SE3910 – REAL TIME SYSTEMS

Queuing Theory

Long Lines

ROADMAP

- Today
  - Queuing Theory
- Monday
  - Memory Utilization
- Wednesday
  - Toyota systems failure

Friday
Review

Final Exam: Monday 14:00–16:00

OBJECTIVES

- Explain the concept of a Poisson queue
- Explain how to calculate the average servicing time for a system
- Explain the concepts of an M/M/1 queue
- For an M/M/1 queue system, calculate the average response time and the average number of customers in the system
- Calculate the mean response time for an M/M/2 queue.
- Calculate the Average time spent in an M/M/infinite queue system.

Applicable

IN CLASS EXERCISE

- Working with a partner
  - One person will solve these easy math problems
  - One person will time how long it takes to solve all of these math problems
  - We'll record the times at the end for data analysis purposes.

THE GROCERY STORE CHECKOUT

"Good"

"Good"

QUEUING THEORY

- The classical applied statistics problem is the producer – consumer problem
  - Producer
    - Makes something that needs to be processed
  - Consumer
    - Processes something that has been made by the producer

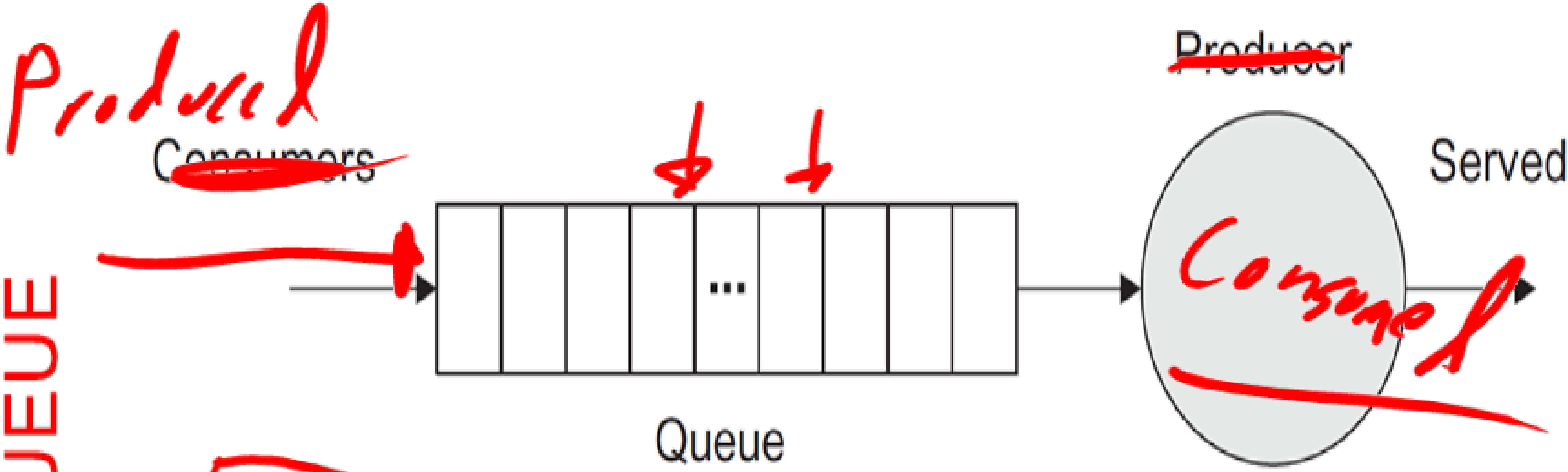
3 things

- Standard description is a tuple
  - Probability distribution for the arrival of "customers" / Probability distribution of the time needed to service each customer / Number of processors consuming the data

How long

MM1 queue

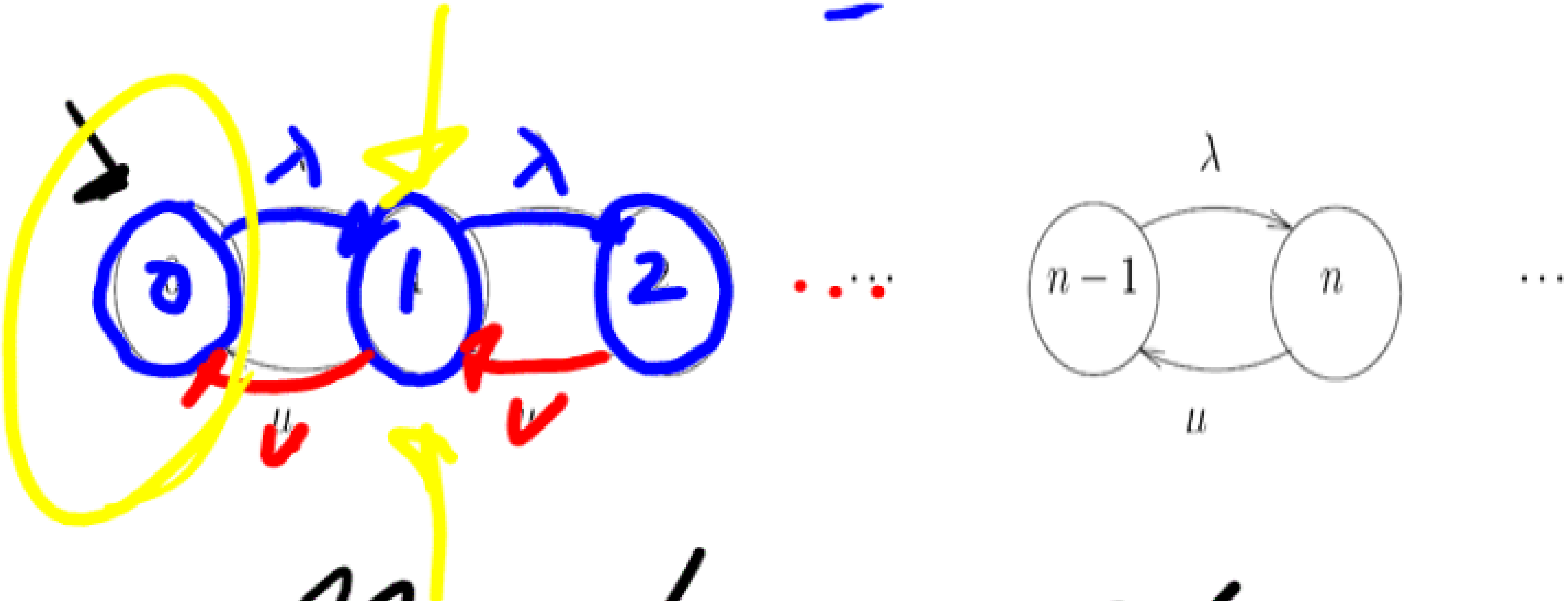
THE M/M/1 QUEUE



- M/M/1 queue is the simplest system — times are exponentially distributed
  - M represents exponentially distributed systems with a Poisson distribution
- $Mean\ arrival\ time = \frac{1}{\lambda}$ ← How often we arrive
- $Mean\ processing\ time = \frac{1}{\mu}$ ← How long to process.
- $\frac{1}{\lambda} > \frac{1}{\mu}$

DERIVATION OF THE FORMULA

- Let N be the average number of customers in the queue
- Let p_n be the probability of there being n customers in the queue



Markov Chain

DERIVATION

- Lets write an equation for n=0

$P_0 \lambda = P_1 \mu$

$\rho = \frac{\lambda}{\mu}$

$P_1 = \rho P_0$

- Lets write a balance equation for n=1

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

$P_2 = \rho P_1$

- Lets write a global balance equation for n=I

- What is the sum of p0...pn?

$P_0 (1 + \rho + \rho^2 + \rho^3 + \dots) = 1$

$\rho < 1$

$\lambda < \mu$

- If λ<μ, then p0/(1-ρ)=1

$\Rightarrow P_0 = (1 - \rho)$

$P_i = (1 - \rho) \rho^i$

DERIVATION

- The average number of customers in the queue is thus:

$$\bar{N} = \frac{\rho}{1-\rho}$$

- Little's Theorem

WHAT IS THE MEAN TIME SOMEONE WOULD SPEND IN THE QUEUE?

- $T = \frac{1/\mu}{1-\rho}$

$$T = \frac{\frac{1}{\mu}}{1-\rho}$$

WHAT IS THE PROBABILITY OF K CUSTOMERS IN THE QUEUE?

- $\Pr[\geq k \text{ in system}] = \rho^k$

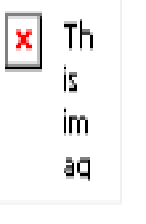
$\rho^k \Leftarrow \left(\frac{\lambda}{\mu}\right)^k$

REAL WORLD EXAMPLE

- We have a web server
  - Single thread for right now
  - 10 MB/s Ethernet connection feeding from the site.
  - What is the mean time it will take to download files?

File Size (KB)	Download Time (10 MB/s etherne (e)	Probability of being accessed	Product
5	0.004	0.05	.0002
10	0.008	0.05	.0004
15	0.012	0.1	.0012
20	0.016	0.15	.0024
100	0.08	0.1	.008
500	0.4	0.25	.1
1000	0.8	0.15	.12
1500	1.2	0.1	.12
2000	1.6	0.05	.08

Σ 1 .4322



WEB SERVER FILE REQUESTS

- Lets say that a web request is received every 500ms by the server.
  - What is the probability that the server will be idle?

$Pr(>=1)$ in system.

$\rho^K \qquad \rho = \frac{\lambda}{\mu}$

$\frac{1}{\lambda} = .5 \Rightarrow \lambda \Rightarrow 2$

$\frac{1}{\mu} = .4324 \Rightarrow \mu = 2.313$

$\left(\frac{2}{2.313}\right)^1 \Rightarrow 86.47\%$

Idle $\Rightarrow 1 - .8647 \Rightarrow .1353$

WHAT IS THE EXPECTED NUMBER OF ITEMS IN THE QUEUE?

$$N = \frac{\rho}{1-\rho} \Rightarrow \frac{\frac{2}{2.313}}{1 - \frac{2}{2.313}}$$

$$\boxed{N = 6.38}$$

WHAT IS THE MEAN RESPONSE TIME?

- $T = \frac{1/\mu}{1-\rho}$

$$T = \frac{\frac{1}{\mu}}{1-\rho} \Rightarrow \frac{\frac{1}{2.313}}{1-\frac{2}{2.313}} = 3.19$$

WHAT HAPPENS IF WE HAVE MULTIPLE PROCESSORS?

- M/M/∞ Queue

$$Average\ Number\ of\ customers = \frac{\lambda}{\mu} \Rightarrow \frac{2}{2.313} \Rightarrow .86$$

$$Average\ Time\ Spent = \frac{1}{\mu} \Rightarrow \frac{1}{\mu} \Rightarrow \frac{1}{2.313} = .4323$$

THE M/M/2 QUEUE

- Mean number of customers in the system:

$$L = \frac{2\rho}{1 - \rho^2}.$$

- Mean time to go through the system:

$$W = \frac{2\rho}{\lambda(1 - \rho^2)} = \frac{1}{\mu(1 - \rho^2)}.$$

- Mean waiting time in the queue:

$$W_q = W - \frac{1}{\mu} = \frac{\rho^2}{\mu(1 - \rho^2)}.$$

- Mean number of customers in the queue:

$$L_q = \lambda W_q = \frac{2\rho^3}{1 - \rho^2}.$$

WHAT IS THE MEAN RESPONSE TIME IF OUR SYSTEM HAS 2 THREADS PROCESSING REQUESTS?

$$\frac{\rho^{2}}{w-(1-\rho^{2})} = 1.7095$$