Queueing Systems: Lecture 2

Amedeo R. Odoni

October 13, 2004

Lecture Outline

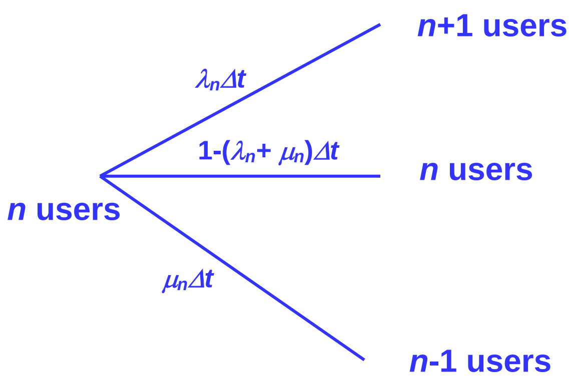
- Birth-and-death processes
- State transition diagrams
- Steady-state probabilities
- M/M/1
- M/M/m
- M/M/∞

Reference: Chapter 4, pp. 194-206

Birth-and-Death Queueing Systems

1. m parallel, identical servers.
2. Infinite queue capacity (for now).
3. Whenever *n* users are in system (in queue plus in service) arrivals are Poisson at rate of λ_n per unit of time.
4. Whenever *n* users are in system, service completions are Poisson at rate of μ_n per unit of time.
5. FCFS discipline (for now).

The Fundamental Relationship



$P_n(t)$ = Prob [*n* users in system at time *t*]

$$P_n(t+\Delta t) = P_{n+1}(t)\cdot\mu_{n+1}\cdot\Delta t + P_{n-1}(t)\cdot\lambda_{n-1}\cdot\Delta t + P_n(t)\cdot[1-(\mu_n+\lambda_n)\cdot\Delta t]$$

The differential equations that determine the state probabilities

$$P_n(t+\Delta t) = P_{n+1}(t)\cdot\mu_{n+1}\cdot\Delta t + P_{n-1}(t)\cdot\lambda_{n-1}\cdot\Delta t + P_n(t)\cdot[1-(\mu_n+\lambda_n)\cdot\Delta t]$$

After a simple manipulation:

$$\frac{dP_n(t)}{dt} = -(\lambda_n+\mu_n)\cdot P_n(t) + \lambda_{n-1}\cdot P_{n-1}(t) + \mu_{n+1}\cdot P_{n+1}(t) \qquad \textbf{(1)}$$

(1) applies when *n* = 1, 2, 3,….; when *n* = 0, we have:

$$\frac{dP_0(t)}{dt} = -\lambda_0\cdot P_0(t) + \mu_1\cdot P_1(t) \qquad \textbf{(2)}$$

• The system of equations (1) and (2) is known as the Chapman-Kolmogorov equations for a birth-and-death system

The “state balance” equations

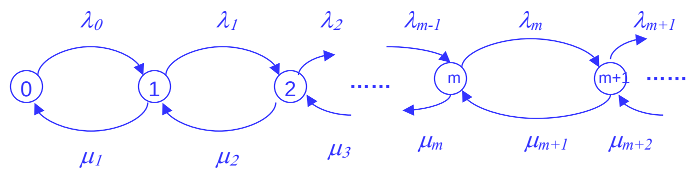
- **We now consider the situation in which the queueing system has reached “steady state”, i.e., *t* is large enough to have $P_n(t) = P_n$, independent of t, or** $\frac{dP_n(t)}{dt} = 0$
- **Then, (1) and (2) provide the state balance equations:**

$$\lambda_0\cdot P_0 = \mu_1\cdot P_1 \qquad n=0 \qquad (3)$$

$$(\lambda_n+\mu_n)\cdot P_n = \lambda_{n-1}\cdot P_{n-1} + \mu_{n+1}\cdot P_{n+1} \qquad n=1,2,3,.. \qquad (4)$$

- **The state balance equations can also be written directly from the state transition diagram**

Birth-and-Death System: State Transition Diagram



- **We are interested in the characteristics of the system under equilibrium conditions ("steady state"), i.e., when the state probabilities $P_n(t)$ are independent of t for large values of t**

- **Can write system balance equations and obtain closed form expressions for P_n, L, W, L_q, W_q**

Solving…..

Solving (3) and (4), we have:

$$P_1 = \frac{\lambda_0}{\mu_1} \cdot P_0; \quad P_2 = \frac{\lambda_1}{\mu_2} \cdot P_1 = \frac{\lambda_1 \cdot \lambda_0}{\mu_2 \cdot \mu_1} \cdot P_0 \qquad etc.$$

and, in general,

$$P_n = \frac{\lambda_{n-1} \cdot \lambda_{n-2} \cdot \ldots\ldots \cdot \lambda_1 \cdot \lambda_0}{\mu_n \cdot \mu_{n-1} \cdot \ldots\ldots \cdot \mu_2 \cdot \mu_1} \cdot P_0 = K_n \cdot P_0$$

But, we also have: $1 = \sum_{n=0}^{\infty} P_n = P_0 \cdot (1 + \sum_{n=1}^{\infty} K_n)$

Giving, $P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} K_n}$

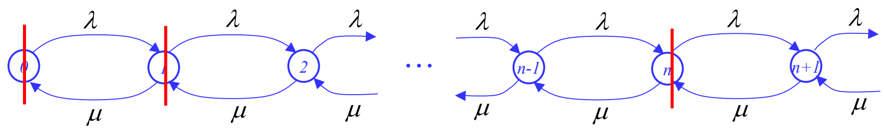
Condition for steady state:

$$\sum_{n=1}^{\infty} K_n < \infty$$

M/M/1: Observing State Transition Diagram from Two Points

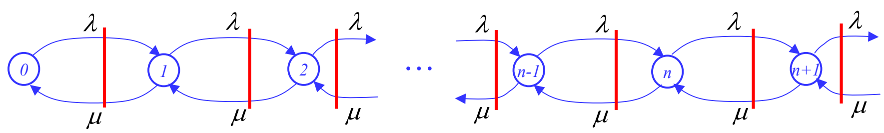
- **From point 1:**

$\lambda P_0 = \mu P_1 \quad (\lambda+\mu)P_1 = \lambda P_0 + \mu P_2 \qquad (\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+1}$



- **From point 2:**

$\lambda P_0 = \mu P_1 \qquad \lambda P_1 = \mu P_2 \qquad \lambda P_n = \mu P_{n+1}$



M/M/1: Derivation of P_0 and P_n

Step 1: $P_1 = \frac{\lambda}{\mu}P_0, \quad P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0, \quad \cdots, \quad P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$

Step 2: $\sum_{n=0}^{\infty} P_n = 1, \quad \Rightarrow \quad P_0 \sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = 1 \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n}$

Step 3: $\rho = \frac{\lambda}{\mu}$, then $\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty}\rho^n = \frac{1-\rho^{\infty}}{1-\rho} = \frac{1}{1-\rho} \quad (\because \rho < 1)$

Step 4: $P_0 = \frac{1}{\sum_{n=0}^{\infty}\rho^n} = 1-\rho$ and $P_n = \rho^n(1-\rho)$

M/M/1: Derivation of *L*, *W*, *W*$_q$, and *L*$_q$

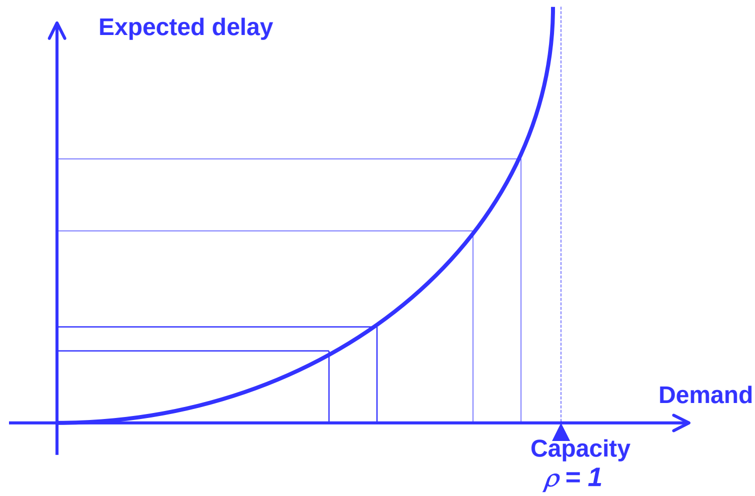
- $L = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n (1-\rho) = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho\sum_{n=1}^{\infty} n\rho^{n-1}$

$$= (1-\rho)\rho \frac{d}{d\rho}\left(\sum_{n=0}^{\infty}\rho^n\right) = (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

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

- $W = \frac{L}{\lambda} = \frac{\lambda}{\mu-\lambda}\cdot\frac{1}{\lambda} = \frac{1}{\mu-\lambda}$
- $W_q = W - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$
- $L_q = \lambda W_q = \lambda\cdot\frac{\lambda}{\mu(\mu-\lambda)} = \frac{\lambda^2}{\mu(\mu-\lambda)}$

High Sensitivity of Delay at High Levels of Utilization



M/M/1: An alternative, direct derivation of *L* and *W*

- For an M/M/1 system, with FCFS discipline:

$$W = \sum_{n=0}^{\infty} \frac{(n+1)}{\mu} \cdot P_n = E[\frac{n+1}{\mu}] = \frac{E[n]+1}{\mu} = \frac{L+1}{\mu} \quad \textbf{(1)}$$

- But from Little's theorem we also have:

$$L = \lambda \cdot W \qquad \textbf{(2)}$$

- It follows from (1) and (2) that, as before:

$$L = \frac{\lambda}{\mu - \lambda}; \qquad W = \frac{1}{\mu - \lambda}$$

Does the queueing discipline matter?

Additional important M/M/1 results

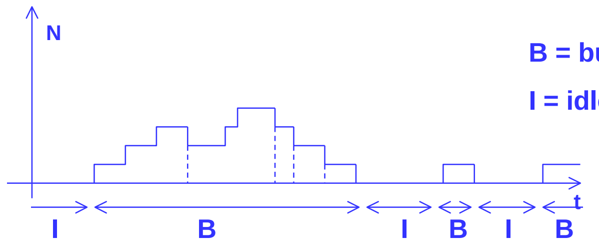
- The pdf for the total time in the system, *w,* can be computed for a M/M/1 system (and FCFS):

$$f_w(w) = (1-\rho)\mu e^{-(1-\rho)\mu w} = (\mu - \lambda)e^{-(\mu-\lambda)w} \text{ for } w \geq 0$$

Thus, as already shown, $W = 1/(\mu - \lambda) = 1/[\mu(1-\rho)]$

- The standard deviation of N, w, N_q, w_q are all proportional to $1/(1-\rho)$, just like their expected values (L, W, L_q, W_q, respectively)
- The expected length of the "busy period", *E[B]*, is equal to $1/(\mu - \lambda)$

M/M/1: *E*[*B*], the expected length of a busy period



B = busy period

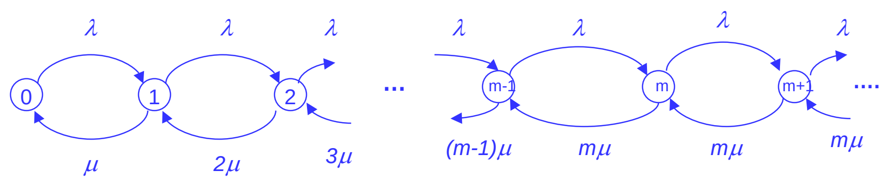
I = idle period

$$P_0 = \frac{E[Idle]}{E[Busy] + E[Idle]}$$

But, $P_0 = 1 - \rho \qquad E[Idle] = 1/\lambda$

Therefore, $E[B] = E[Length\ of\ busy\ period] = \frac{1}{\mu} \cdot \frac{1}{(1-\rho)} = \frac{1}{\mu - \lambda}$

M/M/m (infinite queue capacity)

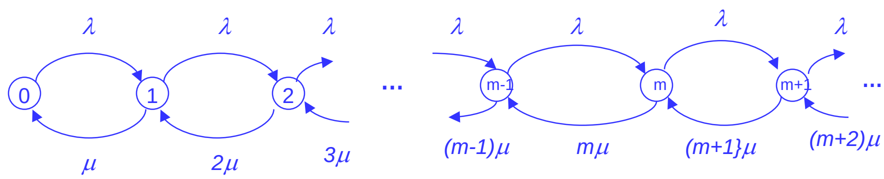


$$P_n = \frac{(\frac{\lambda}{\mu})^n}{n!} P_0 \quad for \quad n = 0, 1, 2, ..., m-1$$

$$P_n = \frac{(\frac{\lambda}{\mu})^n}{m^{n-m} \cdot m!} P_0 \quad for\ n = m, m+1, m+2, ...$$

- **Condition for steady state?**

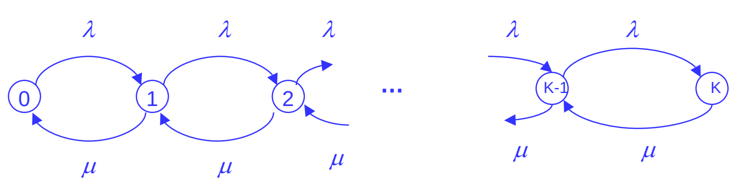
M/M/∞ (infinite no. of servers)



$$P_n = \frac{(\frac{\lambda}{\mu})^n \cdot e^{-(\frac{\lambda}{\mu})}}{n!} \quad for\ n = 0, 1, 2, ...$$

- ***N*** **is Poisson distributed!**
- ***L*** **=** λ / μ **;** ***W*** **= 1 / μ ;** L_q **= 0;** W_q **= 0**
- **Many applications**

M/M/1: finite system capacity, K; customers finding system full are lost

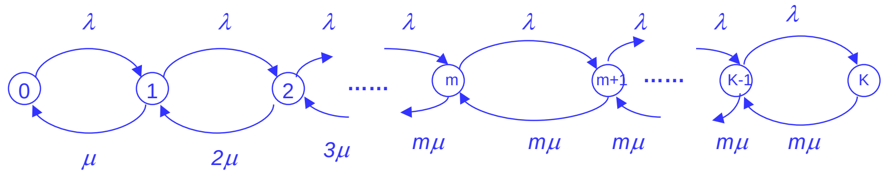


$$P_n = \frac{\rho^n \cdot (1-\rho)}{1-\rho^{K+1}} \quad for\ n = 0, 1, 2, ..., K$$

- **Steady state is always reached**
- **Be careful in applying Little's Law! Must count only the customers who actually join the system:**

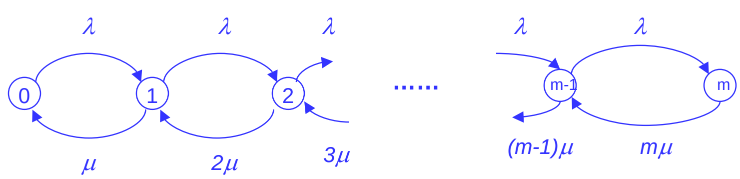
$$\lambda' = \lambda \cdot (1 - P_K)$$

M/M/m: finite system capacity, K; customers finding system full are lost



- Can write system balance equations and obtain closed form expressions for P_n, L, W, L_q, W_q
- Often useful in practice

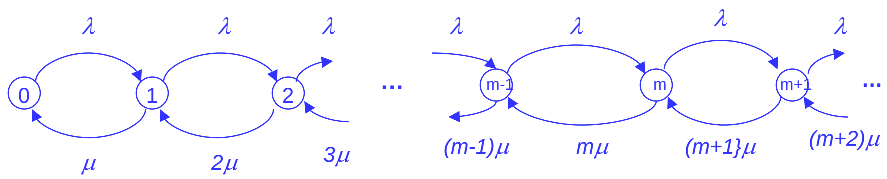
M/M/m: finite system capacity, m; special case!



$$P_n = \frac{\dfrac{(\frac{\lambda}{\mu})^n}{n!}}{\sum_{i=0}^{m} \dfrac{(\frac{\lambda}{\mu})^i}{i!}} \quad for\ n = 0, 1, 2, \ldots m$$

- Probability of full system, P_m, is "Erlang's loss formula"
- Exactly same expression for P_n of M/G/m system with K=m

M/M/∞ (infinite no. of servers)



$$P_n = \frac{(\frac{\lambda}{\mu})^n \cdot e^{-(\frac{\lambda}{\mu})}}{n!} \quad for\ n = 0, 1, 2, ...$$

- *N* is Poisson distributed!
- $L = \lambda / \mu$; $W = 1 / \mu$; $L_q = 0$; $W_q = 0$
- Many applications

Variations and extensions of birth-and-death queueing systems

- Huge number of extensions on the previous models
- Most common is arrival rates and service rates that depend on state of the system; some lead to closed-form expressions
- Systems which are not birth-and-death, but can be modeled by continuous time, discrete state Markov processes can also be analyzed [“phase systems”]
- State representation is the key (e.g. $M/E_k/1$)