Queueing Theory (Part 3)

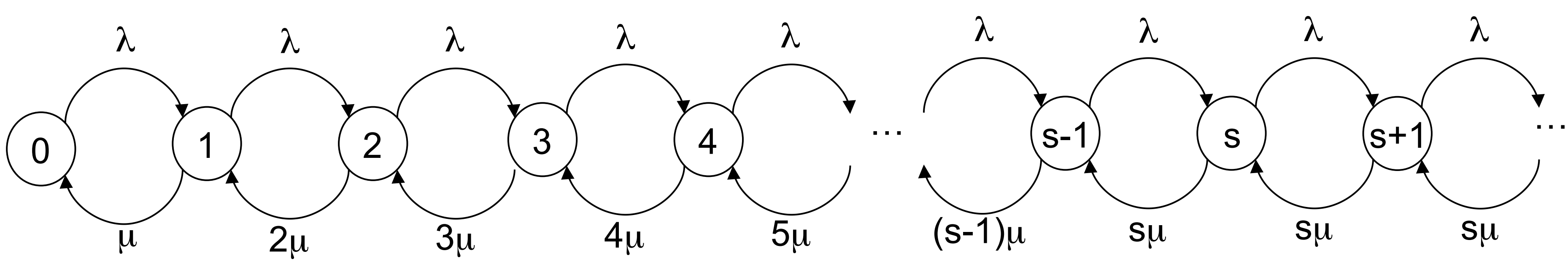
M/M/s Queueing Systems with Variations

(M/M/s, M/M/s//K, M/M/s///N)

M/M/s Queueing System

- We define
  λ = mean arrival rate
  μ = mean service rate
  s = number of servers (s > 1)
  $\rho = \lambda / s\mu$ = utilization ratio
- We require $\lambda < s\mu$, that is $\rho < 1$ in order to have a steady state

Rate Diagram



M/M/s Queueing System

Steady-State Probabilities

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!}\left(\frac{1}{1-\lambda/s\mu}\right)}$$

and $\quad P_n = C_n P_0$

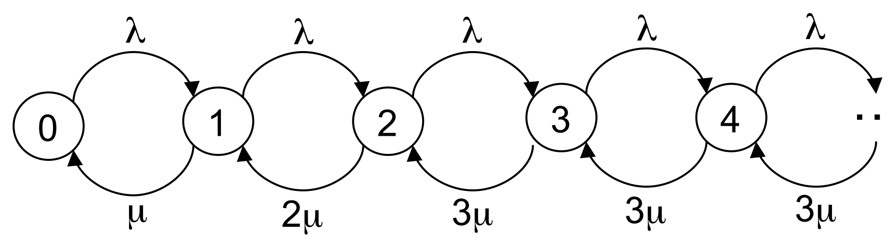
where $C_n = \frac{\lambda_{n-1}\lambda_{n-2}\ldots\lambda_0}{\mu_n\mu_{n-1}\ldots\mu_1} = \begin{cases} \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{n!} & n = 1, 2, \ldots, s \\[2ex] \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{s!\,s^{n-s}} & n = s+1, s+2, \ldots \end{cases}$

Use Birth Death Processes

Rate In = Rate Out

Coefficients are easy to remember if you think of rate diagram

Example: s = 3



$$C_0 = 1$$

$$C_1 = \frac{\lambda}{\mu}$$

$$C_2 = \frac{\lambda}{\mu}\frac{\lambda}{2\mu} = \left(\frac{\lambda}{\mu}\right)^2 \frac{1}{2!}$$

$$C_3 = \left(\frac{\lambda}{\mu}\right)^3 \frac{1}{3!}$$

$$C_4 = \left(\frac{\lambda}{\mu}\right)^4 \frac{1}{3}\frac{1}{3!}$$

$$C_5 = \left(\frac{\lambda}{\mu}\right)^5 \left(\frac{1}{3}\right)^2 \frac{1}{3!}$$

$$C_n = \left(\frac{\lambda}{\mu}\right)^n \left(\frac{1}{s}\right)^{n-s} \frac{1}{s!}$$

M/M/s Queueing System

L, L_q, W, W_q

$$L_q = \frac{P_0(\lambda/\mu)^s \rho}{s!(1-\rho)^2}$$

$$= \frac{P_0 \lambda^{s+1}}{(s-1)!\mu^{s-1}(s\mu-\lambda)^2}$$

$$P(\omega > t) = e^{-\mu t}\left[1 + \frac{P_0(\lambda/\mu)^s}{s!(1-\rho)}\left(\frac{1-e^{-\mu t(s-1-\lambda/\mu)}}{s-1-\lambda/\mu}\right)\right]$$

$$P(\omega_q > t) = \left(1 - \sum_{n=0}^{s-1} P_n\right) e^{-s\mu(1-\rho)t}$$

If $s - 1 - \lambda/\mu = 0$
then this term is (μt)

How to find L? W? W_q?

Use L_q to find W_q ($L_q = \lambda W_q$):

$W_q = L_q/\lambda$

Use W_q to find W:

$W = W_q + 1/\mu$

Use $L=\lambda W$ to find L in terms of L_q:

$L = \lambda W$

$= \lambda(W_q + 1/\mu)$

$= \lambda(L_q/\lambda + 1/\mu)$

$= L_q + \lambda/\mu$

M/M/s Example: A Better ER

- As before, we have
  - Average arrival rate = 1 patient every ½ hour
    λ = 2 patients per hour
  - Average service time = 20 minutes to treat each patient
    μ = 3 patients per hour
- Now we have 2 doctors
  $s = 2$
- Utilization

  $\rho = \lambda/2\mu = 2/6 = 1/3$ (Before $s=1$, $\rho=2/3$)

M/M/s Example: ER

Questions

In steady state, what is the…

1. probability that both doctors are idle?

   probability that exactly one doctor is idle?

2. probability that there are *n* patients?

3. expected number of patients in the ER?

M/M/s Example: ER
Questions

In steady state, what is the…

1. probability that both doctors are idle?

$$P_0 = \frac{1}{\frac{\left(\frac{\lambda}{\mu}\right)^0}{0!} + \frac{\left(\frac{\lambda}{\mu}\right)^1}{1!} + \frac{\left(\frac{\lambda}{\mu}\right)^2}{2!}\frac{1}{1-\rho}} = \frac{1}{1 + \frac{2}{3} + \frac{4}{9\cdot 2}\frac{1}{2/3}} = \frac{1}{2}$$

probability that exactly one doctor is idle?

$$P_1 = \frac{(\lambda/\mu)^1}{1!}P_0 = \frac{2}{3}\frac{1}{2} = \frac{1}{3}$$

2. probability that there are *n* patients?

$$P_n = \begin{cases} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}P_0 = \left(\frac{2}{3}\right)^n \frac{1}{n!}\frac{1}{2} & \text{if } 0 \le n < 2 \\ \frac{\left(\frac{\lambda}{\mu}\right)^n}{s!\,s^{n-s}}P_0 = \left(\frac{2}{3}\right)^n \frac{1}{2!}\left(\frac{1}{2}\right)^{n-2}\frac{1}{2} = \left(\frac{1}{3}\right)^n & \text{if } n \ge 2 \end{cases}$$

3. expected number of patients in the ER?

$$L = \lambda W = \lambda\left(L_q/\lambda + 1/\mu\right) = L_q + \lambda/\mu = 1/12 + 2/3 = 3/4$$

M/M/s Example: ER

Questions

In steady state, what is the…

4. expected number of patients waiting for a doctor?

5. expected time in the ER?

6. expected waiting time?

7. probability that there are at least two patients waiting in queue?

   probability that a patient waits more than 30 minutes?

M/M/s Example: ER

Questions

In steady state, what is the…

4. expected number of patients waiting for a doctor?

$$L_q = \frac{P_0(\lambda/\mu)^s \rho}{s!(1-\rho)^s} = \frac{(1/2)(2/3)^2(1/3)}{2!(2/3)^2} = \frac{1}{12}$$

5. expected time in the ER?

   W = L/λ = (3/4)/2 = 3/8 hour ≈ 22.5 minutes

6. expected waiting time?

   $W_q = L_q/\lambda$ = (1/12)/2 = 1/24 hour ≈ 2.5 minutes

7. probability that there are at least two patients waiting in queue?

   P(≥ 4 patients in system) $= 1 - P_0 - P_1 - P_2 - P_3$

   = 1 – ½ - 1/3 – 1/9 – 1/27 ≈ 0.0185

8. probability that a patient waits more than 30 minutes?

$$P(\omega_q > t) = (1 - P_0 - P_1)e^{-2\mu(1-\rho)t} = \left(1 - \frac{1}{2} - \frac{1}{3}\right)e^{-2(3)(2/3)t} = \frac{1}{6}e^{-4t}$$

$$P(\omega_q > 30\,\text{min}) = P\left(\omega_q > \frac{1}{2}\text{hour}\right) \approx 0.022$$

Performance Measurements	s = 1	s = 2
ρ	2/3	1/3
L	2	3/4
Lq	4/3	1/12
W	1 hr	3/8 hr
Wq	2/3 hr	1/24 hr
P(at least two patients waiting in queue)	0.296	0.0185
P(a patient waits more than 30 minutes)	0.404	0.022

Travel Agency Example

- Suppose customers arrive at a travel agency according to a Poisson input process and service times have an exponential distribution
- We are given
  - λ= 0.10/minute, that is, 1 customer every 10 minutes
  - μ =0.08/minute, that is, 8 customers every 100 minutes
- If there was only one server, what would happen?

  $\lambda/\mu > 1$

  Customers would balk at long lines – never reach steady state

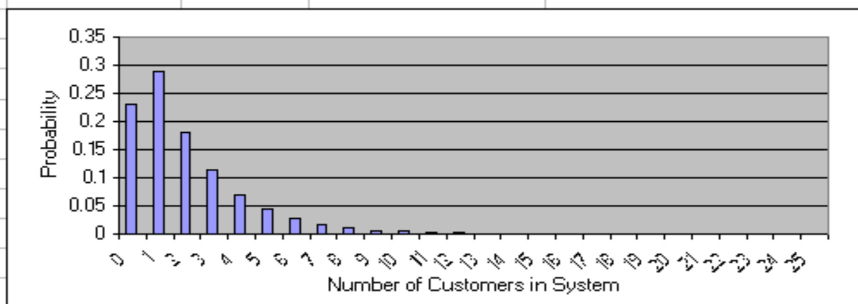
  - lose customers

  - go out of business?
- How many servers would you recommend?

  Calculate P_0, L_q and W_q for s=2, s=3, and s=4

Template for the M/M/s Queueing Model

	Data	
λ =	0.1	(mean arrival rate)
μ =	0.08	(mean service rate)
s =	2	(# servers)
$P(\omega>t)$ =	0.958978	
when t =	1	
$P(\omega_q>t)$ =	0.452771	
when t =	1	



	Results
L =	2.051282051
L_q =	0.801282051
W =	20.51282051
W_q =	8.012820513
ρ =	0.625

n	P_n
0	0.230769231
1	0.288461538
2	0.180288462
3	0.112680288
4	0.07042518
5	0.044015738
6	0.027509836
7	0.017193648
8	0.01074603
9	0.006716269
10	0.004197668
11	0.002623542
12	0.001639714
13	0.001024821
14	0.000640513
15	0.000400321
16	0.000250201
17	0.000156375
18	9.77346E-05
19	6.10841E-05
20	3.81776E-05
21	2.3861E-05
22	1.49131E-05
23	9.3207E-06
24	5.82543E-06
25	3.6409E-06

Range Name	Cells
L	G4
Lambda	C4
Lq	G5
Mu	C5
n	F13:F38
P0	G13
Pn	G13:G38
Rho	G10
s	C6
Time1	C9
Time2	C12
W	G7
Wq	G8

Template for the M/M/s Queueing Model

	Data	
λ =	0.1	(mean arrival rate)
μ =	0.08	(mean service rate)
s =	3	(# servers)
$P(\omega>t)$ =	0.93426	
when t =	1	
$P(\omega_q>t)$ =	0.135161	
when t =	1	

	Results
L =	1.361051883
L_q =	0.111051883
W =	13.61051883
W_q =	1.110518834
ρ =	0.416666667

n	P_n
0	0.278606965
1	0.348258706
2	0.217661692
3	0.090692371
4	0.037788488
5	0.015745203
6	0.006560501
7	0.002733542
8	0.001138976
9	0.000474573
10	0.000197739
11	8.23912E-05
12	3.43297E-05
13	1.4304E-05
14	5.96001E-06
15	2.48334E-06
16	1.03472E-06
17	4.31135E-07
18	1.7964E-07
19	7.48498E-08
20	3.11874E-08
21	1.29948E-08
22	5.41449E-09
23	2.25604E-09
24	9.40015E-10
25	3.91673E-10

Range Name	Cells
L	G4
Lambda	C4
Lq	G5
Mu	C5
n	F13:F38
P0	G13
Pn	G13:G38
Rho	G10
s	C6
Time1	C9
Time2	C12
W	G7
Wq	G8

Template for the M/M/s Queueing Model

	Data	
λ =	0.1	(mean arrival rate)
μ =	0.08	(mean service rate)
s =	4	(# servers)
$P(\omega>t)$ =	0.926026	
when t =	1	
$P(\omega_q>t)$ =	0.033881	
when t =	1	

	Results
L =	1.269190145
L_q =	0.019190145
W =	12.69190145
W_q =	0.191901452
ρ =	0.3125

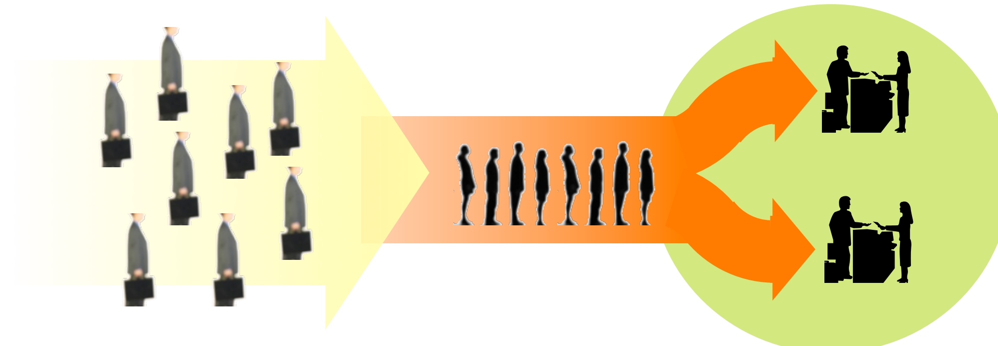
n	P_n
0	0.28532829
1	0.356660362
2	0.222912726
3	0.092880303
4	0.029025095
5	0.009070342
6	0.002834482
7	0.000885776
8	0.000276805
9	8.65015E-05
10	2.70317E-05
11	8.44741E-06
12	2.63982E-06
13	8.24943E-07
14	2.57795E-07
15	8.05608E-08
16	2.51753E-08
17	7.86727E-09
18	2.45852E-09
19	7.68288E-10
20	2.4009E-10
21	7.50281E-11
22	2.34463E-11
23	7.32696E-12
24	2.28968E-12
25	7.15524E-13

Range Name	Cells
L	G4
Lambda	C4
Lq	G5
Mu	C5
n	F13:F38
P0	G13
Pn	G13:G38
Rho	G10
s	C6
Time1	C9
Time2	C12
W	G7
Wq	G8

Single Queue vs. Multiple Queues

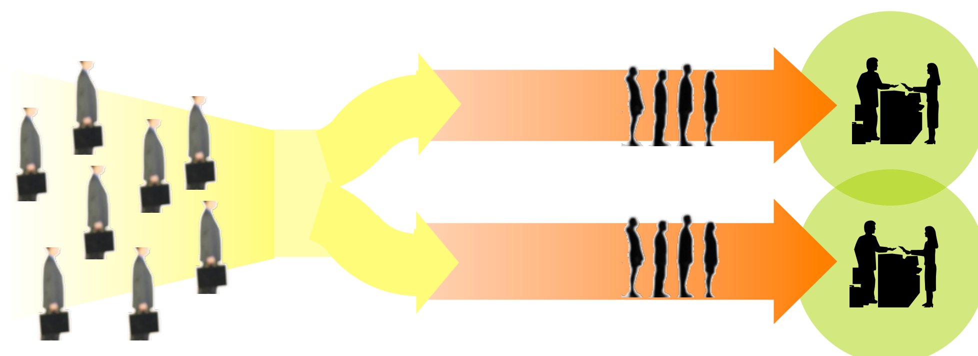
- Would you ever want to keep separate queues for separate servers?

Single queue



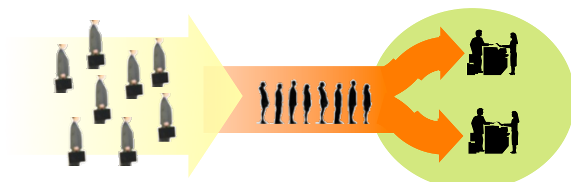
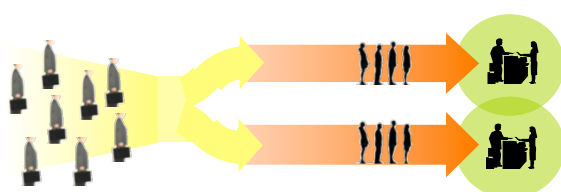
vs.

Multiple queues



Bank Example

- Suppose we have two tellers at a bank
- Compare the single server and multiple server models
- Assume λ = 2, μ = 3,

	L	L_q	W	W_q	P_0	ρ
	0.75	0.083	0.375	0.042	0.5	$\lambda/2\mu$ $=1/3$
	1.0	0.334	0.5	0.167	0.4449	$\lambda`/\mu$ $=(\lambda/\mu)/3$ $=1/3$

Bank Example
Continued

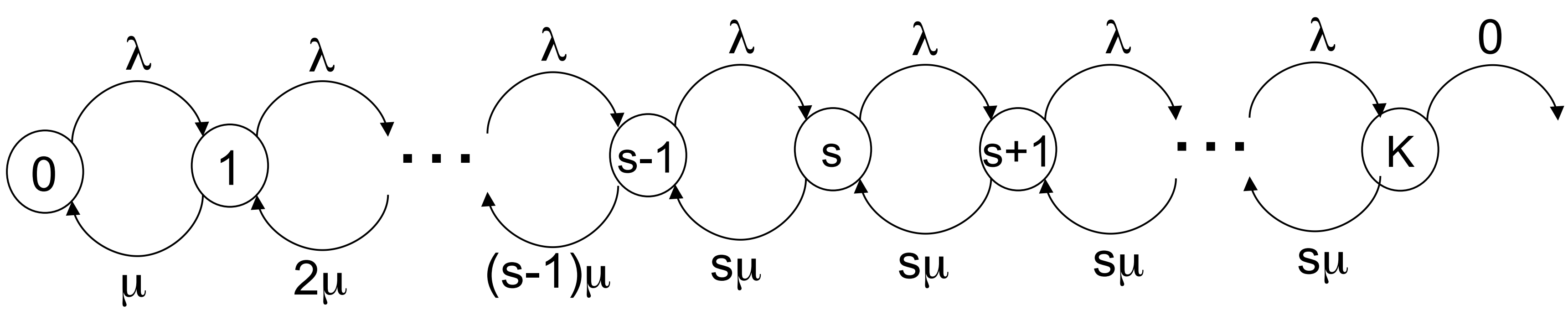
- Suppose we now have 3 tellers
- Again, compare the two models

M/M/3	Three M/M/1 queues	
$\lambda=2,\ \mu=3$	$\lambda' = \lambda/3 = 2/3,\ \mu=3$	
$\rho=\lambda/(s\mu) = 2/9$	M/M/1: $\rho=\lambda'/3 = 2/9$	ρ is the same
$L = 0.676$	$L=0.286$	$3L = 0.858$
$L_q = 0.009$	$L_q=0.063$	$3L_q = 0.189$
$W = 0.338$	$W = 0.429$	
$W_q = 0.005$	$W_q = 0.095$	
$P_0 = 0.5122$	$P_0 = 0.7778$	$(P_0)^3 = 0.47$

M/M/s//K Queueing Model

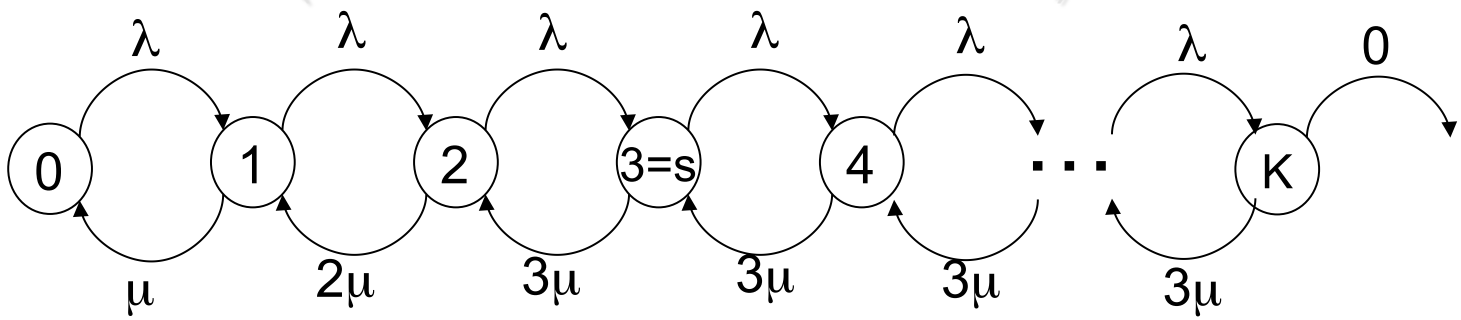
(Finite Queue Variation of M/M/s)

- Now suppose the system has a maximum capacity, *K*
- We will still consider s servers
- Assuming $s \leq K$, the maximum queue capacity is $K - s$
- Some applications for this model:
  - Trunk lines for phone – call center
  - Warehouse with limited storage
  - Parking garage
- Draw the rate diagram for this problem:



M/M/s//K Queueing Model

(Finite Queue Variation of M/M/s)



Balance equations: Rate In = Rate Out

State 0: $\mu P_1 = \lambda P_0$

State 1: $\lambda P_0 + 2\mu P_2 = (\lambda+\mu)P_1$

State 2: $\lambda P_1 + 3\mu P_3 = (\lambda+2\mu)P_2$

State 3: $\lambda P_2 + 3\mu P_4 = (\lambda+3\mu)P_3$

⋮

State K-1: $\lambda P_{K-2} + 3\mu P_K = (\lambda+3\mu)P_{K-1}$

State K: $\lambda P_{K-1} = 3\mu P_K$

$$C_0 = 1$$

$$C_1 = \frac{\lambda}{\mu}$$

$$C_2 = \frac{\lambda^2}{2\mu^2}$$

$$C_3 = \frac{\lambda^3}{3!\mu^3} \qquad (s = 3)$$

$$C_4 = \left(\frac{1}{3!\cdot 3}\right)\left(\frac{\lambda}{\mu}\right)^4$$

$$\vdots$$

$$C_n = \left(\frac{1}{3!\cdot 3^{(n-s)}}\right)\left(\frac{\lambda}{\mu}\right)^n \qquad \text{for } s \le n \le K$$

$$C_{K+1} = 0$$

$$P_0 = \frac{1}{\sum_{n=0}^{K} C_n}$$

$$P_n = C_n P_0$$

M/M/s//K Queueing Model

(Finite Queue Variation of M/M/s)

Solving the balance equations, we get the following steady state probabilities:

$$P_0 = \frac{1}{1 + \sum_{n=1}^{s} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s+1}^{K} \left(\frac{\lambda}{s\mu}\right)^{n-s}}$$

$$P_n = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} P_0 & \text{for } n = 1, 2, \ldots, s \\[2ex] \dfrac{\lambda^n}{s^{n-s} s!\,\mu^n} P_0 & \text{for } n = s, s+1, \ldots, K \\[2ex] 0 & n > K \end{cases}$$

Verify that these equations match those given in the text for the single server case (M/M/1//K)

M/M/s//K Queueing Model

(Finite Queue Variation of M/M/s)

$$L_q = \frac{P_0 (\lambda/\mu)^s \rho}{s!(1-\rho)^2}[1-\rho^{K-s} - (K-s)\rho^{K-s}(1-\rho)], \quad \text{where } \rho = \lambda / s\mu$$

$$L = \sum_{n=0}^{s-1} nP_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n\right)$$

To find W and W_q:
Although $L \neq \lambda W$ and $L_q \neq \lambda W_q$ because λ_n is **<u>not</u>** equal for all n,

$$L = \bar{\lambda} W \quad \text{and} \quad L_q = \bar{\lambda} W_q \quad \text{where} \quad \bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \lambda(1 - P_K)$$

Also, because there is a finite number of states, the steady state equations do hold, even if $\rho>1$

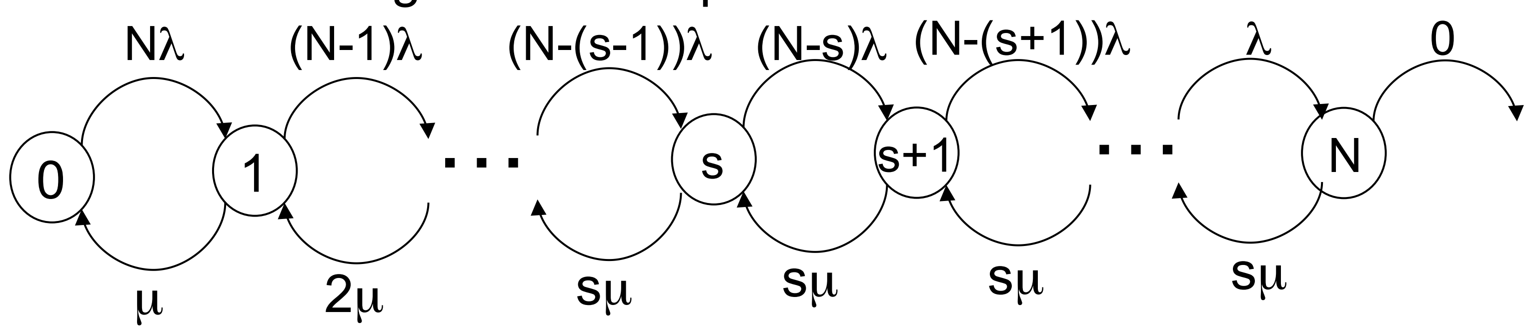
M/M/s///N Queueing Model

(Finite Calling Population Variation of M/M/s)

- Now suppose the calling population is finite, *N*
- We will still consider *s* servers
- Assuming $s \le N$, the maximum number in the queue capacity is $N - s$, so $K \ge N$ does not affect anything

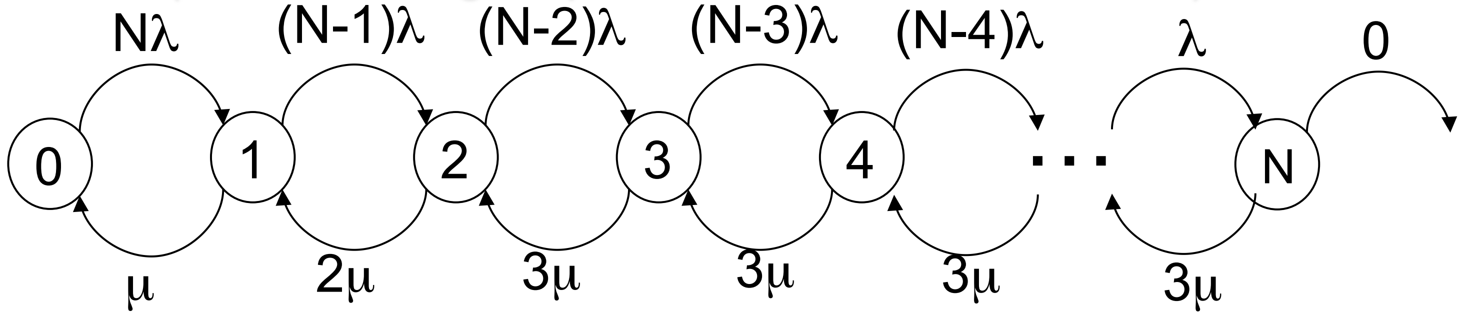
  If *N* is the entire population, then the maximum number in system is *N*. Assume $N \le K$ and $s \le N$
- Application for this model:

  Machine replacement
- Draw the rate diagram for this problem:



M/M/s///N Queueing Model

(Finite Calling Population Variation of M/M/s)



Balance equations: Rate In = Rate Out

State 0: $\mu P_1 = \lambda P_0$ ➔ $P_1 = (N\lambda/\mu)P_0$

State 1: $N\lambda P_0 + 2\mu P_2 = ((N-1)\lambda+\mu)P_1$ ➔ $P_2 = (1/2)(N\lambda/\mu)\,((N-1)\lambda/\mu)P_0$

⋮

$$C_0 = 1$$

$$C_1 = N\left(\frac{\lambda}{\mu}\right)$$

$$C_2 = \frac{N(N-1)}{2}\left(\frac{\lambda}{\mu}\right)^2$$

$$C_3 = \frac{N(N-1)(N-2)}{3!}\left(\frac{\lambda}{\mu}\right)^3$$

M/M/s///N Results

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{N!}{(N-n)!\,n!}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{N} \frac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n}$$

$$P_n = \begin{cases} \frac{N!}{(N-n)!\,n!}\left(\frac{\lambda}{\mu}\right)^n P_0 & \text{for} \quad n = 0,1,\ldots,s \\ \frac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n P_0 & \text{for} \quad s \le \mathrm{n} \le \mathrm{N} \\ 0 & \text{for} \quad \mathrm{n} > \mathrm{N} \end{cases}$$

$$L_q = \sum_{n=s}^{N} (n-s) P_n \qquad L = \sum_{n=0}^{s-1} n P_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n\right)$$