Queuing Theory

Dr. Sk. Md. Masudul Ahsan
Associate Professor, Dept. of CSE, KUET

Queuing System (elements)

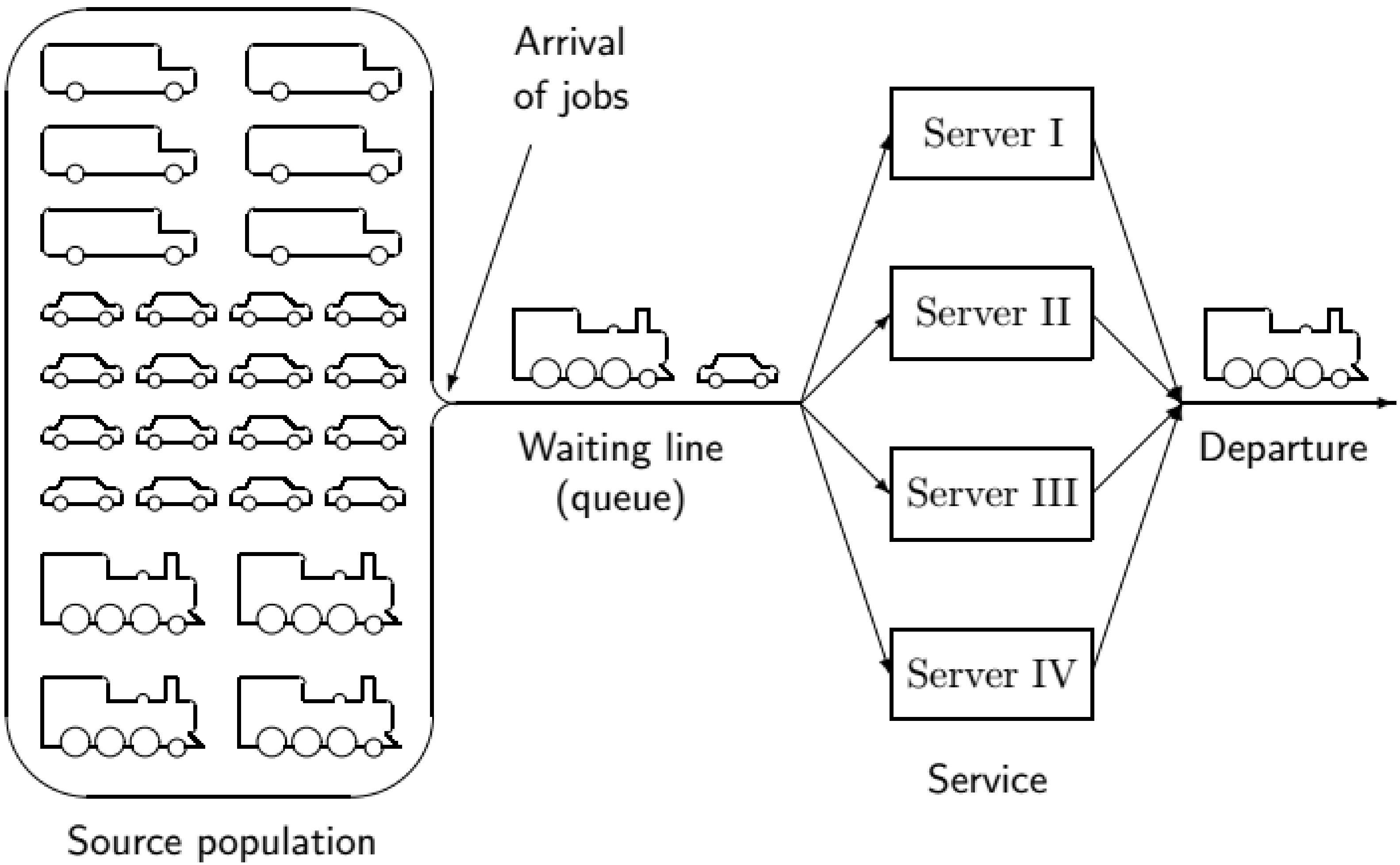
- Waiting for service is part of our daily life. We queue up for
  - Bus, check out counter etc.
- Not only for human, jobs wait to be processed on a machine
- Study of queue quantifies the phenomenon of waiting lines using
  - Average queue length, average waiting time, average facility utilization
  - Thereby reduce the adverse impact of waiting to tolerable levels

Queuing System

- A queuing system is a facility consisting of one or several servers designed to perform certain tasks or process certain jobs and a queue of jobs waiting to be processed.
- Examples of queuing systems are:
  - a shared computer executing tasks sent by its users;
  - an internet service provider whose customers connect to the internet, browse, and disconnect;
  - a printer processing jobs sent to it from different computers;
  - a medical office serving patients; and so on.

Queuing System

- Main components



Queuing System - Input process

- Usually called the **arrival process.**
- Arrivals are called **customers.**
- We assume that no more than one arrival can occur at a given instant.
- If more than one arrival can occur at a given instant, we say that **bulk arrivals** are allowed.
- Models in which arrivals are drawn from a small population are called **finite source models.**
- If a customer arrives but fails to enter the system, we say that the customer has **balked.**

Queuing System - Output process

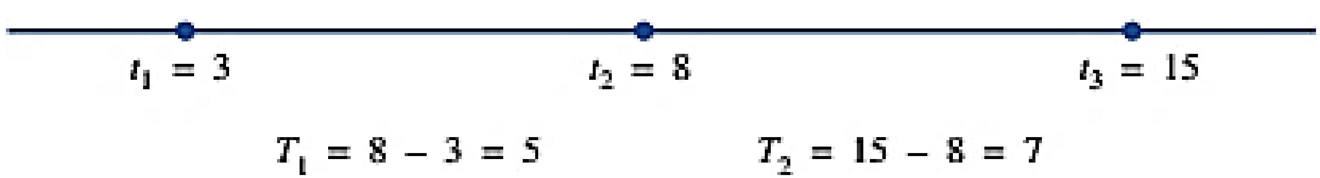
- To describe the output process of a queuing system, we usually specify a probability distribution – the **service time distribution** – which governs a customer's service time.
- Two arrangements of servers:
  - Servers are in **parallel** if all server provide the same type of service and a customer need only pass through one server to complete service.
  - Servers are in **series** if a customer must pass through several servers before completing service.

Queuing System - Queue discipline

- Describes the method used to determine the order in which customers are served.
- The most common - **FCFS discipline** (first come, first served), served in order of their arrival.
- **LCFS discipline** (last come, first served), the most recent arrivals are the first to enter service.
- Customer is randomly chosen from waiting line - **SIRO discipline** (service in random order).
- **Priority queuing disciplines** - classifies each arrival into one of several categories.
  - Each category is then given a priority level, and within each priority level served on an FCFS basis.

Modeling Arrival and Service Processes

- We define t_i to be the time at which the ith customer arrives.
- we define $T_i = t_{i+1} - t_i$ to be the ith interarrival time.
- In modeling the arrival process we assume that the T_i's are independent, continuous random variables described by **A** having a density function $a(t)$.
- The assumption that each interarrival time is governed by the same random variable implies that the distribution of arrivals is independent of the time of day or the day of the week.
- This is the assumption of stationary interarrival times.
- Stationary-(whose distribution characteristics do not change over time)



Modeling Arrival and Service Processes

- The independence assumption means, for example, that the value of T_2 has no effect on the value of T_3, T_4, or any later T_i
- Stationary interarrival ties is often unrealistic, but we may often approximate reality by breaking the time of day into segments.
- A negative interarrival time is impossible. This allows us to write

$$P(\mathbf{A} \leq c) = \int_0^c a(t)dt \text{ and } P(\mathbf{A} > c) = \int_c^\infty a(t)dt$$

- We define λ to be the **arrival rate** or units of arrivals per hour
- **So,** 1/λ to be the mean or average interarrival time or units of hours per arrival.

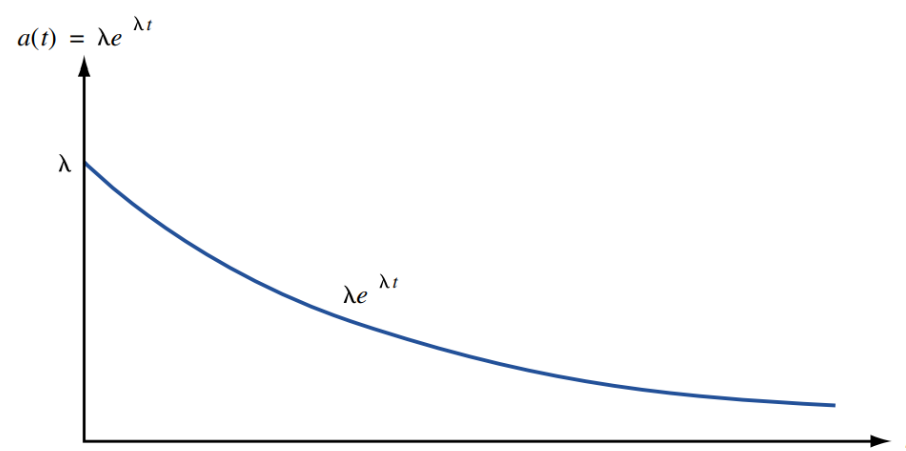
$$\frac{1}{\lambda} = \int_0^\infty t a(t)dt$$

Modeling Arrival and Service Processes

- How to choose **A** - to reflect reality and still be computationally tractable.
- The most common choice for **A** is the **exponential distribution.**
- An exponential distribution with parameter λ has a density

$$a(t) = \lambda e^{-\lambda t}.$$

- We can show that the average or mean by $E(\mathbf{A}) = \dfrac{1}{\lambda}$
- And Variance by $\operatorname{var} \mathbf{A} = \dfrac{1}{\lambda^2}$



Modeling Arrival and Service Processes

- Lemma 1: If **A** has an exponential distribution, then for all nonnegative values of *t* and *h*,

$$P(\mathbf{A} > t + h \mid \mathbf{A} \geq t) = P(\mathbf{A} > h) \qquad (5)$$

- For example, if h=4, then (5) yields, for t=5, t=3, t=2, and t=0

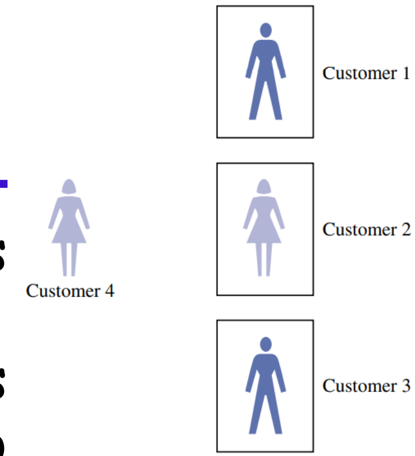
$$P(\mathbf{A} > 9 | \mathbf{A} \geq 5) = P(\mathbf{A} > 7 | \mathbf{A} \geq 3) = P(\mathbf{A} > 6 | \mathbf{A} \geq 2)$$
$$= P(\mathbf{A} > 4 | \mathbf{A} \geq 0) = e^{-4\lambda}$$

- For reasons that become apparent, a density that satisfies the equation is said to have the **no-memory property.**
- The no-memory property of the exponential distribution is important because it implies that if we want to know the probability distribution of the time until the next arrival, then *it does not matter how long it has been since the last arrival.*

Modeling the Service Process

- We assume that the service times of different customers are independent random variables and that each customer's service time is governed by a random variable **S** having a density function *s(t)*.
- We let 1/μ be then mean service time for a customer.
- so μ is service rate or μ has units of customers per hour.
- For example, μ=5 means that if customers were always present, the server could serve an average of 5 customers per hour, and the average service time of each customer would be 1/5 hour.
- As with interarrival times, we hope that service times can be accurately modeled as exponential random variables
- Unfortunately, actual service times may not be consistent with the no-memory property.

$$s(t) = \mu e^{-\mu t}$$



- Suppose all three servers are busy, and a customer is waiting
- What is the probability that the customer who is waiting will be the last of the four customers to complete service?
- One of customers 1–3 (say, customer 3) will be the first to complete service.
- Then customer 4 will enter service.
- By the no-memory property, customer 4's service time has the same distribution as the remaining service times of customers 1 and 2.
- Thus, by symmetry, customers 4, 1, and 2 will have the same chance of being the last customer to complete service.
- This implies that customer 4 has a 1/3 chance of being the last customer to complete service.
- Without the no-memory property, this problem would be hard to solve

The Kendall-Lee Notation for Queuing Systems

- Standard notation used to describe many queuing systems.
- The notation is used to describe a queuing system in which all arrivals wait in a single line until one of s identical parallel servers id free. Then the first customer in line enters service, and so on.
- To describe such a queuing system, Kendall devised the following notation.
- Each queuing system is described by six characters:

  1/2/3/4/5/6

The Kendall-Lee Notation for Queuing Systems

- The first characteristic specifies the nature of the arrival process. The following standard abbreviations are used:

  M = Interarrival times are independent, identically distributed (iid) having an exponential distribution.

  D = Interarrival times are iid and deterministic

  E_k = Interarrival times are iid Erlangs with shape parameter k.

  GI = Interarrival times are iid and governed by some general distribution

The Kendall-Lee Notation for Queuing Systems

- The second characteristic specifies the nature of the service times:

  M = Service times are iid and exponentially distributed

  D = Service times are iid and deterministic

  E_k = Service times are iid Erlangs with shape parameter *k*.

  G = Service times are iid and governed by some general distribution

The Kendall-Lee Notation for Queuing Systems

- The third characteristic is the number of parallel servers.
- The fourth characteristic describes the queue discipline:
  - FCFS = First come, first served
  - LCFS = Last come, first served
  - SIRO = Service in random order
- The fifth characteristic specifies the maximum allowable number of customers in the system.
- The sixth characteristic gives the size of the population from which customers are drawn.

The Kendall-Lee Notation for Queuing Systems

- In many important models 4/5/6 is $GD/\infty/\infty$. If this is the case, then 4/5/6 is often omitted.
- $M/E_2/8/\text{FCFS}/10/\infty$ might represent a health clinic with 8 doctors, exponential interarrival times, two-phase Erlang service times, an FCFS queue discipline, and a total capacity of 10 patients.

Birth-Death Processes

- We will use it to answer questions about several different types of queuing systems.
- Number of people present in any queuing system at time *t* to be the **state** of the queuing systems at time *t*.
- $P_{ij}(t)$ which is defined as the probability that j people will be present in the queuing system at time t, given that at time 0, i people are present.
- We call π_j the **steady state**, or equilibrium probability, of state *j*.
- The behavior of $P_{ij}(t)$ before the steady state is reached is called the **transient behavior** of the queuing system.
- A **birth-death process** is a continuous-time stochastic process for which the system's state at any time is a nonnegative integer.

Laws of Motion for Birth-Death

- Law 1
  - With probability $\lambda_j \Delta t + o(\Delta t)$, a birth occurs between time t and time $t+\Delta t$.
  - A birth increases the system state by 1, to $j+1$.
  - The variable λ_j is called the **birth rate** in state j.
  - In most queuing systems, a birth is simply an arrival.
- Law 2
  - With probability $\mu_j \Delta t + o(\Delta t)$, a death occurs between time t and time $t + \Delta t$. A death decreases the system state by 1, to j-1. The variable μ_j is the death rate in state j. In most queuing systems, a death is a service completion.
  - Note that $\mu_0 = 0$ must hold, or a negative state could occur.
- Law 3
  - Births and deaths are independent of each other.

†Recall from Section 20.2 that $o(\Delta t)$ means that $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

- M/M/1/FCFS/∞/∞ queuing system in which interarrival times are exponential with parameter λ and service times are exponentially distributed with parameter μ.
- If the state (number of people present) at time t is j, then the no-memory property of the exponential distribution implies that the probability of a birth during the time interval [t,t+ Δt] will not depend on how long the system has been in state j.
- This means that the probability of a birth occurring during [t,t+ Δt] will not depend on how long the system has been in state j and thus may be determined as if an arrival had just occurred at time t.
- Then the probability of a birth occurring during [t,t+ Δt] is

$$\int_0^{\Delta t} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda \Delta t}$$

By the Taylor series expansion given in Section 11.1,

$$e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t)$$

$[t, t + \Delta t]$ is $\lambda \Delta t + o(\Delta t)$.

M/M/1/FCFS/∞/∞ queuing system

If the state at time t is $j \geq 1$, then we know (since there is only one server) that exactly one customer will be in service. The no-memory property of the exponential distribution then implies that the probability that a customer will complete service between t and $t + \Delta t$ is given by

$$\int_0^{\Delta t} \mu e^{-\mu t} dt = 1 - e^{-\mu \Delta t} = \mu \Delta t + o(\Delta t)$$

Thus, for $j \geq 1$, $\mu_j = \mu$. In summary, if we assume that service completions and arrivals occur independently, then an $M/M/1$/FCFS/∞/∞ queuing system is a birth–death process. The birth and death rates for the $M/M/1$/FCFS/∞/∞ queuing system may be represented in a rate diagram (see Figure 9).

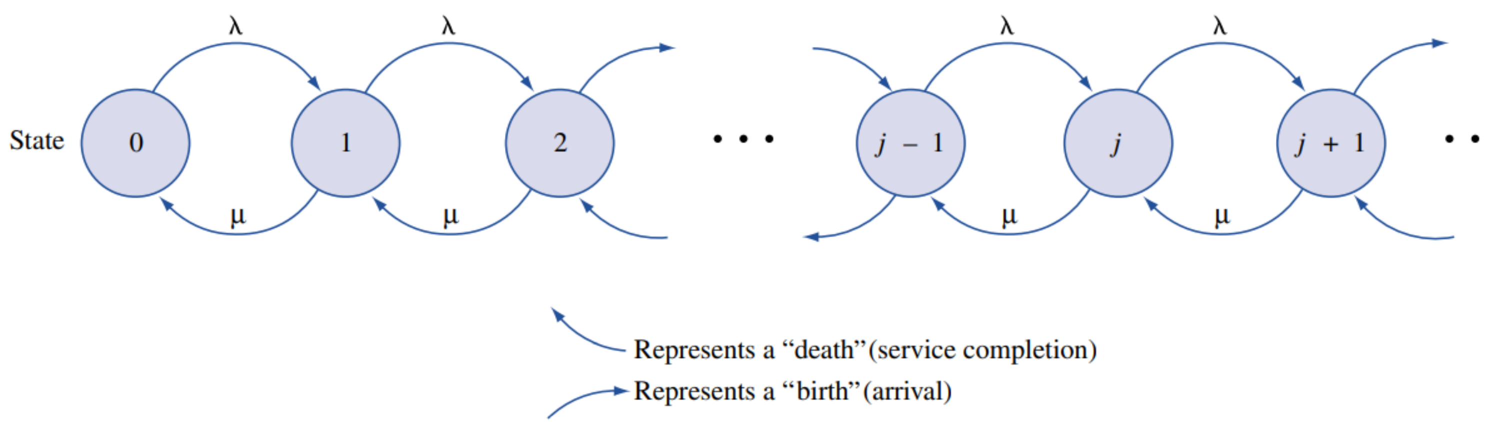
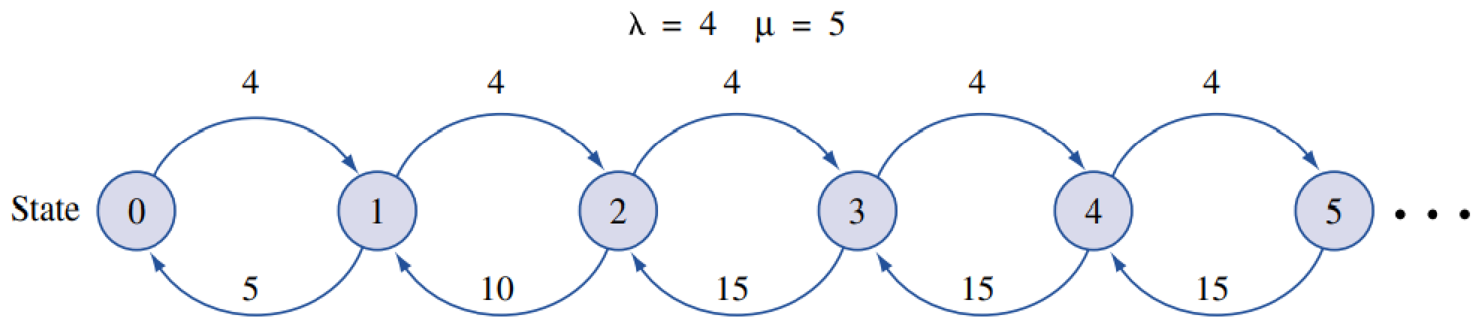


FIGURE 9
Rate Diagram for $M/M/1$/FCFS/∞/∞ Queuing System

M/M/3/FCFS/∞/∞ queuing system

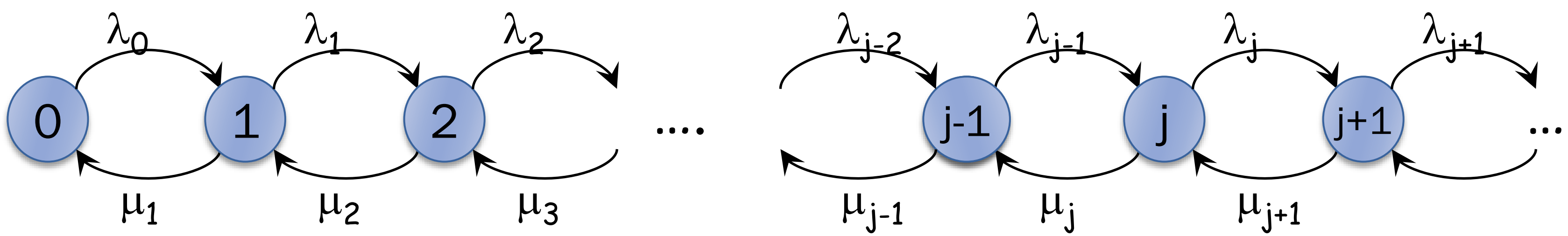
- Interarrival times are exponential with $\lambda = 4$ and service times are exponential with $\mu=5$.

FIGURE 10
Rate Diagram for *M/M/3/FCFS/∞/∞* Queuing System



- If either interarrival times or service times are nonexponential, then the birth–death process model is not appropriate.
- Suppose, for example, that service times are not exponential and we are considering an M/G/1/FCFS/∞/∞ queuing system.
- Since the service times for an M/G/1/FCFS/∞/∞ system may be nonexponential, the probability that a death (service completion) occurs between t and t+ Δt will depend on the time since the last service completion.
- This violates law 2, so we cannot model an M/G/1/FCFS/∞/∞ system as a birth–death process.

Birth-Death Processes



Computations of Probability That State at Time $t + \Delta t$ Is j

State at Time t	State at Time $t + \Delta t$	Probability of This Sequence of Events
$j - 1$	j	$P_{i,j-1}(t)\ (\lambda_{j-1}\Delta t + o(\Delta t)) = \text{(I)}$
$j + 1$	j	$P_{i,j+1}(t)\ (\mu_{j+1}\Delta t + o(\Delta t)) = \text{(II)}$
j	j	$P_{i,j}(t)\ (1 - \mu_j\ \Delta t - \lambda_j \Delta t - 2o(\Delta t)) = \text{(III)}$
Any other state	j	$o(\Delta t) = \text{(IV)}$

Steady-State Probabilities for Birth-Death Processes

- We now show how the π_j's may be determined for an arbitrary birth-death process.
- The key role is to relate (for small Δt) $P_{ij}(t+\Delta t)$ to $P_{ij}(t)$.

$$\pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} = \pi_j(\lambda_j + \mu_j)(j = 1,2,\ldots)$$

$$\pi_1\mu_1 = \pi_0\lambda_0$$

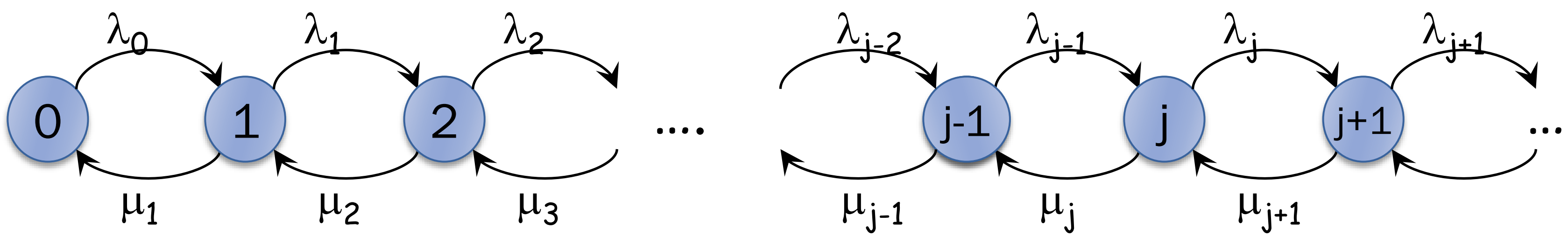
- The above equations are often called the **flow balance equations,** or **conservation of flow equations**, for a birth-death process.

$\pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1}$ = Expected no. of entrances into state j

$\pi_j(\lambda_j + \mu_j)$ = Expected no. of departures from state j

Solution of Balance Eq.

- We obtain the flow balance equations for a birth-death process:



$$(j=0)\pi_0\lambda_0=\pi_1\mu_1$$
$$(j=1)(\lambda_1+\mu_1)\pi_1=\lambda_0\pi_0+\mu_2\pi_2$$
$$(j=2)\ \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\ (\lambda_2+\mu_2)\pi_2=\lambda_1\pi_1+\mu_3\pi_3$$
$$\vdots$$
$$(j\text{th equation})(\lambda_j+\mu_j)\pi_j$$
$$=\lambda_{j-1}\pi_{j-1}+\mu_{j+1}\pi_{j+1}$$

Solution of Balance Eq.

$$j = 0 \qquad \pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

$$j = 1 \qquad \pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 + \frac{1}{\mu_2}(\mu_1\pi_1 - \lambda_0\pi_0) \quad = \quad \frac{\lambda_1}{\mu_2}\pi_1 \quad = \quad \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0$$

$$j = 2 \qquad \pi_3 = \frac{\lambda_2}{\mu_3}\pi_2 + \frac{1}{\mu_3}(\mu_2\pi_2 - \lambda_1\pi_1) \quad = \quad \frac{\lambda_2}{\mu_3}\pi_2 \quad = \quad \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0$$

$$\vdots$$

$$j = j - 1 \qquad \pi_j = \frac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{j-1}}{\mu_1\mu_2\mu_3\ldots\mu_j}\pi_0$$

$$let \quad C_j = \frac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{j-1}}{\mu_1\mu_2\mu_3\ldots\mu_j} \qquad for \quad j = 1, 2, 3, \ldots \text{ and } \mathrm{C_j} = 1 \text{ for } \mathrm{j} = 0 \text{ then}$$

$$\pi_j = C_j\pi_0 \quad for \quad j = 0, 1, 2, \ldots$$

The requirement is that

$$\sum_{\mathrm{j}=0}^{\infty} \pi_j = 1 \quad => \left(\sum_{\mathrm{j}=0}^{\infty} C_j\right)\pi_0 = 1$$

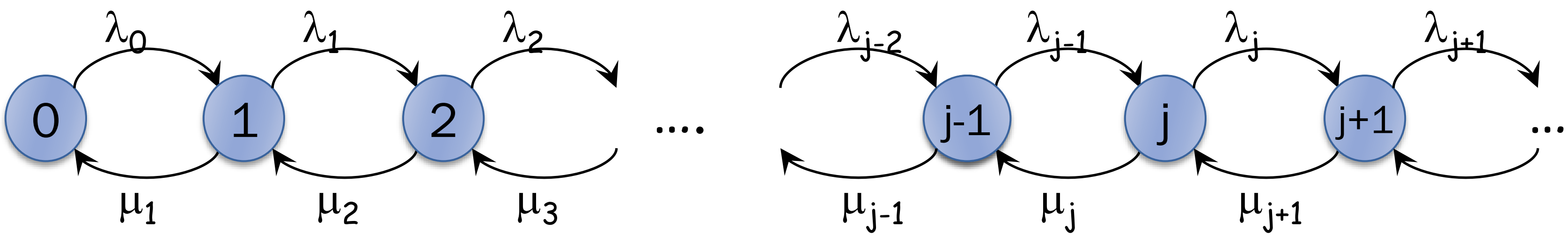
Solution of Balance Eq.

- If $\sum_{j=0}^{j=\infty} c_j$ is finite, we can solve for π_0:

$$\pi_0 = \frac{1}{\sum_{j=0}^{j=\infty} c_j} = \frac{1}{1 + \sum_{j=1}^{j=\infty} c_j}$$

- It can be shown that if $\sum_{j=1}^{j=\infty} c_j$ is infinite, then no steady-state distribution exists.
- The most common reason for a steady-state failing to exist is that the arrival rate is at least as large as the maximum rate at which customers can be served.

Birth Death Process



$$\pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} = \pi_j(\lambda_j + \mu_j)(j = 1,2,\ldots)$$

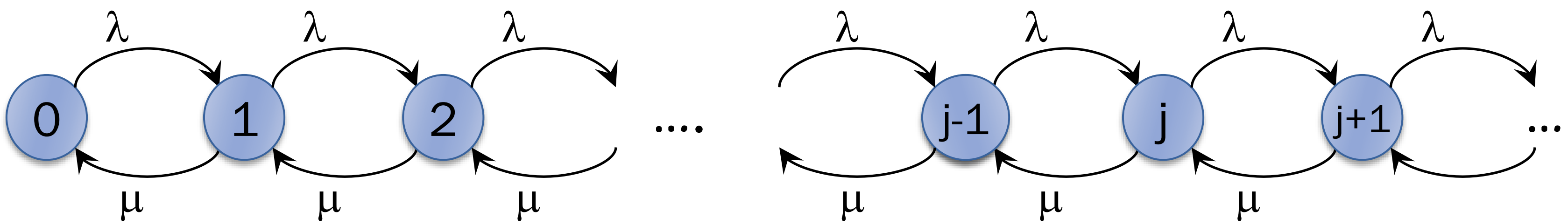
$$\pi_1\mu_1 = \pi_0\lambda_0$$

$$\pi_j = C_j\pi_0 \quad for \quad j \geq 0$$

$$C_j = \begin{cases} 1 & for \quad j = 0 \\ \dfrac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{j-1}}{\mu_1\mu_2\mu_3\ldots\mu_j} & for \quad j > 0 \end{cases}$$

$$\pi_0 = \left(\sum_{j=0}^{j=\infty} C_j\right)^{-1}$$

$M/M/1/GD/\infty/\infty$ Queuing System



Here, $\lambda_j = \lambda \qquad j = 0, 1, 2, \ldots$

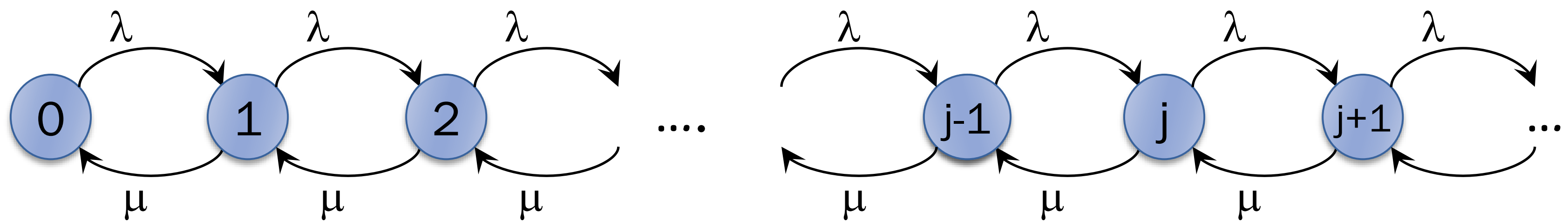
$\mu_0 = 0$

$\mu_j = \mu \qquad j = 1, 2, \ldots$

Let, the **traffic intensity** be $\rho = \lambda / \mu$ and $\mathbf{0 \leq \rho < 1}$

$$C_j = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_j} = \frac{\lambda^j}{\mu^j} = \rho^j \quad for \;\; j \geq 0$$

$M/M/1/GD/\infty/\infty$ Queuing System



$$\pi_0 = \left(\sum_{j=0}^{j=\infty} C_j\right)^{-1} = \left(\sum_{j=0}^{j=\infty} \rho^j\right)^{-1} \quad = \left(1+\rho \quad +\rho^2+\ldots+\rho^j+\ldots\right)^{-1}$$

$$= \left(\frac{1}{1-\rho}\right)^{-1} = 1-\rho$$

$$\pi_j = \rho^j(1-\rho)$$

If $\rho \geq 1$, the infinite sum "blows up", thus, no steady-state distribution exists.

So, rest of this section, we assume $\rho < 1$,

$M/M/1/GD/\infty/\infty$ Queuing System

- The steady state parameters we are interested in
  - L = average number of customers in the queuing system
  - L_q = expected no. of customers waiting in the queue
  - L_S = expected no. of customers in service

    = average number of busy server = C'
  - $L = L_q + L_S$

$M/M/1/GD/\infty/\infty$ Queuing System

- The steady state parameters we are interested in
  - W = average time a customer spends in the system
    = average response time
  - W_q = average time a customer spends in line
  - W_s = average time a customer spends in service

Little's Queuing Formula $L = \lambda_{eff} W$

- For any queuing system in which a steady-state distribution exists, the following relations hold:

$$L = \lambda W$$
$$L_q = \lambda W_q$$
$$L_s = \lambda W_s$$

- Utilization = C′/C

$M/M/1/GD/\infty/\infty$ Queuing System

- **Derivation of L**
- We know that, $E(X) = \sum[x\, f(x)]$ and $P(X=x) = f(x)$
- X = no. of customer in the system during steady state
- X can have values j = 0, 1, 2, … with probability π_j

$$L = E(X) = \sum_{j=0}^{j=\infty} j\pi_j = \sum_{j=0}^{j=\infty} j\rho^j (1-\rho)$$

$$= (1-\rho)\sum_{j=0}^{j=\infty} j\rho^j = (1-\rho)\rho \sum_{j=0}^{j=\infty} j\rho^{j-1}$$

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

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

$M/M/1/GD/\infty/\infty$ Queuing System

Derivation of L

- Another way

$$L = \sum_{j=0}^{j=\infty} j\pi_j = \sum_{j=0}^{j=\infty} j\rho^j (1-\rho)$$

$$= (1-\rho)\sum_{j=0}^{j=\infty} j\rho^j = (1-\rho)S'$$

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

$$S' = \sum_{j=0}^{j=\infty} j\rho^j = \rho + 2\rho^2 + 3\rho^3 + \cdots$$

$$\rho S' = \rho^2 + 2\rho^3 + 3\rho^4 + \cdots$$

$$S' - \rho S' = \rho + \rho^2 + \cdots = \frac{\rho}{1-\rho}$$

$$S' = \frac{\rho}{(1-\rho)^2}$$

$M/M/1/GD/\infty/\infty$ Queuing System

- **Derivation of L_q**
  - Let, the system in state j=0,1,2,… with probability π_j
    - i.e. there are j customer in the system now
  - Since only 1 server, there are j-1 customer is waiting in queue

$$L_q = \sum_{j=1}^{j=\infty}(j-1)\pi_j = \sum_{j=1}^{j=\infty} j\pi_j - \sum_{j=1}^{j=\infty}\pi_j = \sum_{j=0}^{j=\infty} j\pi_j - (1-\pi_0)$$
$$= L - \rho$$

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

$M/M/1/GD/\infty/\infty$ Queuing System

Derivation of L_s

$$L_s = 0\pi_0 + 1(\pi_1 + \pi_2 + \cdots) = 1 - \pi_0 = 1 - (1 - \rho) = \rho$$

$$W_s = \frac{L_s}{\lambda} = \frac{\rho}{\lambda} = \frac{1}{\mu}$$

If we already know L and L_q

$$L_s = L - L_q = \frac{\rho}{1-\rho} - \frac{\rho^2}{1-\rho} = \frac{\rho(1-\rho)}{1-\rho} = \rho$$

If we already know L and L_s

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

An average of 10 cars per hour arrive at a single-server drive-in teller. Assume that the average service time for each customer is 4 minutes, and both interarrival times and service times are exponential. Answer the following questions:

1 What is the probability that the teller is idle?

2 What is the average number of cars waiting in line for the teller? (A car that is being served is not considered to be waiting in line.)

3 What is the average amount of time a drive-in customer spends in the bank parking lot (including time in service)?

4 On the average, how many customers per hour will be served by the teller?

By assumption, we are dealing with an $M/M/1/GD/\infty/\infty$ queuing system for which $\lambda =$ 10 cars per hour and $\mu = 15$ cars per hour. Thus, $\rho = \frac{10}{15} = \frac{2}{3}$.

1 From (24), $\pi_0 = 1 - \rho = 1 - \frac{2}{3} = \frac{1}{3}$. Thus, the teller will be idle an average of one-third of the time.

2 We seek L_q. From (27),

$$L_q = \frac{\rho^2}{1-\rho} = \frac{(\frac{2}{3})^2}{1-\frac{2}{3}} = \frac{4}{3} \quad \text{customers}$$

3 We seek W. From (28), $W = \frac{L}{\lambda}$. Then from (26).

$$L = \frac{\rho}{1-\rho} = \frac{\frac{2}{3}}{} = 2 \quad \text{customers}$$

Thus, $W = \frac{2}{10} = \frac{1}{5}$ hour $= 12$ minutes (W will have the same units as λ).

4 If the teller were always busy, he would serve an average of $\mu = 15$ customers per hour. From part (1), we know that the teller is only busy two-thirds of the time. Thus, during each hour, the teller will serve an average of $(\frac{2}{3})(15) = 10$ customers. This must be the case, because in the steady state, 10 customers are arriving each hour, so each hour, 10 customers must leave the system.

$M/M/1/GD/\infty/\infty$ Queuing System

- **Example 4 (Winston)**
- Suppose that all car owners fill up when their tanks are exactly half full.
- At the present time, an average of 7.5 customers per hour arrive at a single-pump gas station.
- It takes an average of 4 minutes to service a car.
- Assume that interarrival and service times are both exponential.

1. What is the probability that the teller is idle?

2. For the present situation, compute L , Lq and W, Wq.

$M/M/1/GD/\infty/\infty$ Queuing System

3. Suppose that a gas shortage occurs and panic buying takes place.
   - To model the phenomenon, suppose that all car owners now purchase gas when their tank are exactly three quarters full.
   - Since each car owner is now putting less gas into the tank during each visit to the station, we assume that the average service time has been reduced to 3 1/3 =(10/3)minutes.
   - How has panic buying affected L and W?

Solutions

- We have an $M/M/1/GD/\infty/\infty$ system with

  λ = 7.5 cars per hour μ = 60/4=15 cars per hour.

  $\rho = 7.5/15 = 0.5$

1. Idle probability $\pi_0 = 1 - \rho = 0.5$

2. $L = \rho/(1 - \rho)$= .50/(1 – 0.5) = 1,

   $W = L/\lambda = 1/7.5$ = 0.13 hour = 8 mins.

   $L_q = \rho^2/(1- \rho)= 0.25/(1 - 0.5) = 0.5$

   $W_q = L_q/\lambda = 0.5/7.5 = 1/15\ hr$ = 4 mins.

- Hence, in this situation, everything is under control, and long lines appear to be unlikely.

Solutions

3. We now have an *M/M/1/GD/∞/∞* system with

   λ = 2*7.5 =15 cars per hour(This follows because each car owner will fill up twice as often.)

   μ = 60*3/10=18 cars per hour.

   ρ = *15/18* =*5/6*

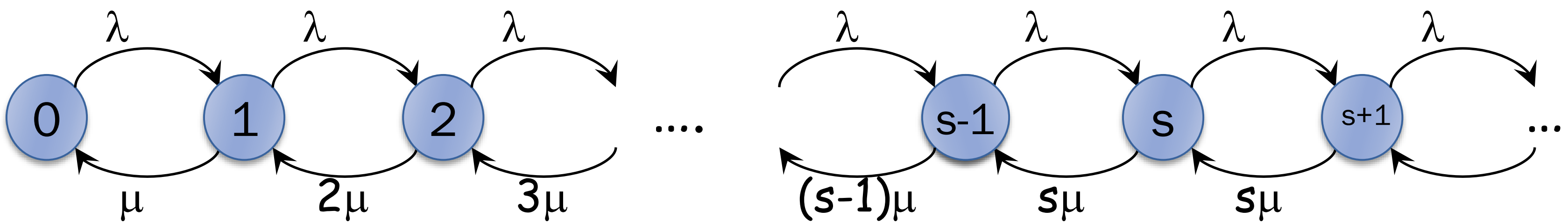
$$L = \frac{\frac{5}{6}}{1-\frac{5}{6}} = 5 \text{ cars and } \quad W = \frac{L}{\lambda} \rightleftarrows = \frac{5}{15} = \frac{1}{3} \text{hours}$$

$$= 20 \text{ minutes}$$

$$L_q = \frac{\left(\frac{5}{6}\right)^2}{1-\frac{5}{6}} = \frac{25}{6} \text{ cars and } \quad W_q = \frac{L_q}{\lambda} \rightleftarrows = \frac{25}{6*15} = \frac{5}{18} \text{ hours} = 16.67 \text{ minutes}$$

- Thus, panic buying has cause long lines.

$M/M/s/GD/\infty/\infty$ Queuing System(Winston: 1103)



Here, $\lambda_j = \lambda \qquad j = 0, 1, 2, \ldots$

$\mu_j = j\mu \qquad j = 0, 1, 2, \ldots, s$

$\mu_j = s\mu \qquad j = s+1, s+2, \ldots$

Let, the **traffic intensity** be $\rho = \lambda/(s\mu)$ and $0 \le \rho < 1$

If $j <= s$ customers are present, then all j customers are in service; if $j > s$ customers are present, then all s servers are occupied, and $j-s$ customers are waiting in line.

$$C_j = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_j}$$

$$= \frac{\lambda^j}{\mu(2\mu)(3\mu)\ldots(j\mu)} = \frac{(s\rho)^j}{j!} \qquad for\ j \le s$$

$M/M/s/GD/\infty/\infty$ Queuing System

$$C_j = \frac{\lambda^j}{\mu(2\mu)(3\mu)\ldots(s\mu)(su)\ldots(su)}$$

$$= \frac{(s\rho)^s}{s!}\rho^{j-s} = \frac{(s\rho)^j}{s!\,s^{j-s}} \qquad for\ j > s$$

$$\pi_0 = \left(\sum_{j=0}^{j=\infty} C_j\right)^{-1} = \left(\sum_{j=0}^{S-1}\frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!}\sum_{j=s}^{\infty}\rho^{j-s}\right)^{-1}$$

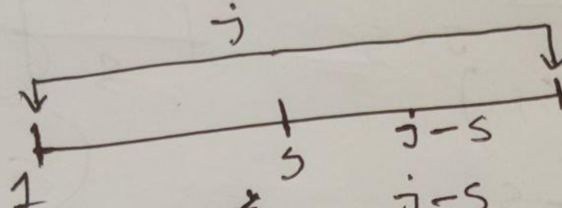
$$= \left(\sum_{j=0}^{S-1}\frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!\,(1-\rho)}\right)^{-1}$$

$$\pi_j = \frac{(s\rho)^j\pi_0}{j!} \qquad (j = 1, 2, \ldots, s)$$

$$\pi_j = \frac{(s\rho)^j\pi_0}{s!s^{j-s}} \qquad (j = s, s+1, s+2, \ldots)$$

If $\rho \geq 1$, the infinite sum "blows up", thus, no steady-state exists.

$$C_j = \frac{\lambda^j}{\mu\cdot(2\mu)(3\mu)\cdots(s\mu)(s\mu)\cdots(s\mu)}$$



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

$$C_j = \frac{\lambda^s \cdot \lambda^{j-s}}{s!\,(\mu^s)\cdot(s\mu)^{j-s}}$$

$$= \frac{\rho^s}{s!} \times \left(\frac{\lambda}{\mu\rho}\right)^s \left(\frac{\lambda}{s\mu}\right)^{j-s}$$

$$= \frac{\rho^s}{s!}\, \rho^s\, s^s \cdot \rho^{j-s}$$

$$= \frac{(s\rho)^s}{s!}\, \rho^{j-s}$$

M/M/s/GD/∞/∞ Queuing System

$$\pi_j = \frac{(s\rho)^j \pi_0}{s!s^{j-s}}$$

$$L_q = \sum_{j=s}^{\alpha}(j-s)\pi_j = \sum_{n=0}^{\alpha} n\pi_{s+n} \quad [taking, \qquad j-s=n]$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \sum_{n=0}^{\alpha} n\rho^{s+n-s} = \frac{(s\rho)^s}{s!}\pi_0 S'$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \frac{\rho}{(1-\rho)^2}$$

$$P(j \geq s) = \sum_{j=s}^{\alpha} \pi_j$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \sum_{j=s}^{\alpha} \frac{(s\rho)^{j-s}}{s^{j-s}}$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \sum_{j=s}^{\alpha} \rho^{j-s}$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \sum_{n=0}^{\alpha} \rho^{n}$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \frac{1}{1-\rho}$$

steady-state probability that all servers are busy

$$P_d = P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

$$L_q = \frac{P(j \geq s)\rho}{1-\rho} \qquad W_q = \frac{L_q}{\lambda} = \frac{P(j \geq s)}{s\mu - \lambda}$$

$$\pi_{s+n} = \frac{(se)^{s+n}}{s!\; s^{s+n-s}}$$

$$= \frac{(se)^{s}\,(se)^{s+n-s}}{s!\; s^{s+n-s}}$$

$$= \frac{(se)^{s}\, e^{s+n-s}}{s!}$$

$M/M/s/GD/\infty/\infty$ Queuing System (Taha version)

- $\rho = \lambda / \mu$, $c = s$, $\pi_0 = p_0$

Thus,

$$p_n = \begin{cases} \dfrac{\lambda^n}{\mu(2\mu)(3\mu)\ldots(n\mu)} p_0 = \dfrac{\lambda^n}{n!\,\mu^n} p_0 = \dfrac{\rho^n}{n!} p_0, & n < c \\ \dfrac{\lambda^n}{\left(\prod_{i=1}^{c} i\mu\right)(c\mu)^{n-c}} p_0 = \dfrac{\lambda^n}{c!\,c^{n-c}\mu^n} p_0 = \dfrac{\rho^n}{c!\,c^{n-c}} p_0, & n \geq c \end{cases}$$

Letting $\rho = \frac{\lambda}{\mu}$, and assuming $\frac{\rho}{c} < 1$, the value of p_0 is determined from $\sum_{n=0}^{\infty} p_n = 1$, which gives,

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\rho}{c}\right)^{n-c} \right\}^{-1}$$

$$= \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \left(\frac{1}{1 - \frac{\rho}{c}} \right) \right\}^{-1}, \frac{\rho}{c} < 1$$

$$L_q = \frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2} p_0$$

- *These are actually same as the previous ones*

$M/M/s/GD/\infty/\infty$ Queuing System

The time spent in the waiting queue: $W_q = \dfrac{L_q}{\lambda} = \dfrac{P_d}{s\mu - \lambda}$

Since $W_s = \dfrac{1}{\mu}$

$$\mathrm{L}_q = \frac{\rho}{1-\rho} P_d$$

$L_s = \dfrac{\lambda}{\mu}$ = sρ = average number of busy servers, s'

$M/M/s/GD/\infty/\infty$ Queuing System

The number of customers in the queueing system:

$$L = L_q + \frac{\lambda}{\mu}$$

$$= \frac{(s\rho)^s}{s!}\pi_0 \frac{\rho}{(1-\rho)^2} + s\rho$$

$$= \frac{\rho}{1-\rho} P_d + s\rho$$

$$P_d = P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

The time spends in the queueing system:

$$W = \frac{L}{\lambda}$$

$$= \frac{L_q}{\lambda} + \frac{1}{\mu}$$

$$= W_q + \frac{1}{\mu}$$

$$= \frac{P(j \geq s)}{s\mu - \lambda} + \frac{1}{\mu}$$

$M/M/s/GD/\infty/\infty$ Queuing System

A community is served by two cab companies. Each company owns two cabs and both share the market equally, as evidenced by the fact that calls arrive at each company's dispatching office at the rate of eight per hour. The average time per ride is 12 minutes. Calls arrive according to a Poisson distribution, and the ride time is exponential. The two companies recently were bought by an investor who is interested in consolidating them into a single dispatching office to provide better service to customers. Analyze the new owner's proposal.

As M/M/2 system

$\lambda = 8$ $\mu = 60/12 = 5$ $s = 2$ $\rho = \lambda/s\mu = 0.8$

$$\pi_0 = \left(\sum_{j=0}^{2-1} \frac{(2*0.8)^j}{j!} + \frac{(2*0.8)^2}{2!\,(1-0.8)} \right)^{-1}$$

$$= \left(1 + 1.6 + \frac{(1.6)^2}{0.4} \right)^{-1} = 1/9 = 0.111$$

$$W_q = Lq/\lambda$$

$$Lq = \frac{(s\rho)^s}{s!} \pi_0 \frac{\rho}{(1-\rho)^2}$$

$$\pi_0 = \left(\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!\,(1-\rho)} \right)^{-1}$$

$M/M/s/GD/\infty/\infty$ Queuing System

As M/M/2 system

$\lambda = 8 \quad \mu = 60/12 = 5 \quad s = 2 \quad \rho = \lambda/s\mu = 0.8$

$$L_q = \frac{(2 * 0.8)^2}{2!} * \frac{1}{9} * \frac{0.8}{(1 - 0.8)^2}$$

$$= \frac{1.28}{9} * 20 = 2.844$$

$$W_q = Lq/8 = 0.356hr = 21.33\text{min}$$

$M/M/s/GD/\infty/\infty$ Queuing System

As M/M/4 system

λ = 2*8=16 μ = 60/12 = 5 s = 2*2 $\rho = \lambda/s\mu$= 0.8

$$\pi_0 = \left(\sum_{j=0}^{4-1} \frac{(4*0.8)^j}{j!} + \frac{(4*0.8)^4}{4!\,(1-0.8)} \right)^{-1}$$

$$= \left(1 + 3.2 + \frac{3.2^2}{2} + \frac{3.2^3}{6} + \frac{(3.2)^4}{4.8} \right)^{-1}$$

$$= (1 + 3.2 + 5.12 + 5.46133 + 21.8453)^{-1}$$

$$= 1/36.62666 = 0.0273$$

$M/M/s/GD/\infty/\infty$ Queuing System

As M/M/4 system

λ = 2*8=16 μ = 60/12 = 5 s = 2*2 $\rho = \lambda/s\mu = 0.8$

$$L_q = \frac{(4*0.8)^4}{4!} * 0.027 * \frac{0.8}{(1-0.8)^2}$$
$$= 4.369 * 0.0273 * 20 = 2.386$$
$$W_q = Lq/16 = 0.149hr = 8.95\text{min}$$

Waiting Time Distribution of M/M/s/FCFS/∞/ ∞

- Average waiting time Wq is independent of queue discipline
- But is probability distribution is not

- For FCFS queue with s=1
  - System waiting time distribution : exponential with parameter μ - λ

$$w(t) = \mu(1-\rho)e^{-\mu(1-\rho)t}$$
$$= (\mu-\lambda)e^{-(\mu-\lambda)t} \quad , \qquad t > 0$$

  - Queue waiting time distribution

$$w_q(t) = 1-\rho \, , \qquad t = 0$$

$$w_q(t) = \mu\rho(1-\rho)e^{-\mu(1-\rho)t}$$
$$= \rho(\mu-\lambda)e^{-(\mu-\lambda)t} \quad , \qquad t > 0$$

Waiting Time Distribution of M/M/s/FCFS/∞/ ∞

- Average waiting time W_q is independent of queue discipline
- But is probability distribution is not

- For FCFS queue with s=1

$$P(t > W) = 1 - \int_0^W w(t)dt$$

$$= 1 - \int_0^W (\mu - \lambda)e^{-(\mu-\lambda)t}dt$$

$$= 1 - \left(1 - e^{-(\mu-\lambda)W}\right)$$
$$= e^{-1} = 0.368$$

For M/M/1

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

About 37% customer wait longer than average waiting time

Waiting Time Distribution of M/M/s/FCFS/∞/ ∞

- Average waiting time Wq is independent of queue discipline
- But is probability distribution is not

- For FCFS queue with s=1
- Gas buying : probability of waiting more than 4 min in queue, where λ = 7.5 cars per hour   μ = 60/4=15

$$P(t > W_q = 4) = 1 - \int_0^4 w_q(t)dt$$

$$= 1 - \rho \int_0^4 (\mu - \lambda)e^{-(\mu-\lambda)t}dt$$

$$= 1 - 0.5\left(1 - e^{-(7.5)4}\right)$$
$$= 0.5$$

Waiting Time Distribution of M/M/s/FCFS/∞/ ∞

- Average waiting time Wq is independent of queue discipline
- But is probability distribution is not

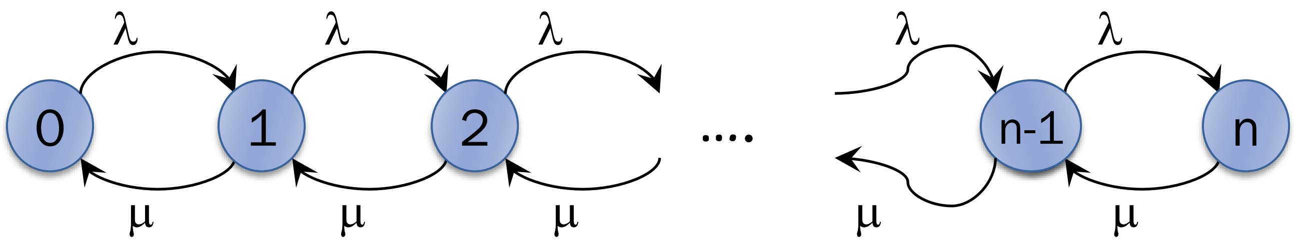
- For FCFS queue with s

$$P(\mathbf{W} > t) = e^{-\mu t}\left\{1 + P(j \geq s)\,\frac{1 - \exp\left[-\mu t(s - 1 - s\rho)\right]}{s - 1 - s\rho}\right\}^{\dagger}$$

$$P(\mathbf{W}_q > t) = P(j \geq s)\exp\left[-s\mu(1 - \rho)t\right]$$

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1 - \rho)}$$

$M/M/1/GD/ \ n \ /\infty$ Queuing System



Here, $\lambda_j = \lambda \qquad j = 0, 1, 2, \ldots n-1$

$\lambda_n = 0$

$\mu_0 = 0$

$\mu_j = \mu \qquad j = 1, 2, \ldots, n$

Let, the **traffic intensity** be $\rho = \lambda / \mu$ and $\rho \geq 0$

$$C_j = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_j} = \rho^j \qquad for\ j = 0,1,\ldots,n$$

The M/M/1/GD/n/∞ system is identical to the M/M/1/GD/∞/∞ system except for the fact that when n customers are present, all arrivals are turned away and are forever lost to the system.

$M/M/1/GD/n/\infty$ Queuing System

$$\pi_0 = \left(\sum_{j=0}^{j=n} C_j\right)^{-1} = \left(\sum_{j=0}^{j=n} \rho^j\right)^{-1} \qquad = (1 + \rho + \rho^2 + \ldots + \rho^n)^{-1}$$

$$= \left(\frac{\rho^{n+1} - 1}{\rho - 1}\right)^{-1} = \frac{1 - \rho}{1 - \rho^{n+1}} \qquad for\ \rho \neq\ 1$$

$$\pi_0 = \frac{1}{n + 1} \qquad for\ \rho =\ 1$$

$$\pi_j = \rho^j \pi_0 \qquad for\ j = 1, \ldots, n \text{ and } for\ any\ \rho$$

Geometric series

For any number *r*,

$1 + r + r^2 + ... + r^n$ = SUM(k=0...n) $r^k = (1 - r^{n+1}) /(1 - r)$, and
$r + r^2 + r^3 + ... + r^n$ = SUM(k=1...n) $r^k = (r - r^{n+1})/(1 - r)$.

Thus, if $|r| < 1$ then taking the limit *n* -> *infinity*,

$1 + r + r^2 + r^3 + . . .$ = SUM(k=0... infinity) $r^k = 1/(1-r)$, and
$r + r^2 + r^3 + . . .$ = SUM(k=1...infinity) $r^k = r/(1-r)$.

$M/M/1/GD/n/\infty$ Queuing System

$$L = \sum_{j=0}^{j=n} j\pi_j = \frac{1-\rho}{1-\rho^{n+1}} \sum_{j=0}^{j=n} j\rho^j$$

$$= \frac{1-\rho}{1-\rho^{n+1}} S$$

$$S = \sum_{j=0}^{j=n} j\rho^j = \rho + 2\rho^2 + \ldots + n\rho^n$$

$$\rho S = \rho^2 + 2\rho^3 + \ldots + n\rho^{n+1}$$

$$S - \rho S = \rho + \rho^2 + \rho^3 + \ldots + \rho^n - n\rho^{n+1}$$

$$= \frac{1-\rho^{n+1}}{1-\rho} - (1 + n\rho^{n+1})$$

$$= \frac{1-\rho^{n+1} - 1 + \rho - n\rho^{n+1} + n\rho^{n+2}}{1-\rho}$$

$$= \frac{\rho[1-(n+1)\rho^n + n\rho^{n+1}]}{1-\rho}$$

$$= \frac{\rho[1-(n+1)\rho^n + n\rho^{n+1}]}{(1-\rho^{n+1})(1-\rho)}$$

$$for\ \rho \neq\ 1$$

$$= \frac{n}{2} \quad (verify)$$

$$for\ \rho =\ 1$$

$M/M/1/GD/n/\infty$ Queuing System

Derivation of L_s

$$L_s = 0\pi_0 + 1(\pi_1 + \pi_2 + \cdots) = 1 - \pi_0$$

$$L_q = L - L_s$$

$$\lambda_{lost} = \lambda \pi_n$$

$$\lambda_{eff} = \lambda - \lambda_{lost} = \lambda(1 - \pi_n)$$

$$W = \frac{L}{\lambda_{eff}} \qquad W_q = \frac{L_q}{\lambda_{eff}} \qquad W_s = \frac{L_s}{\lambda_{eff}}$$

$M/M/1/GD/n/\infty$ Queuing System – Example

- A car wash facility with single bay for wash and 4 parking space, arrival rate = 4 / hr, service time = 10 min / car, all are exponential [Taha Problem 15.6D]
  - Determine the following

(a) Probability that an arriving car will go into the wash bay immediately on arrival.
(b) Expected waiting time until a service starts.
(c) Expected number of empty parking spaces.
(d) Probability that all parking spaces are occupied.

  - Here

$$\lambda = 4 \quad \mu = 6 \quad \rho = 2/3 = 0.667$$
$$n = 4 + 1 = 5$$

  - (a) $$\pi_0 = \frac{1-\rho}{1-\rho^{n+1}} = \frac{{}^{1}/_{3}}{1-\frac{64}{3^6}} = \frac{{}^{1}/_{3}}{\frac{665}{3^6}} = \frac{243}{665} = 0.3654135$$

M/M/1/GD/n/∞ Queuing System - Example

(b) Expected waiting time until a service starts.

(c) Expected number of empty parking spaces.

(d) Probability that all parking spaces are occupied.

- (b)

$$L = \frac{\rho[1-(n+1)\rho^n + n\rho^{n+1}]}{(1-\rho^{n+1})(1-\rho)} = \frac{0.432556}{}$$

$$L_s = 1 - \pi_0 = 0.6345865 \qquad L_q = L - L_s = 0.7879695$$

$$\pi_n = \rho^n \pi_0 \qquad => \pi_5 = (2/3)^5 * 0.3654135 = 0.04812$$

(d)

$$\lambda_{eff} = \lambda(1-\pi_5) = 3.80752 \qquad W_q = \frac{L_q}{\lambda_{eff}} = 0.207hr = 12.42\text{min}$$

(c) Expected empty parking spaces = 4 – Lq = 3.212

M/G/1/GD/∞/ ∞

- interarrival times are exponential, let λ be the arrival rate
- but the service time distribution (**S**) is not
  - Let, $1/\mu = E(\mathbf{S})$ and $\sigma^2 = \text{var } \mathbf{S}$
- It is *not* a birth–death process
  - because the probability that a service completion occurs between t and $t+\Delta t$ when the state of the system at time t is j depends on the length of time since the last service completion.
- Determination of the steady-state probabilities for *M/G/1/GD/∞/∞* queuing system is a difficult matter.
- Fortunately, however, utilizing the results of Pollaczek and Khinchin, we may determine L_q, L, L_s, W_q, W, W_s.

M/G/1/GD/∞/ ∞

- Pollaczek and Khinchin showed that for the *M/G/1/GD/∞/∞* queuing system,

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}$$

$$L = L_q + \rho$$

$$W_q = \frac{L_q}{\lambda}$$

$$W = W_q + \frac{1}{\mu}$$

- It can also be shown that $\pi_0 = 1 - \rho$.
- The result is similar to the one for the *M/M/1/GD/∞/∞* system.

M/G/1/GD/∞/ ∞

- consider an $M/M/1/GD/\infty/\infty$ system with λ=5 customers per hour and μ = 8 customers per hour

$$L = \frac{\lambda}{\mu - \lambda} = \frac{5}{8-5} = \frac{5}{3} \text{ cust}$$

$$L_q = L - \rho = \frac{5}{3} - \frac{5}{8} = \frac{25}{24} \text{ cus}$$

$$W = \frac{L}{\lambda} = \frac{\frac{5}{3}}{5} = \frac{1}{3} \text{ hour}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{24}}{5} = \frac{5}{24} \text{ hour}$$

Considering it as a M/G/1 system

$E(S) = 1/\mu = 1/8$

$\sigma^2 = 1/\mu^2 = 1/64$ [since S is exponential]

$$L_q = \frac{\frac{(5)^2}{64} + \left(\frac{5}{8}\right)^2}{2\left(1 - \frac{5}{8}\right)} = \frac{25}{24} \text{ customers}$$

$$L = L_q + \rho = \frac{25}{24} + \frac{5}{8} = \frac{40}{24} = \frac{5}{3} \text{ cust}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{24}}{5} = \frac{5}{24} \text{ hour}$$

$$W = \frac{L}{\lambda} = \frac{\frac{5}{3}}{5} = \frac{1}{3} \text{ hour}$$

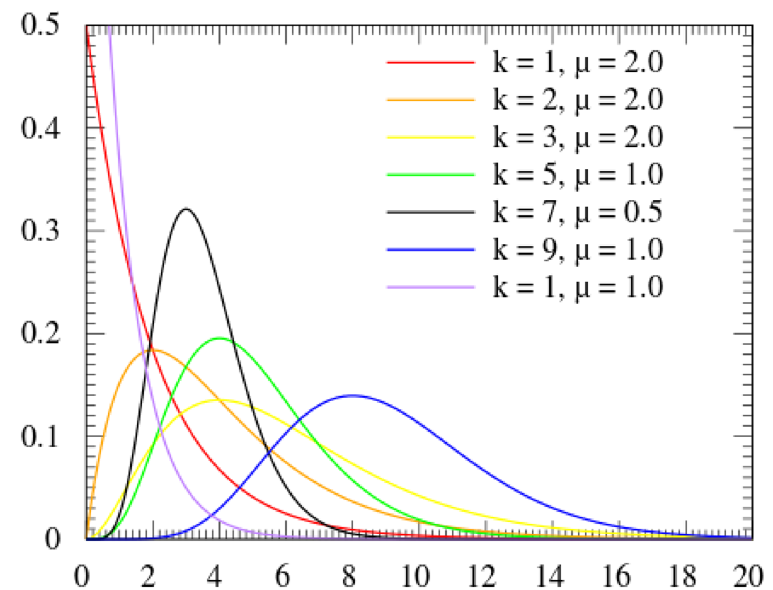
M/G/1/GD/∞/ ∞

- **Erlang Distribution** is a two parameter family of continuous probability distributions with support x ∈ [0 , ∞)
- The two parameters are:
  - a positive integer k the "**shape**", and
  - a positive real number λ , the "**rate**". The "scale", μ , the reciprocal of the rate, is sometimes used instead.
- It is a special case of the Gamma distribution.
- The probability density function is

$$f(x;k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad \text{for } x, \lambda \geq 0.$$

Or

$$f(x;k,\mu) = \frac{x^{k-1} e^{-\frac{x}{\mu}}}{\mu^k (k-1)!} \quad \text{for } x, \mu \geq 0$$



$M/G/1/GD/\infty/\infty$

- ## Erlang Distribution
  - Mean: k / λ.
  - Variance: k / λ^2.

  - Erlang distribution with $k = 1$ simplifies to exponential distribution.

- Consider an $M/G/1/GD/\infty/\infty$ queuing system in which an average of 10 arrivals occur each hour. Suppose that each customer's service time follows an Erlang distribution, with rate parameter 1 customer per minute and shape parameter 4. [Winston page 1098]
  a) Find the expected number of customers waiting in line.
  b) Find the expected time that a customer will spend in the system.
  c) What fraction of the time will the server be idle?

M/G/1/GD/∞/ ∞

Since, Erlang Distribution

$$k = 4 \quad \lambda_s = 1/\min = 60/hr$$

$$mean \quad service \quad time = \frac{k}{\lambda_s} = \frac{4}{60}hr = \frac{1}{15}hr = 4\min$$

$$\sigma^2 = \frac{k}{{\lambda_s}^2} = \frac{4}{60^2} = \frac{1}{30^2}hr^2$$

M/G/1/GD/∞/ ∞

a) Find the expected number of customers waiting in line.

b) Find the expected time that a customer will spend in the system.

c) What fraction of the time will the server be idle?

$$arrival\ rate\ \lambda = 10/hr$$
$$\text{service rate}\ \mu = 60/4 = 15/hr$$
$$\rho = 2/3$$

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}$$
$$= \frac{(10/30)^2 + (^2/_3)^2}{^2/_3}$$
$$= \frac{^5/_9}{^2/_3} = \frac{5}{6}\quad cust$$

$$L = L_q + \rho = 5/6 + 2/3$$
$$= 3/2\quad cust$$
$$W = \frac{L}{\lambda} = \frac{3}{20} hr = 9\text{min}$$

$$\pi_0 = 1 - \rho = 1/3$$

$M/D/1/GD/\infty/\infty$

- consider an $M/D/1/GD/\infty/\infty$ system with λ=5 customers per hour and μ = 8 customers per hour
- $E(\mathbf{S}) = 1/\mu = 1/8$
- $\sigma^2 = 0$

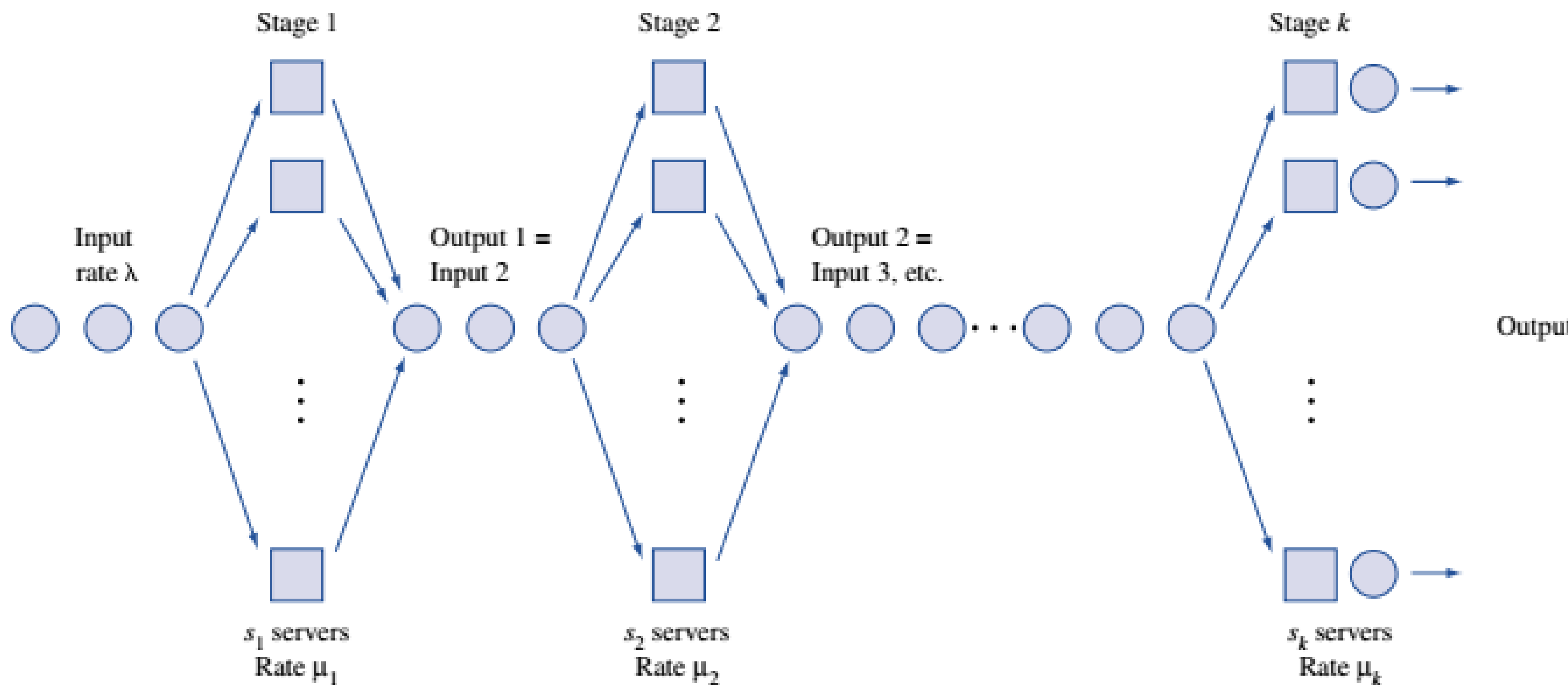
$$L_q = \frac{\left(\frac{5}{8}\right)^2}{2\left(1 - \frac{5}{8}\right)} = \frac{25}{48} \text{ customer}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{48}}{5} = \frac{5}{48} \text{ hour}$$

- a typical customer will spend only half as much time in line as in an $M/M/1/GD/\infty/\infty$ queuing system with identical arrival and service rates.
- even if mean service times are not decreased, a decrease in the variability of service times can substantially reduce queue size and customer waiting time

Exponential Queues in Series Networks

- In the queuing models that we have studied so far, a customer's entire service time is spent with a single server.
- In many situations the customer's service is not complete until the customer has been served by more than one server.
- k-stage series queuing system



Exponential Queues in Series Networks

- **Theorem**
- **If**
  1) interarrival times for a series queuing system are exponential with rate λ,
  2) service times for each stage $\boldsymbol{i}$ server are exponential, and
  3) each stage has an infinite-capacity waiting room,
- **then** interarrival times for arrivals to each stage of the queuing system are exponential with rate λ.

Exponential Queues in Series Networks

- The last two works in a car manufacturing process are installing the engine and putting on the tires. An average of 54 cars per hour arrive requiring these two tasks.
- One worker is available to install the engine and can service an average of 60 cars per hour.
- After the engine is installed, the car goes to the tire station and waits for its tires to be attached. Three workers serve at the tire station. Each works on one car at a time and can put tires on a car in an average of 3 minutes.
- Both interarrival times and service times are exponential.
  - Determine the mean queue length at each work station.
  - Determine the total expected time that a car spends waiting for service

Exponential Queues in Series Networks

- This is a series queuing system with
  - λ=54 cars per hour, s_1 = 1, μ_1 = 60 cars per hour,
  - s_2 = 3, and μ_2 = 20 cars per hour
  - Since $\lambda < \mu_1$ and $\lambda < 3\mu_2$, neither queue will "blow up,"
- For stage 1 (engine), ρ = 54/60 = 0.9

$$L_q \text{ (for engine)} = \left(\frac{\rho^2}{1-\rho}\right) = \left[\frac{(.90)^2}{1-.90}\right] = 8.1 \text{ cars}$$

$$W_q \text{ (for engine)} = \frac{L_q}{\lambda} = \frac{8.1}{54} = 0.15 \text{ hour}$$

Exponential Queues in Series Networks

- For stage 2 (Tires), ρ = 54/(3*20) = 0.9

$$L_q = \frac{(s\rho)^s}{s!}\pi_0 \frac{\rho}{(1-\rho)^2} \qquad L_q = \frac{P(j \geq s)\rho}{1-\rho} \qquad P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

$$\pi_0 = \left(\sum_{j=0}^{S-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!\,(1-\rho)}\right)^{-1}$$
$$= (1 + 2.7 + 3.645 + 32.805)^{-1} = 0.025$$

$$P(j \geq s) = 0.82$$

$$L_q = \frac{0.82 * 0.9}{1-0.9} = 7.4 \quad cars$$

total expected waiting time is

0.15 +0.137 = 0.287 hour

$$W_q = \frac{7.4}{54} = 0.137 hrs$$