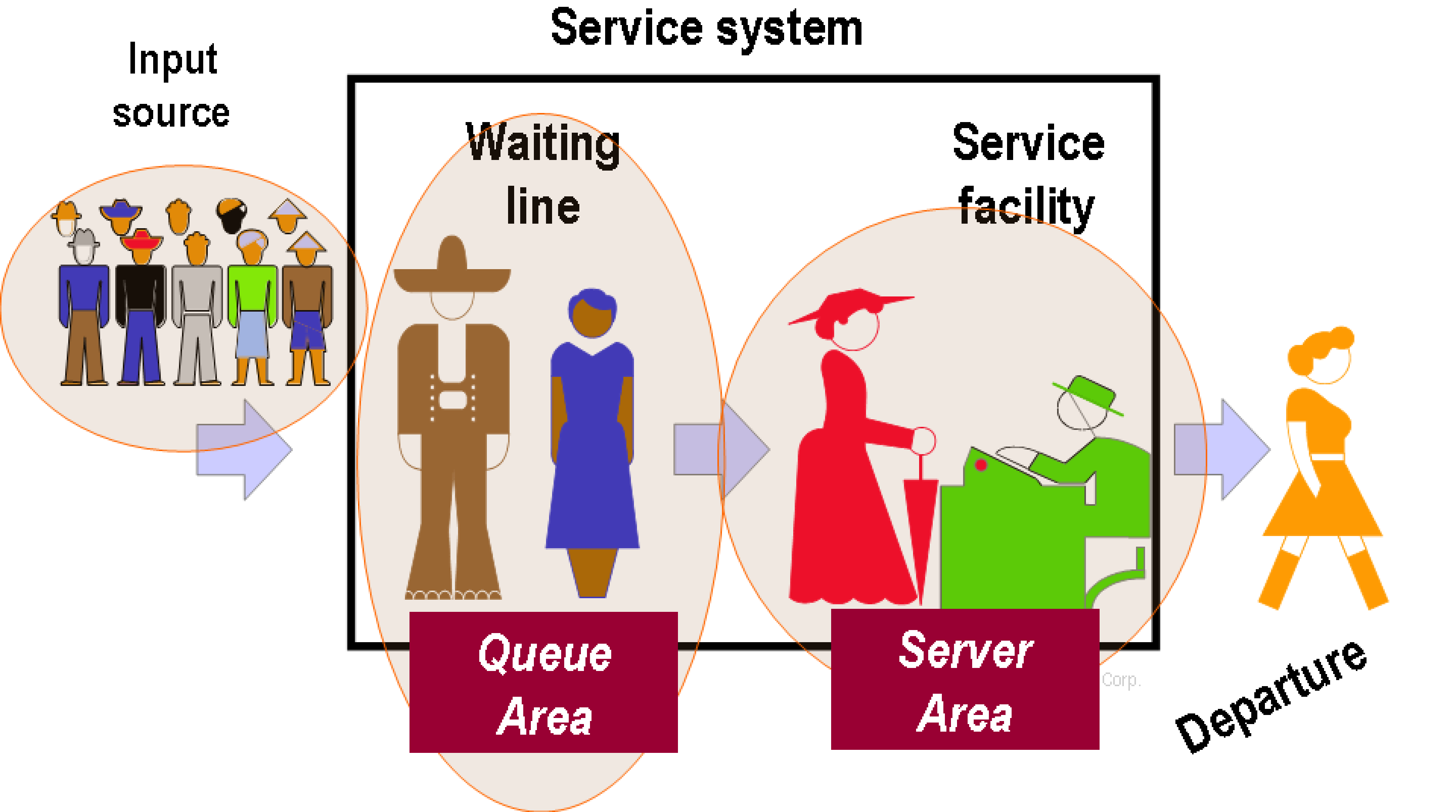
Queuing Theory

Queuing System : Performance Measures

- Average queue time, W_q
- Average queue length, L_q
- Average time in system, W
- Average number in system, L
- Probability of idle service facility, P_0
- System utilization, ρ
- Probability of *k* units in system, $P_{n > k}$

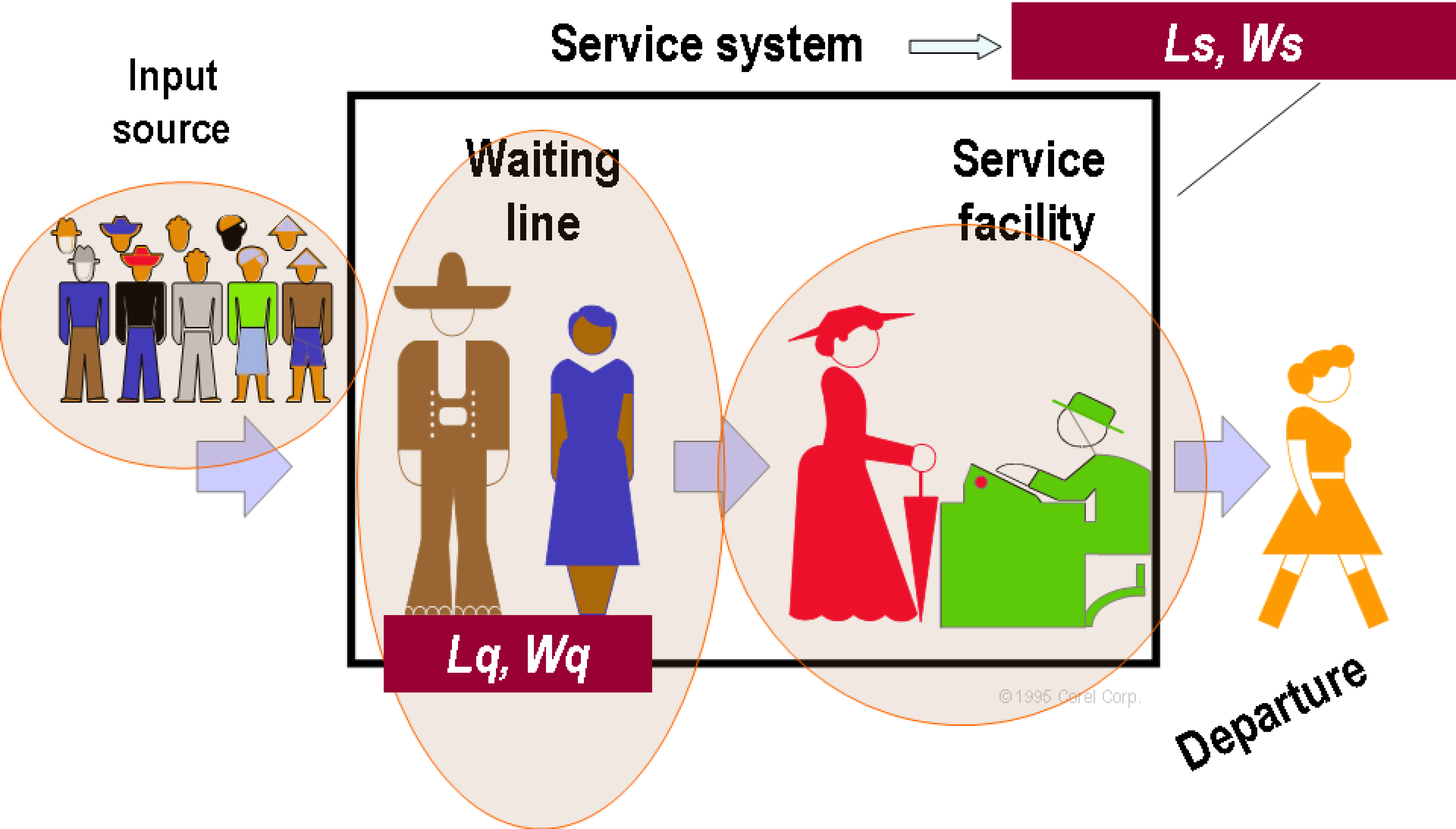
Queuing System

System comprises Queue and Servers



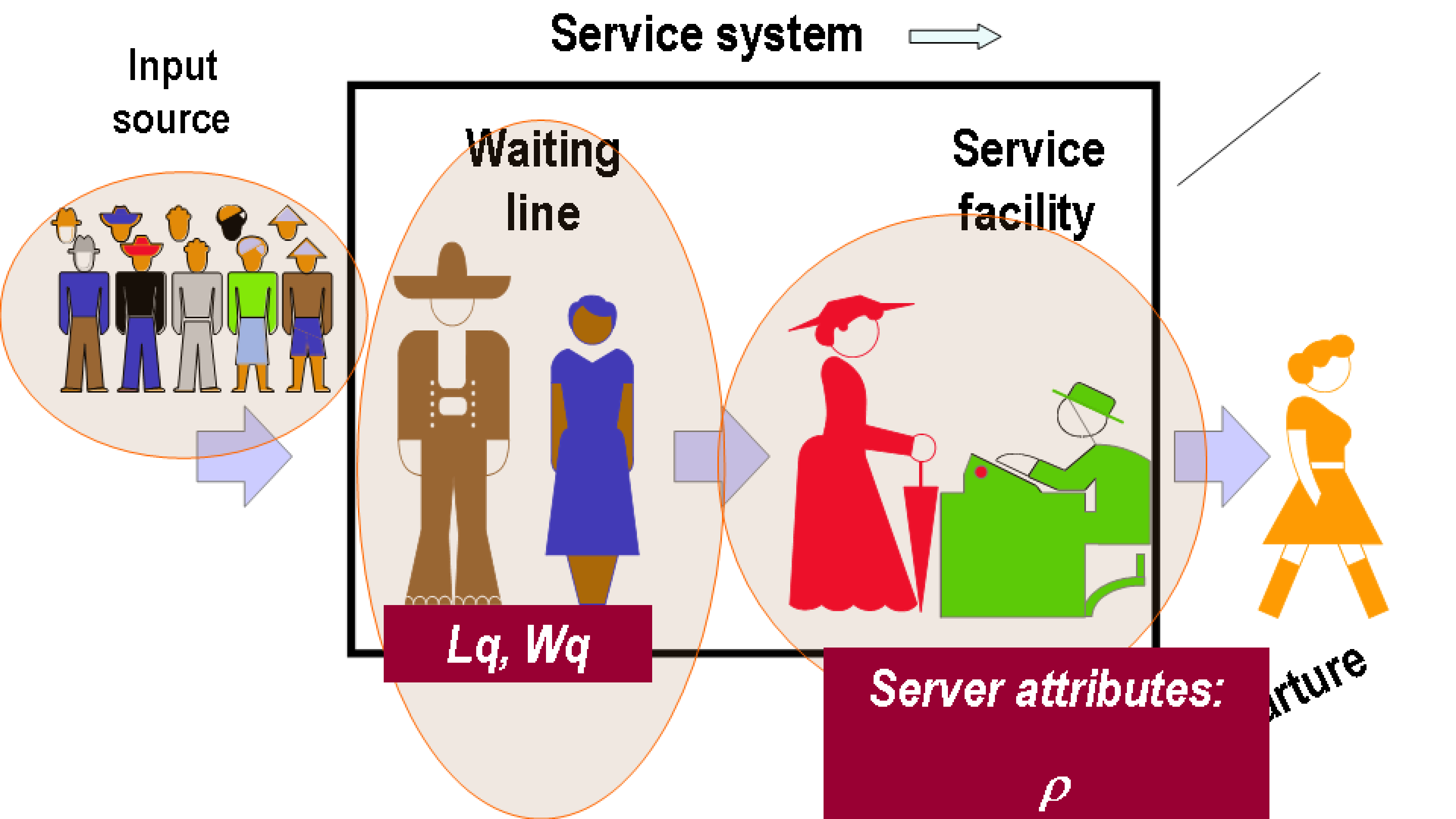
Waiting Line System

System comprises Queue and Servers



Waiting Line System

System comprises Queue and Servers



Assumptions of the Basic Simple Queuing Model

- Arrivals are served on a first come, first served basis (FCFS) – no priorities
- Arrivals are independent of preceding arrivals
- Arrival rates are described by the Poisson probability distribution, and customers come from a very large population
- Service times vary from one customer to another, and are independent of one and other; the average service time is known
- Service times are described by the negative exponential probability distribution
- The service rate is greater than the arrival rate

Types of Queuing Models

- Infinite Capacity
- Simple (M/M/1)
  - Example: Information booth at shopping mall
- Multi-channel (M/M/S)
  - Example: Airline ticket counter
- Constant Service (M/D/1)
  - Example: Automated car wash
- Finite Capacity
  - Example: Department with only 7 drills

Simple (M/M/1) Model Characteristics

- Type: Single-channel, single-phase system
- Input source: Infinite; no balks, no reneging
- Arrival distribution: Poisson
- Queue: Unlimited; single line
- Queue discipline: FIFO (FCFS)
- Service distribution: Negative exponential
- Relationship: Independent service & arrival
- Service rate > arrival rate

Simple (M/M/1) Model Equations

The M/M/1 Waiting line system has a single channel, single phase, Poisson arrival rate, exponential service time, infinite capacity and First-in First-out queue discipline.

Infinite Capacity Queues

Let P_i = Probability that the process is in state i (S_i).

It is a measure of the proportion of time that

the process stays in state i. where $i = 0,1,2,..........$

Similarly let , λ = Arrival rate and μ = Departure rate

For a steady state process,

the rate at which process enters State i will equal

the rate at which process leaves State i.

State Transitions (S_0 to S_1 and vice versa)

When system is empty, the process state is set equal to zero (S_0).

This state of the process will change to 1 (S_1) when an arrival occurs.

If the process is in state 1, the process will revert to state 0 when a departure occurs.

Hence, if P_0 is the proportion of time that process stays in State 0 and λ is the rate of arrival, then;

$$\text{Rate at which process leaves State } 0 = \lambda P_0$$

Similarly, if P_1 is the proportion of time the process stays in state 1 and μ is the rate of departure of the process from the state, then;

$$\text{Rate at which process enters State } 0 = \mu P_1$$

State Transitions (S_0 to S_1) : Steady State Equations

For a steady state process;

Rate at which process leaves a state = Rate at which process enters a state

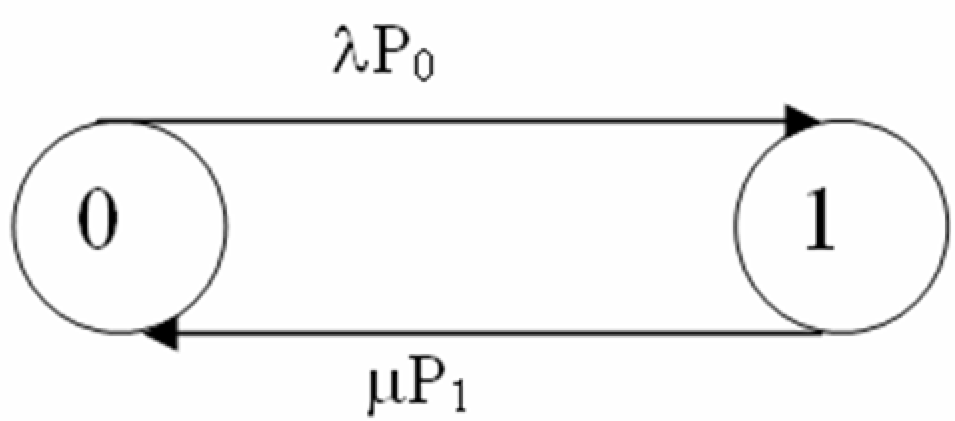
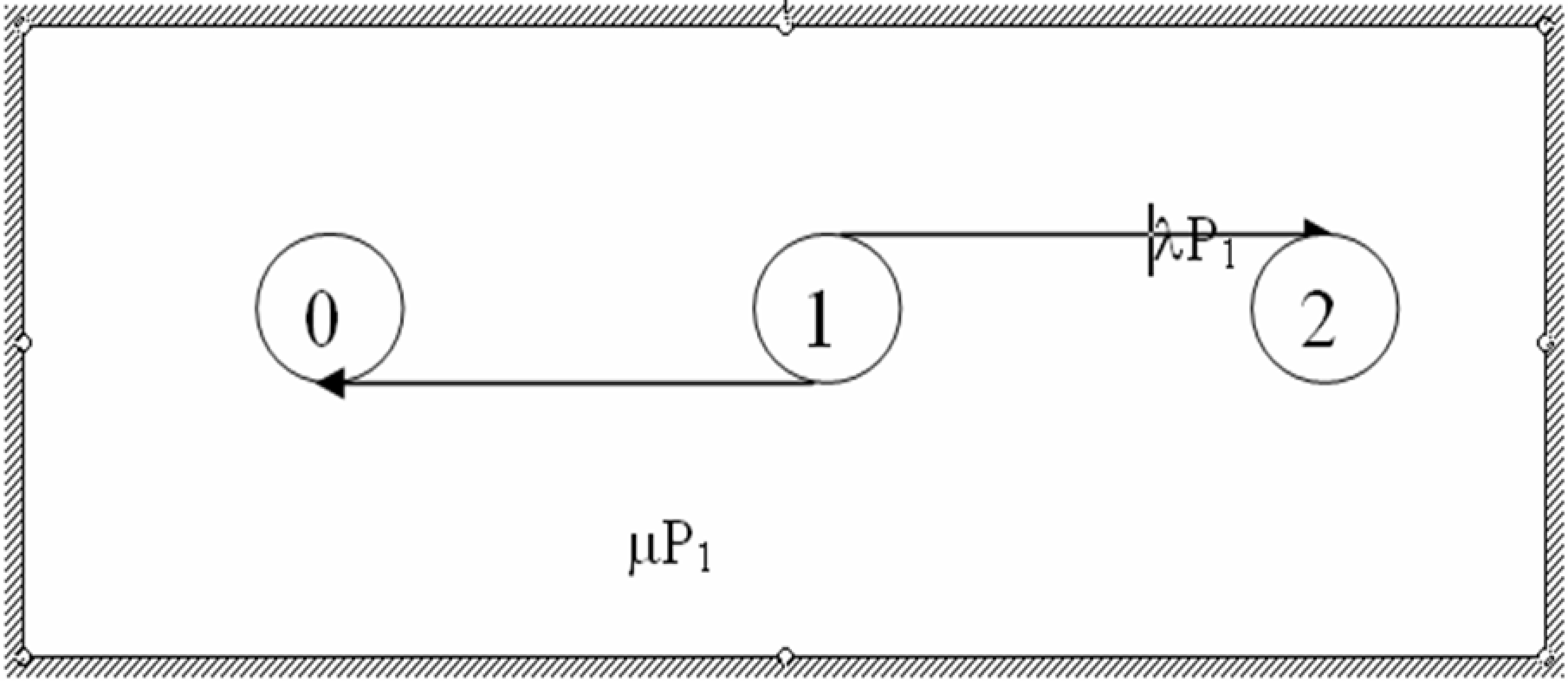


Fig.1. State transition (S_0 to S_1)

So, $\lambda P_0 = \mu P_1$ or, $P_1 = (\lambda/\mu) P_0$

Rate of Leaving State 1

When process is in state S_1, the state of the process will change to S_2 when an arrival occurs or it will revert to state 0 (S_0) when a departure occurs. The rate at which process will *leave* state 1 (S_1) will be equal to $\lambda P_1 + \mu P_1$

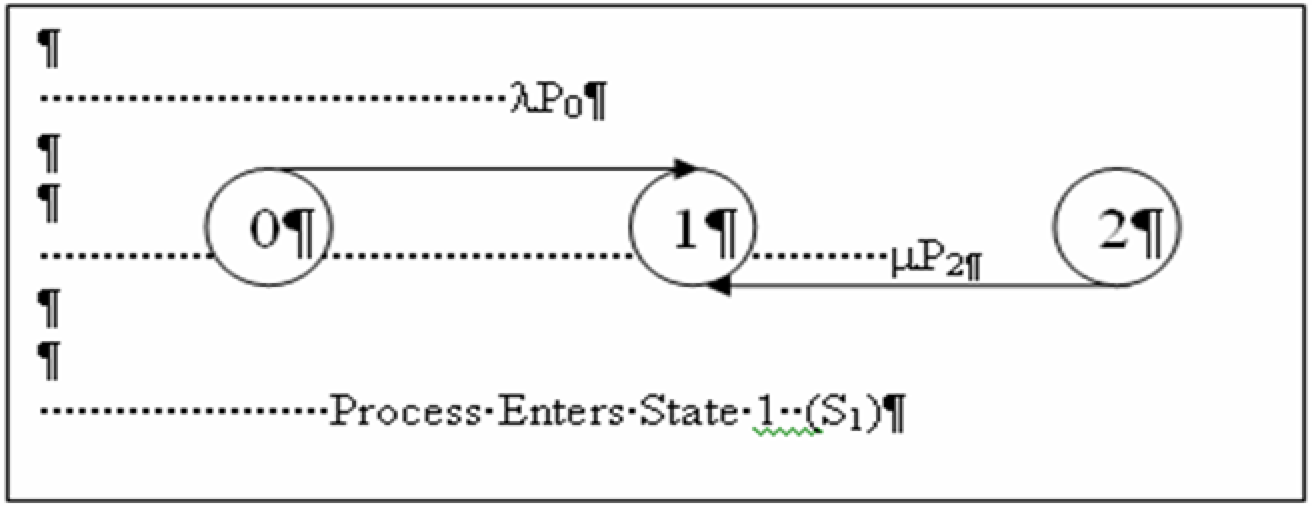


Similarly the process will enter state 1 (S_1); either when an arrival occurs and process is in state 0 (S_0)

OR

when a departure occurs and process is in state 1 (S_1).

So the rate at which the process will enter state 1 (S_1) will be equal to $\lambda P0 + \mu P2$



Rate of Entering State 1

Steady State Equation For State 2

For a steady state process;

Rate at which process leaves state S1 =

Rate at which process enters state S1

So, $\lambda P_1 + \mu P_1 = \lambda P_0 + \mu P_2$

But, $\mu P_1 = \lambda P_0$

So, $\lambda P_1 + \lambda P_0 = \lambda P_0 + \mu P2$

or, $\lambda P_1 = \mu P_2$

$$P_2 = (\lambda/\mu)\, P_1 = (\lambda/\mu)^2\, P_0$$

Rate of Leaving State *k*

Continuing the arguments, suppose the process is in state k (S_k).

Clearly, the process will *leave* state k in two ways;

either an arrival occurs or a departure occurs.

If a departure occurs, process will revert to state k-1 (S_{k-1}).

If an arrival occurs, the process will assume the new state k+1 (S_{k+1}).

If P_{k-1} and P_{k+1} are the respective probabilities of the corresponding states k-1

then the rate at which process *leaves* process k will be $\lambda P_k + \mu P_k$.

Probabilities of Leaving State *k*

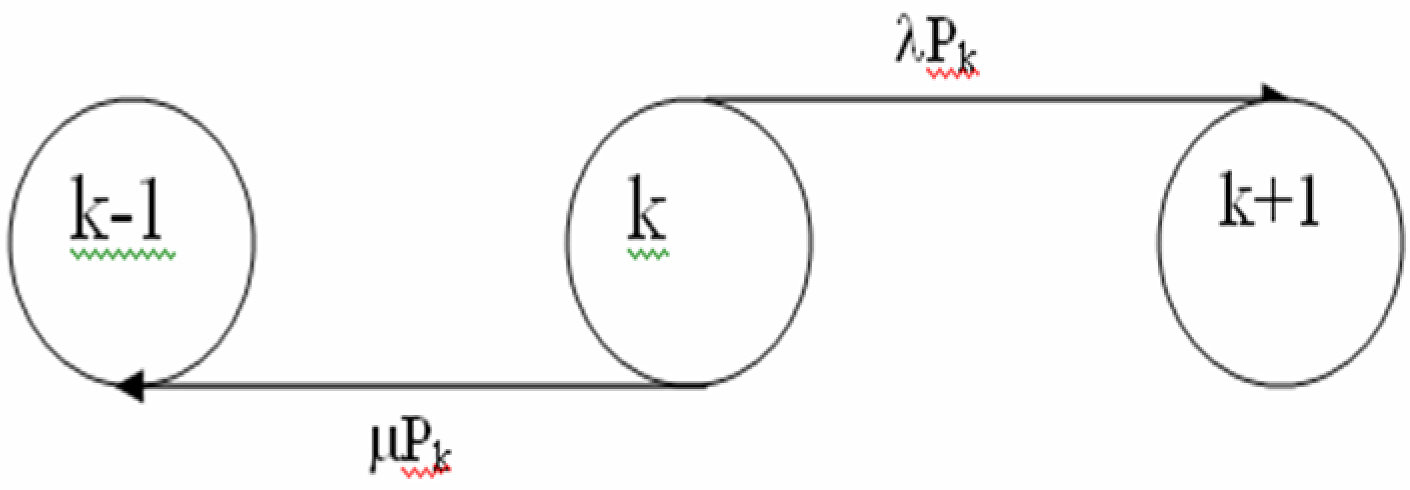


Fig.4. Process *Leaves* State k (S_k)

Rate of Entering State *k*

Similarly the process will *enter* state k (S_k);

either when an arrival occurs and process is in state k-1 (S_{k-1}) or

when a departure occurs and process is in state k+1 (S_{k+1}).

So the rate at which the process will *enter* state k (S_k) will be equal to

$\lambda P_{k-1} + \mu P_{k+1}$ as shown below in Fig.5.

Probabilities of Entering State *k*

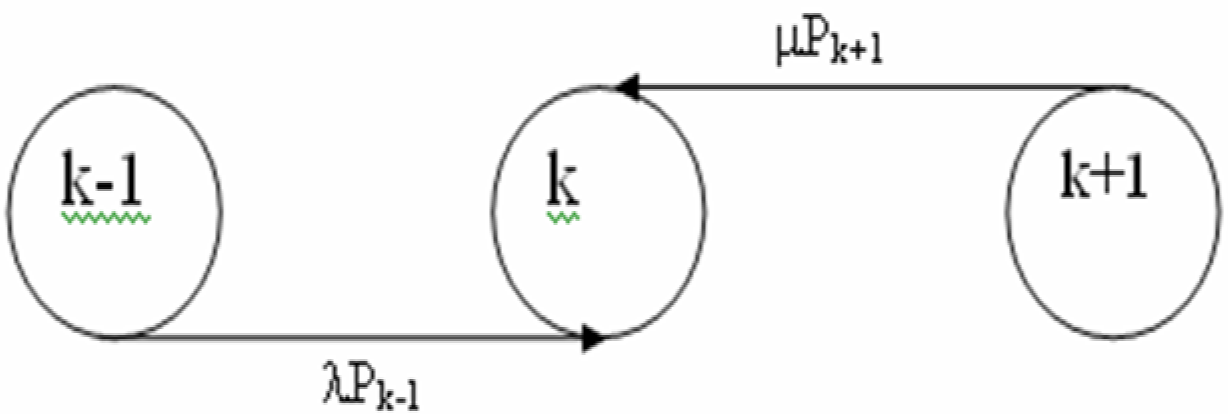


Fig.5. Process *Enters* State k (S_k)

Probability For State k

Using flow rate balance for process k;

$$\lambda P_k + \mu P_k = \lambda P_{k-1} + \mu P_{k+1}$$

Using induction from Eq(3) ,

$$P_k = (\lambda/\mu)^k P_0$$

For an infinite capacity queue, k=0,1,2,..........................,

$$\sum_{k=0}^{\infty} P_k = 1 = \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k P_0 = 1$$

∞

From axioms of Probability theory , P(S) =1

$$\sum_{k=0}^{\infty} a^k \qquad \frac{1}{1-a}$$

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

$$P_k = (\lambda/\mu)^k P_0$$

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

, { Note that ($\lambda < \mu$) in this expression ?)

L

Let, L = Expected value of process states in the long run. Then, using expectation theory,

$$L = \sum_{k=0}^{\infty} kP_k = \sum_{k=0}^{\infty} k\left(\frac{\lambda}{\mu}\right)^k \left(1-\frac{\lambda}{\mu}\right)$$

$$L = \left(1-\frac{\lambda}{\mu}\right) \sum_{k=0}^{\infty} k\left(\frac{\lambda}{\mu}\right)^k$$

using math :

$$\sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}$$

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

Lq

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

$$= \sum_{k=1}^{\infty} kP_k - \sum_{k=1}^{\infty} P_k$$

$$= L - (1 - P_0) = L - (1 - 1 + \frac{\lambda}{\mu})$$

$$= \mathrm{L} - \frac{\lambda}{\mu} = \frac{\lambda}{(\mu-\lambda)} - \frac{\lambda}{\mu} = \frac{\lambda\mu - \lambda\mu + \lambda^2}{\mu(\mu-\lambda)}$$

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

Similarly, let W= Expected life of the different states of a process
.

Then, ratio of L/W $= \lambda$. So, $L = \lambda W$ or $W = L/\lambda$

Example :
Suppose L is the expected number of parts that stay in the machine shop over a long period of time and
W is the expected time a part spends in the shop.
In particular, L = 100 and W = 25 minutes.
Then, L/W= 100/25. It is equal to 4 parts/minute.
Hence, it is concluded that average arrival rate of the parts in the job shop is 4.
So, λ = 4 parts/minute.

Example : Customers arrive at a Poisson rate of one per every 12 minutes, and that the service time is exponential at a rate of one per 8 minutes. What are L and W ?

Solution: Since $\lambda = 1/12$, $\mu = 1/8$,

$$L = \frac{\lambda}{\mu - \lambda} = \frac{1/12}{1/8 - 1/12} = 2 \quad , W = \frac{L}{\lambda} = \frac{2}{(12)^{-1}} = 24$$

Suppose , arrival occurs after every 10 minutes. Then L = 4 and W = 40

Relationship between L & W

To better understand the relationship of L, W, λ and μ, rewrite the Eqs (6) and (7) as follows

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

As can be seen, the values of L and W are highly sensitive to the ratio λ/μ. As the ratio $\lambda/\mu \to 1$, a slight increase in the ratio of λ/μ will result in drastic increase in the values of L and W.

Finite Capacity Queues

- In real server/customer systems, there is an implied upper limit on the number of customers in an accumulation line. If N is the system capacity, it means that there can be no more than N customers in the system at any time. In such a situation, an arriving customer will not join the queue if there are already N customers present.
- As before, let P_k defines the probability that there are k customers in the system. Note that k is bounded by [0,N] now. Using the principle that rate at which customers enter the system equals the rate at which customers departs the system, we obtain the relationship between transition probabilities as follows:

Steady State Equations

State	Rate at which process *leaves* State k = Rate at which process *enters* State k
k=0	$\lambda P_0 = \mu P_1$
$1 \leq k \leq N-1$	$(\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+1}$
k=N	$\mu P_N = \lambda P_{N-1}$

Note the contrast of state transition probabilities in this case. Since there is upper limit of N customers in the system, the process can have a maximum of N states. The rate at which process can enter the last state N (S_N) is the product of arrival rate λ and the probability P_{N-1} ; i.e., rate at which process enters state N = λP_{N-1}. Similarly, the process will leave the state N at rate μP_N. Using the rate equality principle, the balance equation for last state N for finite capacity system is; $\mu P_N = \lambda P_{N-1}$

State Probabilities

When $k = 0$, $\lambda P_0 = \mu P_1$. Simplifying $P_1 = \frac{\lambda}{\mu} P_0$

For any state k=n, the rate equality balance equation is

$$(\lambda+\mu)\, P_n = \lambda P_{n-1} + \mu P_{n+1}$$

Rearranging, $P_{n+1} = \frac{\lambda}{\mu} P_n + \left(P_n - \frac{\lambda}{\mu} P_{n-1} \right)$.

State Probabilities with Respect To P_o

When k=1, $(\lambda+\mu) P_1 = \lambda P_0 + \mu P_2$ and, $P_2 = \frac{\lambda}{\mu} P_1 + \left(P_1 - \frac{\lambda}{\mu} P_0 \right) = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu} \right)^2 P_0$

When k=2, $(\lambda+\mu) P_2 = \lambda P_1 + \mu P_3$ and, $P_3 = \frac{\lambda}{\mu} P_2 + \left(P_2 - \frac{\lambda}{\mu} P_1 \right) = \frac{\lambda}{\mu} P_2 = \left(\frac{\lambda}{\mu} \right)^3 P_0$

Similarly, When k=3, $(\lambda+\mu) P_3 = \lambda P_2 + \mu P_4$ and,

$$P_4 = \frac{\lambda}{\mu} P_3 + \left(P_3 - \frac{\lambda}{\mu} P_2 \right) = \frac{\lambda}{\mu} P_3 = \left(\frac{\lambda}{\mu} \right)^4 P_0.$$

P_N and P_o

So, when k=N-2, $(\lambda+\mu) P_{N-2} = \lambda P_{N-3} + \mu P_{N-1}$ and,

$$P_{N-1} = \frac{\lambda}{\mu} P_{N-2} + \left(P_{N-2} - \frac{\lambda}{\mu} P_{N-3} \right) = \left(\frac{\lambda}{\mu} \right)^{N-1} P_0 .$$

$$\text{Finally, } P_N = \frac{\lambda}{\mu} P_{N-1} = \left(\frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^{N-1} P_0 = \left(\frac{\lambda}{\mu} \right)^{N} P_0 .$$

State Probabilities

By using the fact, $\sum_{k=0}^{N} P_k = 1$, the value of P_0 can be estimated as follows:

$$1 = \sum_{k=0}^{N} \left(\frac{\lambda}{\mu}\right)^k P_0$$

$$1 = \left(\frac{1-(\lambda/\mu)^{N+1}}{1-\lambda/\mu}\right) P_0 \qquad \{\text{Remember } \sum_{k=0}^{N} x^k = \frac{1-x^{N+1}}{1-x} \}$$

$$\text{So, } P_0 = \frac{(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}} \qquad \text{Eq(8)}$$

Expected Length (L)

$$P_0 = \frac{(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}} \quad , \qquad P_k = \frac{\left(\frac{\lambda}{\mu}\right)^k \left(1-\frac{\lambda}{\mu}\right)}{1-\left(\frac{\lambda}{\mu}\right)^{N+1}} \quad \text{where, } k = 0,1,2,\ldots\ldots, N$$

Length of units in the line, L is expressed by; $L = \sum_{k=0}^{N} kP_k$

$$L = \frac{(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}} \sum_{k=0}^{N} k\left(\frac{\lambda}{\mu}\right)^k \qquad L = \frac{\lambda\left[1+N\left(\frac{\lambda}{\mu}\right)^{N+1} - (N+1)\left(\frac{\lambda}{\mu}\right)^{N}\right]}{(\mu-\lambda)\left[1-\left(\frac{\lambda}{\mu}\right)^{N+1}\right]}$$

Expected Time (W)

Next, we find an expression for W. As W is the expected amount of time a customer spends in the line, it is equal to L/λ. But we should be careful in estimating the value of λ for finite capacity systems. Since, maximum capacity of the system is N, all arriving customers do not join the system if system is full. Since P_N is the fraction of customers who do not join the system, proportion of arriving customers that join the system are $1 - P_N$. Hence, if λ_a is the actual arrival rate, then $\lambda_a = (1 - P_N)\,\lambda$. Then Expected time a customer spends in the system is,

$$W = L/\lambda_a = L/(1 - P_N)\,\lambda$$

Example

A work station receives parts automatically from a conveyor. An accumulation line has been provided at the work station and has a storage capacity of 5 parts (N=6). Parts arrive at a poisson rate of 1 per minute; service time is exponentially distributed with a mean of 45 seconds. If queue length is full (N=6), parts are diverted to another station. Find L and W.

Solution

Note the system has a capacity of N=6, the accumulation line has a capacity of 5.

Arrival rate $\lambda = 1$ and service rate, $\mu = 4/3$

Using Eq (9), $P_0 = 0.28851$, $P_1 = 21638$, $P_2 = 16229$, $P_3 = 12172$, $P_4 = 0.09121$
$P_5 = 0.06847$, $P_6 = 0.05135$; Note that $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1$

The system will be empty 28.85% of the time. It means that the server (work station) will be idle 28.85% of the time. Note $P_6 = 0.0513$. It means the accumulation line (Queue) will remain full 5.13% of the time. It implies that arriving customers will find system in 'Full House' 5.13% of the time. So 5.13% of the arriving customers will not join the system.

L and Lq

Note that $$L = \sum_{k=0} kP_k = 1.92167$$

Similarly, number of units in the accumulation line (Queue) L_q is

$$L_q = \sum_{k=1}^{6} (k-1)P_k = 1.21019$$

In finite capacity single server systems, an arriving customer does not join the system if total units in the system is N. In such situations, actual arrival rate is not λ. It is termed as λ_a and, is given by the the product of $\lambda(1-P_N)$. For the example cited above $P_N = 0.05135$.

So, $\lambda a = (1 - P_N)\ \lambda = (1 - 0.05135)\ (1) = 0.9487$ / minute.