# Lecture 6
# Operational Analysis

Network of Queues

Observations

Operational Laws

Bottleneck Analysis

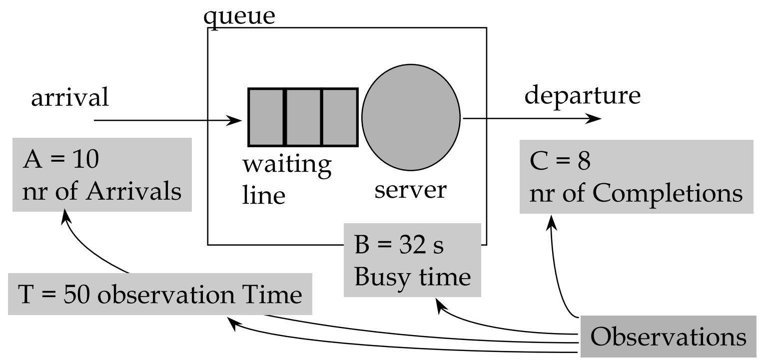


## Operational Analysis

(operaatio-analyysi)

- Simple to use
- Limited gain
- "Back in the envelope" calculations
- Based on (simple) measurements or observations
- What happens (happened) in the system?
- Can be used to give bounds for predicted performance

# Single Server Queue

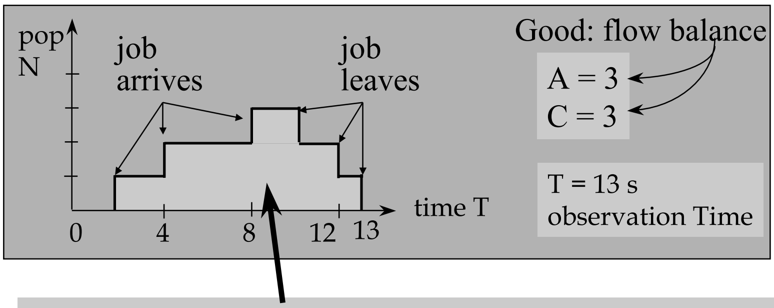



| | |
|---|---|
| mean service time | S = B/C = 32/8 = 4 s |
| server utilization | U = B/T = 32/50 = 0.64 |
| throughput | X = C/T = 8/50 = 0.16 |



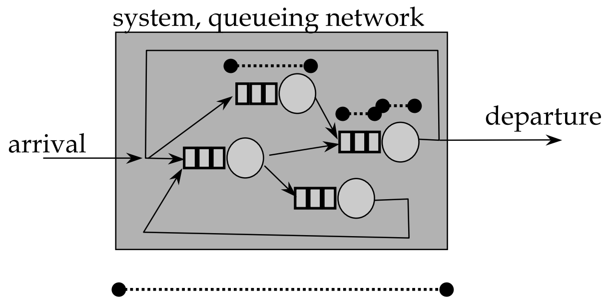
# Waiting Time Integral

(odotusaikaintegraali)




total waiting time (queue + service) = W = 21
= area under population curve
average response time = R = W/C = 21/3 = 7 s
average population level N = W/T = 21/13 = 1.61
average throughput = C/T = N/R = 3/13 = 0.23 transactions/s

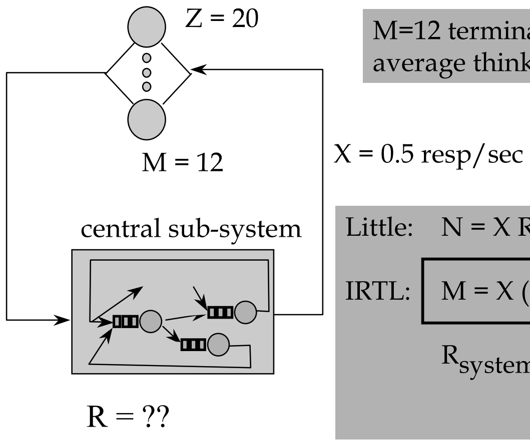
# Little's Law: N = XR Apply to Utilization, Waiting line, Node, Whole network






# Interactive Response Time Law (IRTL) (1)

(interaktiivinen vastausaikalaki)




M=12 terminals,
average thinking time Z=20 s

Little: $N = X R_{total}$

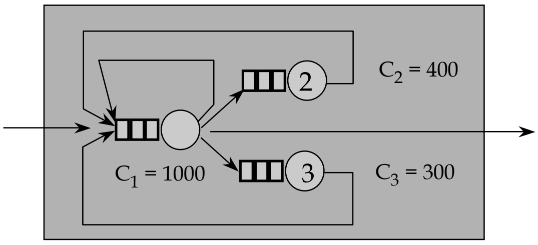
IRTL: $M = X (R_{system} + Z)$

$R_{system} = M/X - Z$

$= 12/0.5 - 20 = 4$ s

# Forced Flow Law (FFL) (2)

(pakkovuolaki)




T = 10 s

$X_0 = 20$ (trans/s)

$C_0 = 200$

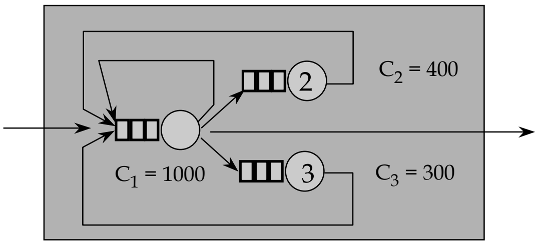
Visit ratio (FFL) $V_k = \frac{C_k}{C_0} = \frac{C_k/T}{C_0/T} = \frac{X_k}{X_0}$ i.e., $X_k = V_k X_0$

$V_1 = 1000/200 = 5$
$V_2 = 400/200 = 2$
$V_3 = 300/200 = 1.5$

$X_1 = 5 \bullet 20 = 100$
$X_2 = 2 \bullet 20 = 40$
$X_3 = 1.5 \bullet 20 = 30$

18.3.2002  7

# Branching Probabilities




T = 10 s

$X_0 = 20$ (trans/s)

$C_0 = 200$

Branching Probability $q_{ij} = \frac{C_{ij}}{C_i}$

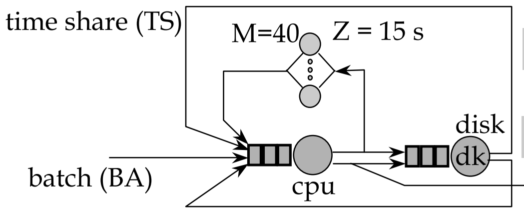
$q_{10} = 200/1000 = 0.2 = 20\%$
$q_{12} = 400/1000 = 0.4 = 40\%$
$q_{13} = 300/1000 = 0.3 = 30\%$
$q_{11} = 1.0 - q_{10} - q_{12} - q_{13} = 0.1 = 10\%$
$q_{21} = 1.0 = q_{31}$

easy when arrivals only from one other server, or only to one other server

difficult when arrivals from many other servers (solve system of linear equations)

18.3.2002  8

# Example (Denning & Buzen, 1978) (5)




$R_{0,TS}$ = 5 sec

$V_{DK,TS}$ = 10

$X_{0,BA}$ = ?

$V_{DK,BA}$ = 5

$S_{DK}$ = 40 ms
$U_{DK}$ = 0.9

$$X_{0,TS} \overset{RTL}{=} \frac{M}{Z+R_{0,TS}} = 40/(15+5) = 2 \text{ jobs/s}$$

$$X_{DK,TS} \overset{FFL}{=} V_{DK,TS} X_{0,TS} = 10*2 = 20 \text{ jobs/s}$$

$$X_{DK,ALL} \overset{Little}{=} \frac{U_{DK}}{S_{DK}} = \frac{0.9}{0.040} = 22.5 \text{ jobs/s}$$

$$X_{DK,BA} = X_{DK,ALL} - X_{DK,TS} = 22.5 - 20 = 2.5 \text{ jobs/s}$$

$$X_{0,BA} \overset{FFL}{=} \frac{X_{DK,BA}}{V_{DK,BA}} = \frac{2.5}{5} = 0.5 \text{ jobs/s}$$



# Example Contd (Denning & Buzen, 1978)

What happens if batch throughput triples, i.e., $X_{0,BA}$ = 1.5?

$$X_{DK,BA} \overset{FFL}{=} V_{DK,BA} * X_{0,BA} = 5*1.5 = 7.5 \text{ jobs/s}$$

$$X_{DK}^{max} \overset{Little}{=} \frac{U^{max}}{S_{DK}} = \frac{1}{0.04} = 25 \text{ jobs/s}$$

$$X_{DK,TS}^{max} = 25 - 7.5 = 17.5 \text{ jobs/s}$$

$$x_{0,TS}^{max} \overset{FFL}{=} \frac{X_{DK,TS}^{max}}{V_{DK,TS}} = \frac{17.5}{10} = 1.75 \text{ jobs/s}$$

$$R_{0,TS}^{min} \overset{RTL}{=} \frac{M}{x_{0,TS}^{max}} - Z = \frac{20}{1.75} - 15 = 7.9s$$

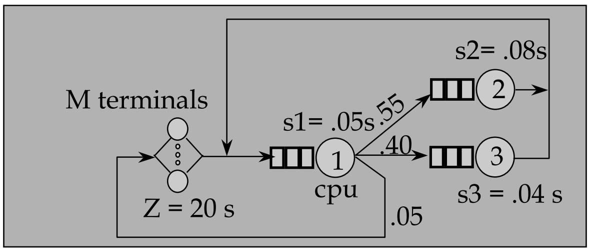
vs. 5 s before

Assuming that M, Z, $V_i$'s and $S_i$'s did not change!

# Bottleneck Analysis

(pullonkaulanalyysi)

- ABA, Asymptotic Bound Analysis

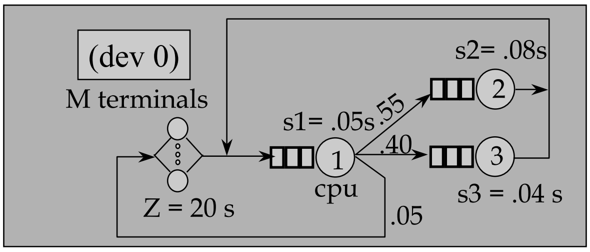



- $q_{ij}$'s know, can compute $V_i$'s
  - $V_1 = 20$, $V_2 = 11$, $V_2 = 8$
- Resource Demands $D_i = V_i S_i$
  - $D_1 = 1.0$, $D_2 = 0.88$, $D_2 = 0.32$

(deductions, or lin. equ's)



# Compute $V_i$'s, $D_i$'s from $q_{ij}$'s




↓

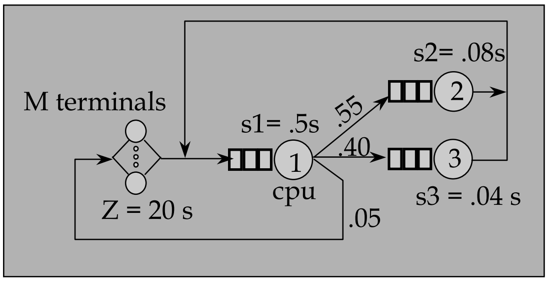
$V_0 = 0.05\, V_1$
$V_1 = V_0 + V_2 + V_3$
$V_2 = 0.55 V_1$
$V_3 = 0.4 V_1$

linear equations

→

$V_1 = 20 \quad V_2 = 11 \quad V_3 = 8$
$D_1 = 1.0 \quad D_2 = 0.88 \quad D_3 = 0.32$
$R_0 = D_1 + D_2 + D_3 = 2.2$ sec

# Bottleneck Analysis (5)




- Resource Demands: $D_1$= 1.0, $D_2$= 0.88, $D_2$= 0.32
- Minimum response time $R_0 = \Sigma D_i = 2.2$ s
- For each device, $U_i \overset{\text{Little}}{=} X_i S_i \overset{\text{FFL}}{=} V_i S_i X_0 = D_i X_0$
- Q: If system load increases, which device will saturate first? I.e., which device will first reach $U_i = 1.0$?
- A: The one with the highest $D_i$: $D_1$ = Bottleneck device



# Bottleneck Analysis

$$X_0 = \frac{U_i}{D_i} < \frac{1.0}{D_i} \quad \text{for all devices i}$$

upper limits for system throughput

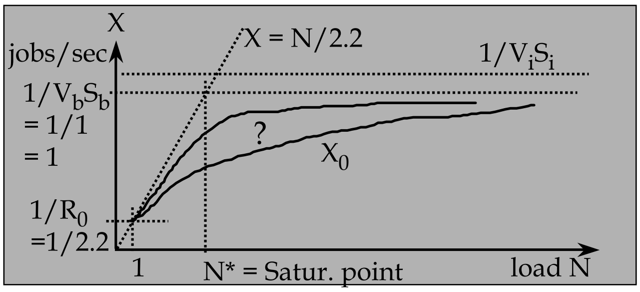
Device b is the bottleneck device if $D_b = \max D_i$

$$X_0 < \frac{1.0}{D_b} = \frac{1.0}{V_b S_b}$$

Also,

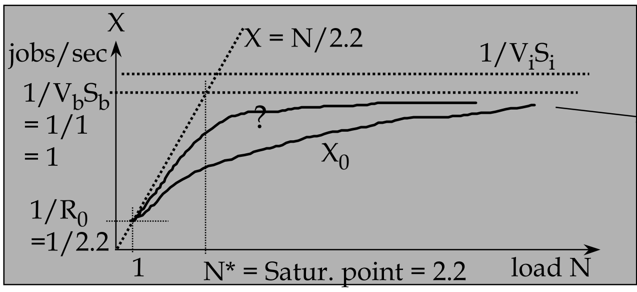
$X_0(N=1) = 1/R_0 = 1/\Sigma D_i$ and $X_0(N=k) \le k/R_0$

# Bottleneck Analysis



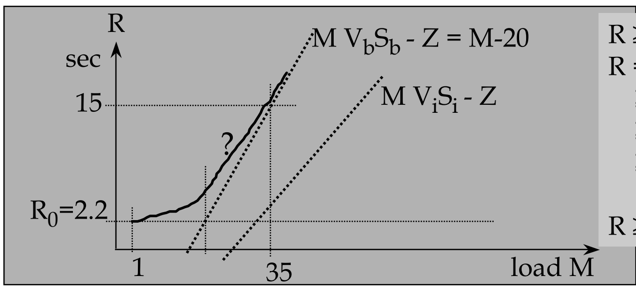



# Bottleneck Analysis




$X_0 \leq 1/D_b$

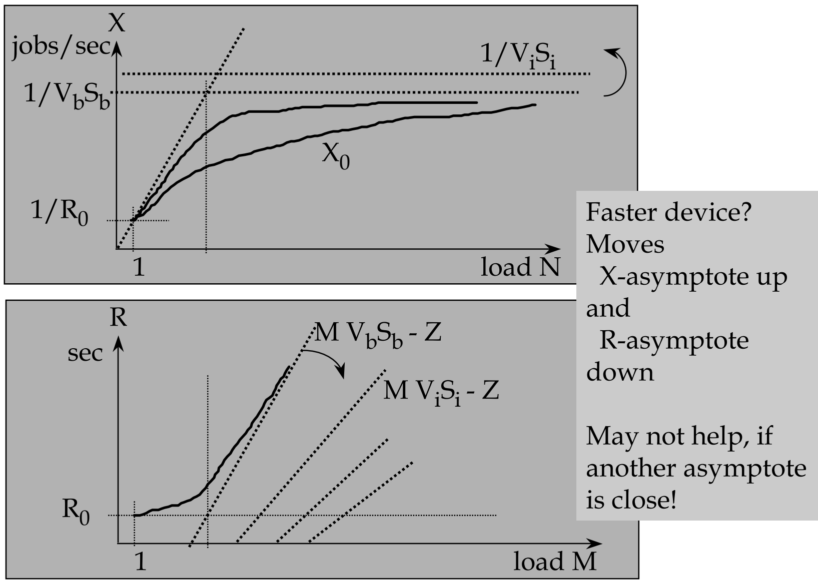
lower limits for R?




$R \geq R_0$
$R = M/X_0 - Z$
$\geq M\ D_b - Z$
$\geq M\ D_i - Z$
$\geq M\ V_i\ S_i - Z$

$R \geq M\ V_b\ S_b - Z$

# Bottleneck Analysis




18.3.2002 Copyright Teemu Kerola 2002 17

# Balanced Job Bounds

- BJB, Balanced Job Bounds
  - make 2 models: one faster and one slower
  - both new models very easy to solve
  - get fast, rough bounds

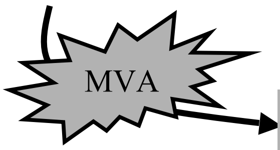
18.3.2002 Copyright Teemu Kerola 1998 18

# BJB - Balanced Job Bounds

- Original model — $D_c$=10, $D_f$=10, $D_s$=15 (secs)
- Slower bound model — $D_c$=15, $D_f$=15, $D_s$=15 (secs)
- Faster bound model — $D_c$=11.7, $D_f$=11.7, $D_s$=11.7
- Solve with constant demand $D_{const}$

ave max

$V_i S_i = D_i = D_{const}$




$$X_0(N) = \frac{N}{D_{const}(K+N-1)}$$

- Figs 5.10, 5.11 [Men 94]



# Better BJB's

- Original model: $D_c$=10, $D_f$=10, $D_s$=15 (secs)
- With max demand $D_{max}$ consider also total demand D: $D = \sum_{i=0}^{K} D_i$ — D = 35, K=3
- New slower bound model:
  - $D/D_{max}$ servers with max load ($D_{max}$)
  - other servers with zero load (0)
  - intuitively OK when $D/D_{max}$ integer
  - math works even if not

# Better BJB's

- New lower limit:

$$X_0(N) = \frac{N}{D_{\max}\left(\frac{D}{D_{\max}} + N - 1\right)} = \frac{N}{D + D_{\max}(N-1)}$$

Figs 5.7a and 5.7b [LZGS84]