 - Z$, so $R_0 \geq M V_b S_b - Z$.