

Lecture 5

Intuitive Solutions for Simple Models

M/M/1 Queue
Markov Chains
Little's Law

18.3.2002

Copyright Teemu Kerola 2002

1

Solution Methods in Overall Picture

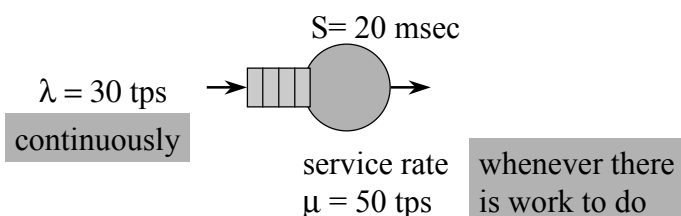
- Baseline model
- Prediction model
- Fig. 4.1

18.3.2002

Copyright Teemu Kerola 2002

2

Single Server System



- Server utilization U ?
- Server queue length Q ?
- Server response time R ?
- Server & system throughput X ?

18.3.2002

Copyright Teemu Kerola 2002

3

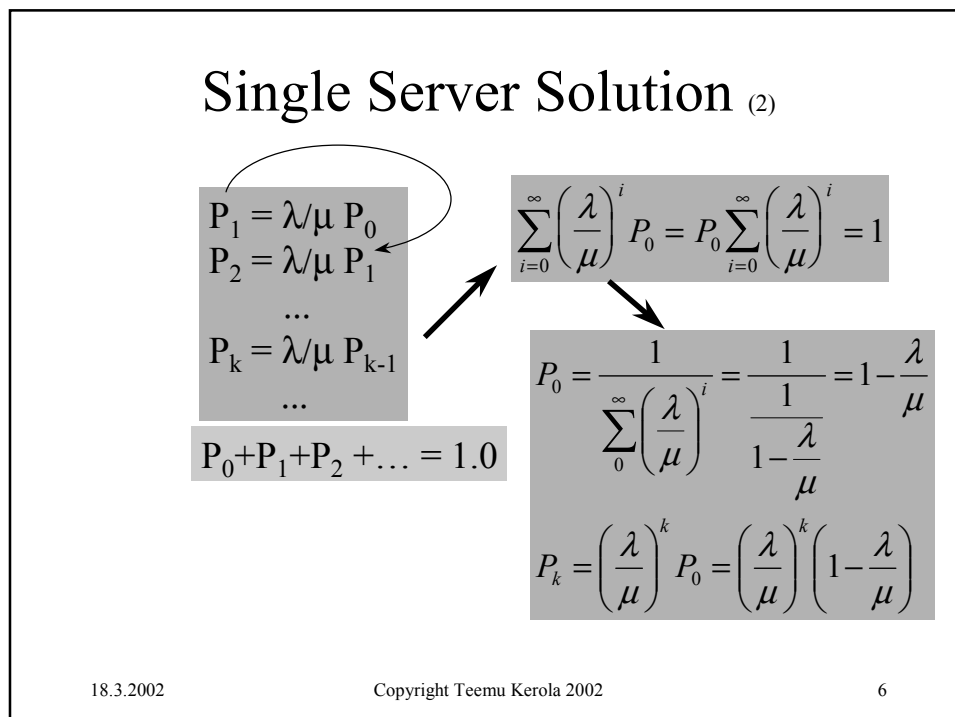
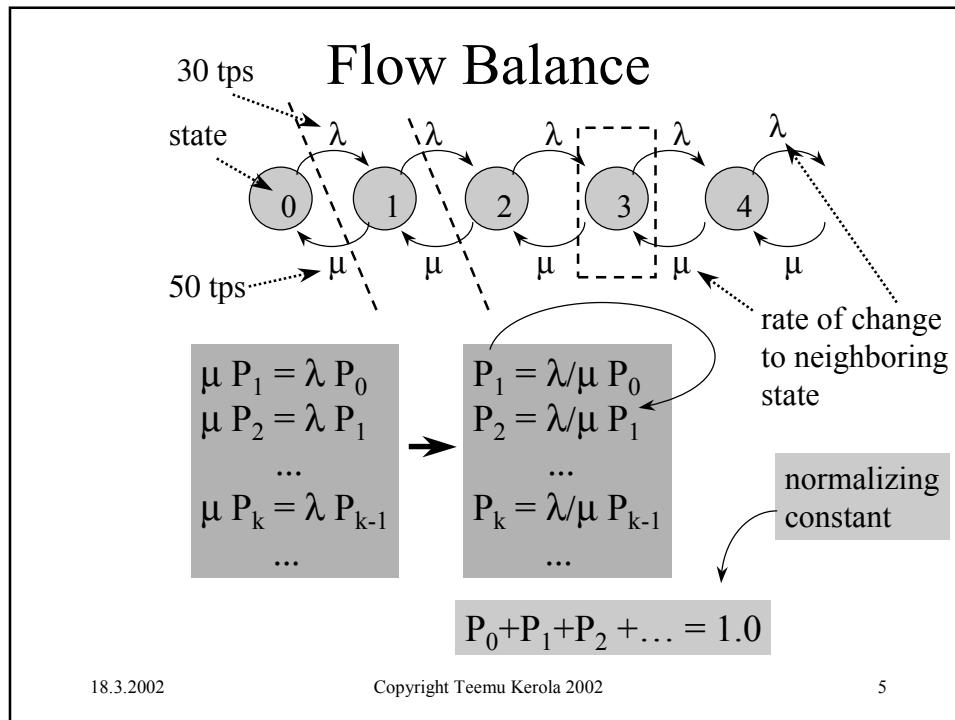
Birth-Death System

- Markov chain (history is irrelevant)
- Birth-death (change to adjacent states only)
- Stochastic process (randomness)
- Fig. 4.3
- System states, nr of jobs (k) in system
 $k = 0, 1, 2, \dots$
- Need: System statistics
 - average U, R, X, Q
- Can compute them from state probabilities $P_k \forall k=0, 1, 2, \dots$

18.3.2002

Copyright Teemu Kerola 2002

4



Single Server Solution (contd) ⁽¹⁾

$$P_0 = \frac{1}{\sum_0^{\infty} \left(\frac{\lambda}{\mu}\right)^i} = \frac{1}{\frac{1}{1 - \frac{\lambda}{\mu}}} = 1 - \frac{\lambda}{\mu}$$

$$\lambda = 30, \mu = 50$$

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

aver. popul?
R?

$$P_0 = 1 - \frac{30}{50} = 0.4 = \boxed{40\%}$$

$$U = 1 - P_0 = \boxed{60\%} \quad (= \lambda/\mu)$$

$$X = \sum_{i=0}^{\infty} \mu_i P_i = \sum_{i=1}^{\infty} \mu_i P_i = \mu \sum_{i=1}^{\infty} P_i$$

$$= \mu U = \mu \frac{\lambda}{\mu} = \lambda = \boxed{30tps}$$

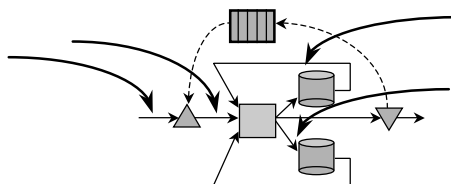
18.3.2002

Copyright Teemu Kerola 2002

7

Arrival Theorem

- Arriving customer to some node (I.e., a customer in transition) sees the system as steady state system with itself removed
 - infinite population: overall steady state
 - finite population: steady state for system with one job less



18.3.2002

Copyright Teemu Kerola 2002

8

Single Server Solution (contd)

number of
customers
in system
(aver.
population)

$$\begin{aligned}\bar{N} &= \sum_0^{\infty} iP_i = \sum_1^{\infty} iP_i = \sum_1^{\infty} i \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right) \\ &= \left(1 - \frac{\lambda}{\mu}\right) \frac{\frac{\lambda}{\mu}}{\left(1 - \frac{\lambda}{\mu}\right)^2} = \frac{\lambda}{\mu - \lambda} = \frac{30}{50 - 30} = 1.5\end{aligned}$$

response time = time in queue + time in service

$$R = \bar{N} \frac{1}{\mu} + \frac{1}{\mu} = \frac{\lambda}{\mu - \lambda} \frac{1}{\mu} + \frac{1}{\mu} = \frac{1}{\mu - \lambda} = \frac{1}{20} = 0.05 \text{ sec}$$

arrival theorem → nr of jobs in front of "me"

18.3.2002

Copyright Teemu Kerola 2002

9

Single Server Solution (contd)

aver nr of jobs in queue (not in service)
= nr of jobs in system - utilization

$$\bar{N} - U = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = 1.5 - 0.6 = 0.9$$

18.3.2002

Copyright Teemu Kerola 2002

10

Use Solution for Predictions

(IAT=22 ms, S=20 ms)
 $\lambda = 45, \mu = 50$

$U = \frac{\lambda}{\mu} = 90\%$
 $X = \lambda = 45\text{tps}$
 $\bar{N} = \frac{\lambda}{\mu - \lambda} = 9$
 $R = \frac{1}{\mu - \lambda} = \frac{1}{5} = 0.2\text{sec}$

(IAT=22 ms, S=13 ms)
 $\lambda = 45, \mu = 75$

(λ, μ)	(30,50)	(45,50)	(45,75)
U	60%	90%	60%
X	30	45	45 tps
$\bar{N}$	1.5	9	1.5
R	0.05	0.2	0.033 sec

Limited buffer example ⁽¹⁾

- Limited buffer size: 3
 - if packet (job,transaction) arrives and buffer full, then that packet is lost
 - flow balance:

$\mu P_1 = \lambda P_0$
 $\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1$
 $\lambda P_1 + \mu P_3 = \lambda P_2 + \mu P_2$
 $\mu P_2 = \lambda P_3$

The diagram shows a buffer with four states, 0, 1, 2, and 3, each in a circle. Transitions are labeled with arrival rate λ (from 0 to 1, 1 to 2, 2 to 3) and service rate μ (from 1 to 0, 2 to 1, 3 to 2). Dashed boxes group states 0-1, 1-2, and 2-3. Dotted arrows connect the flow balance equations to the corresponding transitions in the diagram.

Limited buffer (contd) ⁽¹⁰⁾

$$\begin{aligned}\mu P_1 &= \lambda P_0 \\ \lambda P_0 + \mu P_2 &= \lambda P_1 + \mu P_1 \\ \lambda P_1 + \mu P_3 &= \lambda P_2 + \mu P_2 \\ \mu P_2 &= \lambda P_3\end{aligned}$$

$$\begin{aligned}\lambda &= 30 \text{ tps}, \mu = 50 \text{ tps} \\ \lambda/\mu &= 0.6\end{aligned}$$

$$P_0 + P_1 + P_2 + P_3 = 1.0$$

$$\lambda P_0 + \mu P_2 = \lambda (\lambda/\mu P_0) + \mu (\lambda/\mu)^2 P_0$$

$$P_0 + 0.6 P_0 + 0.6^2 P_0 + 0.6^3 P_0 = 1.0$$

$$(1 + 0.6 + 0.36 + 0.216) P_0 = 1.0$$

$$2.176 P_0 = 1.0$$

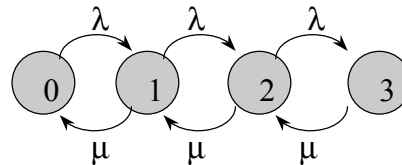
$$P_0 = 0.46, P_1 = 0.275, P_2 = 0.165, P_3 = 0.10$$

18.3.2002

Copyright Teemu Kerola 2002

13

Limited buffer (contd) ⁽⁶⁾



$$P_0 = 0.46, P_1 = 0.275, P_2 = 0.165, P_3 = 0.10$$

$$\text{utilization} = 1 - P_0 = 0.54 = U$$

$$\begin{aligned}\text{throughput} &= 0 * P_0 + \mu P_1 + \mu P_2 + \mu P_3 \\ &= \mu (P_1 + P_2 + P_3) = 50 * 0.54 = 27 \text{ tps} = X\end{aligned}$$

$$\begin{aligned}\text{average population} &= 0 * P_0 + 1 P_1 + 2 P_2 + 3 P_3 \\ &= 0.275 + 2 * 0.165 + 3 * 0.10 = 0.91 = \bar{N}\end{aligned}$$

$$\text{average queue length} = \text{aver pop} - \text{util} = 0.91 - 0.54 = 0.35$$

$$\text{response time} = (1/\mu)\bar{N} + (1/\mu) = 0.02 * 0.91 + 0.02 = 0.4 = R$$

$$\text{loss rate} = \lambda P_3 = 30 * 0.10 = 3 \text{ tps} \quad \text{Prob(loss)} = P_3 = 0.10$$

18.3.2002

Copyright Teemu Kerola 2002

14

Queue length vs. population

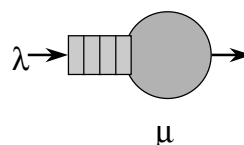
- Two terminologies
- "Queue length" = population Menasce
 - queue includes those in service
- Population N = queue + those in service
 - queue includes only those not in service
 - N = queue length + U LZGS
- No big problem – watch out.

18.3.2002

Copyright Teemu Kerola 2002

15

M/M/1 Queue



one server

Exponential death process: $\mu_i = \mu$ Exponential birth process: $\lambda_i = \lambda$

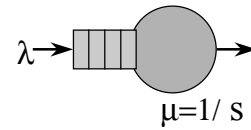
- Single server example = M/M/1
- M/M/2 M/M/m M/M/ ∞
- M/M/1/B M/M/m/B (finite buffer B)
- M/M/1/B/K (finite customer population K)
- M/G/1 G/G/1 M/G/m/B/K

18.3.2002

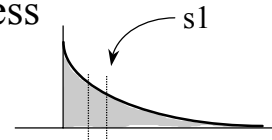
Copyright Teemu Kerola 2002

16

M/M/1 Queue



- Markovian birth-death process
- Exponential distributions
 - birth & death
 - memoryless: $\Pr(S \geq s) = \Pr(S^{\text{rem}} \geq s \mid S^{\text{serv}}=s)$
- Infinite queue (line, buffer) space
- First-in-first-out queueing discipline
- $\text{Prob}(\text{empty}) = P_0 > 0$ (true when $\mu > \lambda$)

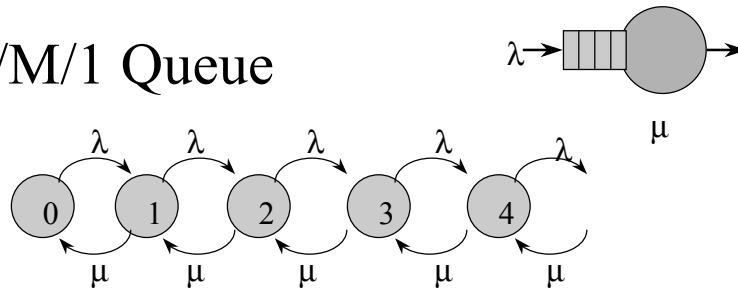


18.3.2002

Copyright Teemu Kerola 2002

17

M/M/1 Queue



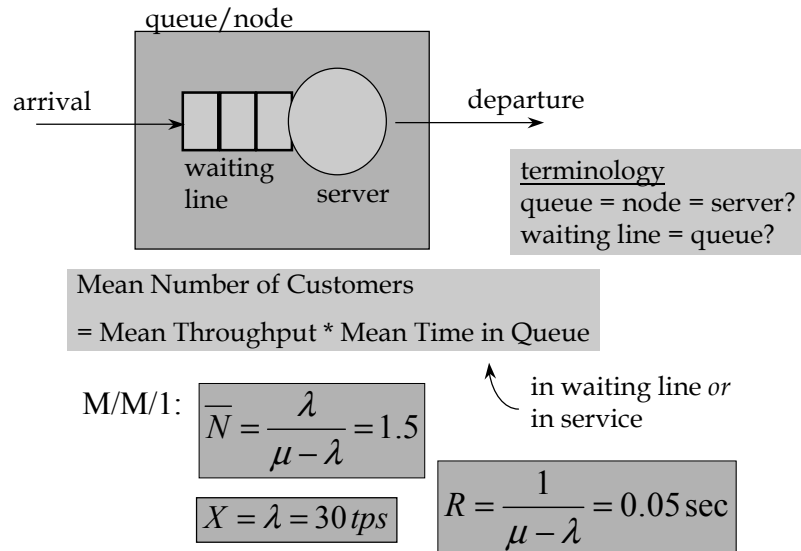
- Transition probability matrix $\mathbf{Q}$ $\mathbf{Q}[1,2] = \lambda / (\lambda + \mu)$
- Ergodicity: recurrent, aperiodic states if $\lambda < \mu$
- Probability vector of being in state at time t : $\mathbf{p}(t)$
- Stable limit probability $\lim_{t \rightarrow \infty} \mathbf{p}(t) = \Pi$ $\Pi_i = P_i$
- Steady state solution $\Pi = \Pi \mathbf{Q}$
- Rephrase question: How to find such Π that $\Pi = \Pi \mathbf{Q}$?
 Box 31.1 [Jain 91]
 Box 31.4 [Jain 91]

18.3.2002

Copyright Teemu Kerola 2002

18

Little's Law: $N = XR$ ₍₁₎

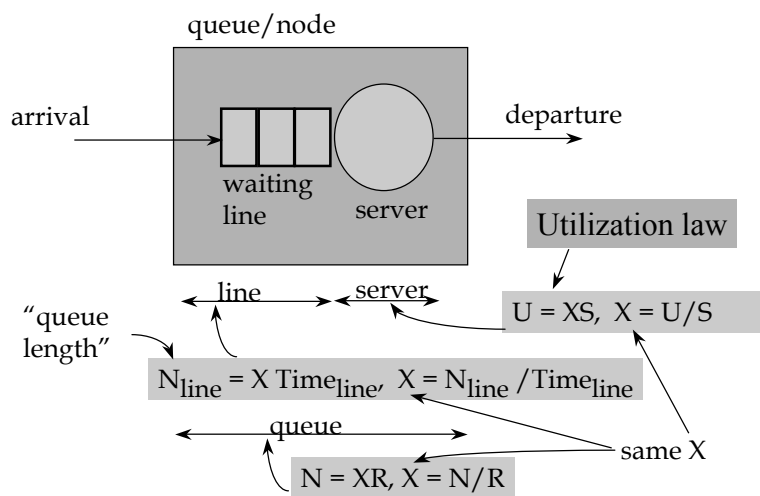


18.3.2002

Copyright Teemu Kerola 2002

19

Little's Law: $N = XR$ Apply to Queue, Server, or Line

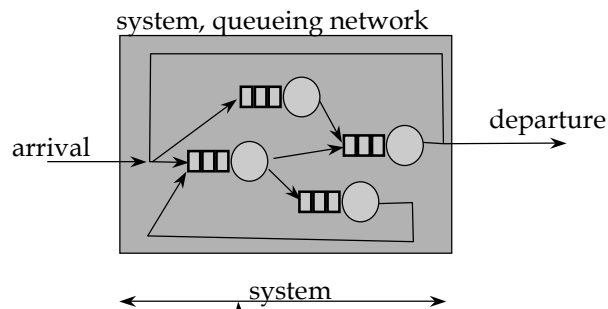


18.3.2002

Copyright Teemu Kerola 2002

20

Little's Law: $N = XR$ Apply to Whole System ⁽¹⁾



$$N = N_{\text{system}} = X_{\text{system}} R_{\text{system}} = X_0 R$$

$$X_0 = N/R$$

Even larger system?
Include human users
in "system".

18.3.2002

Copyright Teemu Kerola 2002

21

18.3.2002

Copyright Teemu Kerola 2002

22