6. QUEUEING SYSTEMS

A queueing system is a system where entities (customers, parts, etc.) arrive and require a service. The entity may or may not be allowed entry to the system (If the system is full the entity "balks"). The entity may or may not have to wait for service (If the service unit or server is idle upon entry, the customer immediately enters service). If the server(s) are busy, the customer waits for service in a queue. Some typical examples of queueing systems include bank teller's windows, drive-in fast food service, grocery store check out counters and theater ticket windows.

The most simple queueing system contains a single service channel as shown in the figure below. Entities arrive to the system at an average rate of λ (entities/unit time). In most queueing systems, the entity enters the queue (in the case where the system is not empty) and joins the waiting line at its end (This is commonly referred to as First-In-First-Out service system). When it is an entity's turn to be serviced it enters service where service takes place at a average rate of μ (entities/unit time). When service is complete, the entity leaves the system.

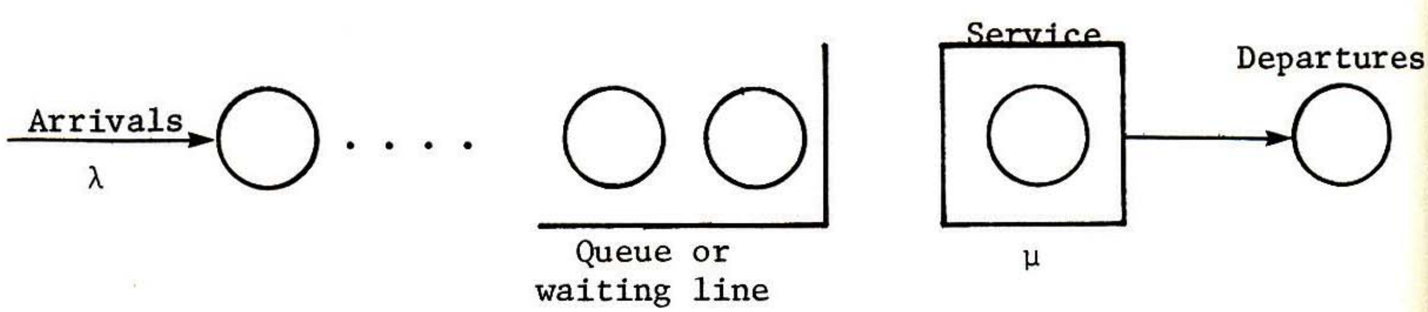

Figure 1. Single Channel Queueing System

The parameters λ and μ are descriptive parameters for the system. If $\lambda \geq \mu$, entities arrive to the system faster than they can be served. The description of such a system is trivial in that the queue will grow to its maximum capacity (or infinity). Normal queueing analysis requires: 1) $\lambda \leq \mu$, 2) λ and μ are homogenous over time, and 3) times between success arrivals and services are independent. The ratio of λ to μ is called the traffic density, ρ and is expressed as

$$\rho = \lambda/\mu$$

In most situations, neither the number of arrivals nor the number of completed services can be predicted with certainty i.e., customers arrive at random times at machining centers, the time required to check-out each customer may be different, etc. Most of the existing queueing formulae assume the λ and μ are random variables which are exponentially distributed. Mathematically stated, if x and y are the times required between successive arrivals and services, then the probability density function of x and y can be expressed by

$$f_1(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$f_2(y) = \mu e^{-\mu x} \qquad y \geq 0$$

Other commonly discussed queueing systems include servers in tandem and parallel servers. Figure 2 and 3 represent these types of systems. A special problem of a queueing system with servers in tandem is that if server k finishes before server k + 1, server k is "blocked" from initiating service on another part.

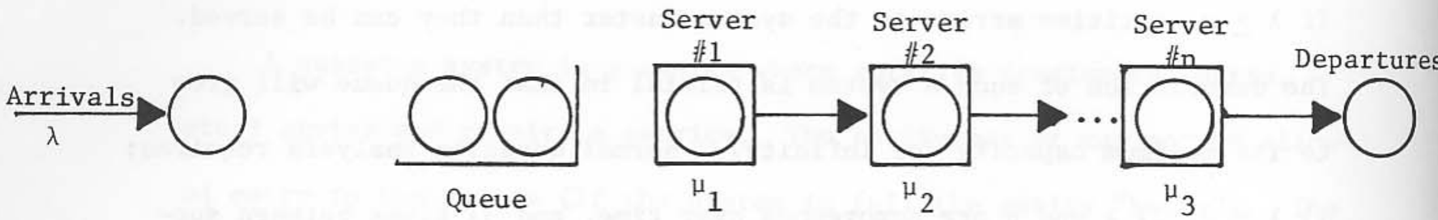

Figure 2. Servers in tandem.

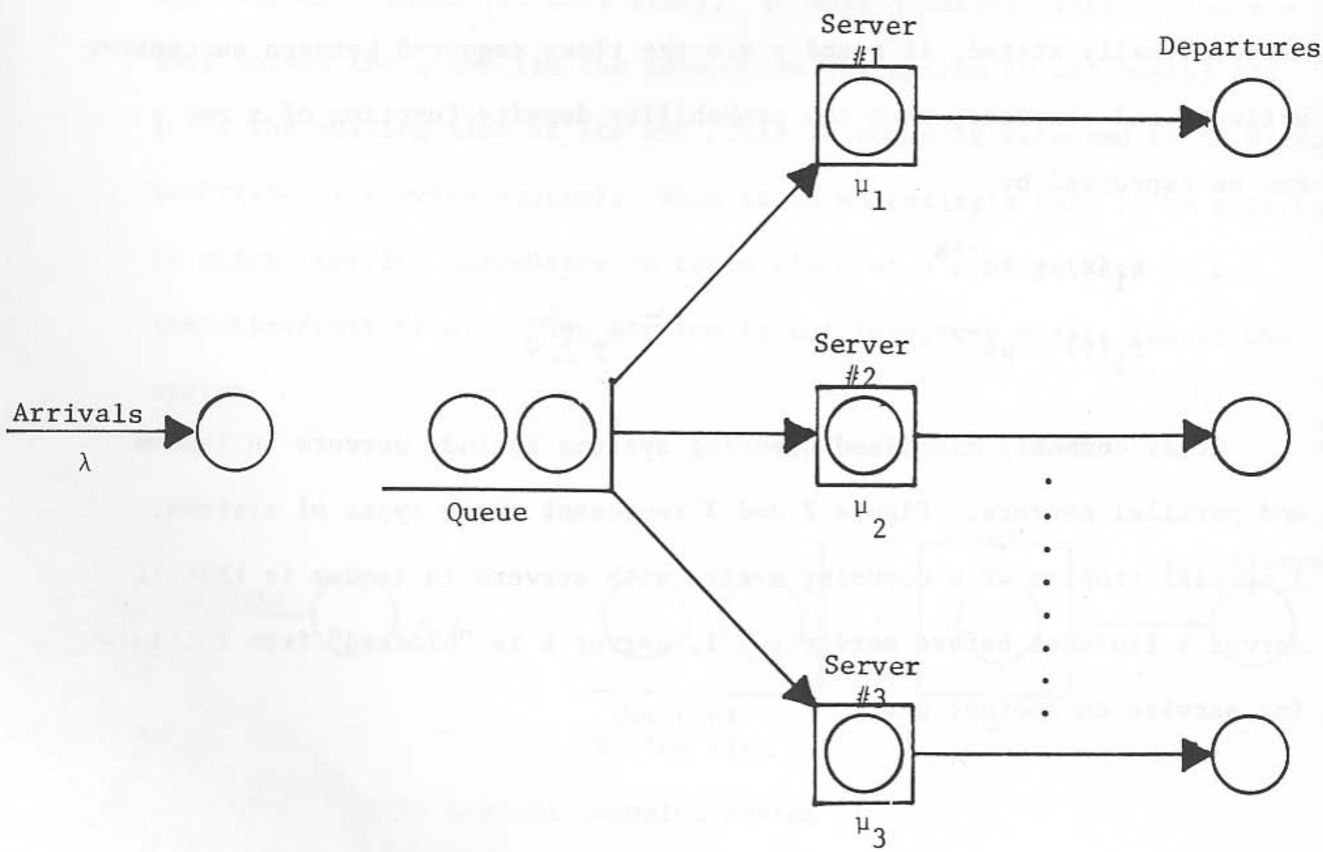

Figure 3. Servers in parallel.

Queueing systems are frequently characterized by using the extended Kendall's notation. Table 2 contains a list of some commonly employed queueing formulae.

Example problem

Customers arrive at a Hot Dog Stand on a major thorough-fare in New York City. The time between customer arrivals is exponential distributed with a mean of 1 min. Customers are served in a FIFO discipline, and service times are exponentially distributed with a mean of .75 minutes. A) What percent of time is the server busy? B) What is the average number in the system? C) How long on average does a customer wait?

A) Percent of time the server is busy

$$\rho = \lambda/\mu = \frac{(1/1)}{(11.75)} = .75 \text{ or } 75\%$$

B) $$L_s = \frac{\lambda}{\mu-\lambda} = \frac{(1/1)}{(1/.75-1/1)} = 3 \text{ people}$$

C) $$W_s = \frac{1}{\mu-\lambda} = 3 \text{ minutes}$$

The notation takes the form

Arrival Process	Service Process	Number of Servers	Limit on number in system	Number in Source	Queue Discipline

The letters used to specify the distribution are

M = negative exponential

E_k = Erlang-K

D = deterministic or constant

G = General

Common queue disciplines include

FIFO - First In, First Out

LIFO - Last In, First Out

RS - Random Selection

PR - PRiority

A number of different outputs can be computed from queueing formulae. Some of the more common include:

L_q - expected number of entities in the queue (this does not include the entity in service) under steady-state

Ls - expected number of entities in the system (including the server) under steady-state.

W_q - expected time spent waiting in the queue (excluding service time)

Ws - expected time spent in the system (including service time)

P_o - *the probability of zero entities in the system.*

P_n - *the probability of n entities in the system.*

B. Selected Queueing Formulas

Notation:

λ - arrival rate

μ - service rate

s - traffic density = $\frac{\lambda}{\mu}$

L_s - mean number in the system

L_q - mean number in the queue

W_s - mean time spent in the system

W_q - mean time spent in the queue

P_o - probability of zero people in the system

P_n - probability of "n" people in the system

	(M/M/1) $(\infty/\infty/FIFO)$	$(M/\infty/FIFO)$	(M/M/FIFO) $(M/\infty/FIFO)$
P_o	$1-\rho$	$\frac{1-\rho}{1-\rho^{n+1}}$	$[\sum_{n=0}^{M}\left(\frac{M!}{(M-n)!}(\frac{\lambda}{\mu})^n\right)]^{-1}$
P_n	$\rho^n(1-\rho)$	$(\frac{1-\rho}{1-\rho^{n+1}})\rho^n$	$\frac{M!}{(M-n)!}(\frac{\lambda}{\mu})^n P_o$
L_s	$\frac{\lambda}{\mu-\lambda}$	$\frac{\rho}{1-\rho} - \frac{(M+1)\rho^{M+1}}{1-\rho^{M+1}}$	$M - \frac{\mu}{\lambda}(1-P_o)$
L_q	$\frac{\lambda^2}{\mu(\mu-\lambda)}$	$L_s - \frac{\lambda(1-P_M)}{\mu}$	$M - \frac{\lambda-\mu}{\lambda}(1-P_o)$
W_s	$\frac{1}{\mu-\lambda}$	$\frac{L_s}{\lambda(1-P_M)}$	$\frac{L_s}{\lambda(M-L_s)}$
W_q	$\frac{\lambda}{\mu(\mu-\lambda)}$	$\frac{L_q}{\lambda(1-P_M)}$	$\frac{L_q}{\lambda(M-L_s)}$
$\{W>t\}$	$e^{-(\mu-\lambda)t}$		$e^{\left(\frac{-\lambda^2}{M\lambda-\mu+\mu P_o}\right)t}$
	$\rho=\frac{\lambda}{\mu}$ for all cases		

	(M/M/s) (∞/∞/FIFO)	(M/∞/FIFO)	(M/M/FIFO) (M/∞/FIFO)
P_o	$\left[\left[\sum_{n=0}^{s-1} \frac{\frac{\lambda n}{\mu}}{n!}\right] + \left[\frac{\frac{\lambda}{\mu}}{s!}^{s} \frac{s\mu}{s\mu-\lambda}\right]\right]^{-1}$	$\left[\sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s(1-(p/s)^{n-s+1}}{s!\,(1-p/s)}\right]$	$\left[\sum_{n=0}^{s-1} \frac{M!}{(M-n)!n!} \frac{\lambda}{\mu}^n + \sum_{n=s}^{M} \frac{M!}{(M-n)!s!s}^{n-s} (\frac{\lambda}{\mu})^n\right]^{-1}$
P_n	$0<n<s$: $\frac{\left(\frac{\lambda}{\mu}\right)^n P_o}{n!}$; $n<s$: $\frac{\left(\frac{\lambda}{\mu}\right)^n P_o}{s^{n-1} s!}$	$0<n<s$: $\frac{\rho^n}{n!} P_o$; $s<n<M$: $\frac{\rho^n}{s!s^{n-s}} P_o$	$0<n<s$: $P_o\left(\frac{M!}{(M-n)!n!}\right)(\frac{\lambda}{\mu})^n$; $s<n<M$: $P_o\left(\frac{M!}{(M-n)!s!s^{n-s}}\right)(\frac{\lambda}{\mu})^n$
L_s	$L_q + P$ values can be read off chart	$L_q + \frac{\lambda(1-P_n)}{\mu}$	$\left(\sum_{n=0}^{s-1} n\right) P_n + L_q + s\,(1 - \sum_{n=0}^{s-1} P_n)$
L_q	$\frac{\rho^{s+1}}{(s-1)!\,(s-\ \ )^2} P_o = \frac{s\rho}{(s-\ \ )^2} P_s$	$P_o\left(\frac{\rho^{s+1}}{(s-1)!\,(s-\rho)^2}\right)\left[\frac{1-\left(\frac{\rho}{s}\right)^{M-s} -(M-s)}{\left(\frac{\rho}{s}\right)^{M-s}\left(1-\frac{\rho}{s}\right)}\right]$	$\sum_{n=s}^{M} (n-s)\, P_n$
W_s	$W_q + \frac{1}{\mu}$	$\frac{L_s}{\lambda(1-P_M)}$	$\frac{L_s}{\lambda\,(M-L_s)}$
W_q	$\frac{L_q}{\lambda}$	$\frac{L_q}{\lambda(1-P_M)}$	$\frac{L_q}{\lambda\,(M-L_s)}$
{W>t}			$e^{-\mu t} \sum_{n=0}^{s-1} P_n + \sum_{n=1}^{M-s} \mu(s\mu)^n \frac{e^{-\mu t}}{(s\mu-\mu)^n} - \sum_{j=1}^{n} \frac{t^{n-j} e^{-s\mu t}}{(n-j)!\,(s\mu-\mu)} P_{n+s+1}$
	$\rho = \frac{\lambda}{\mu}$ *for all cases*		$P_n = \frac{(M-n)\, P_M}{\sum_{n=0}^{M} (M-N) P_M}$

C. Selected References

1) Guttman, I., Wilks, S. S., Introductory Engineering Statistics, John Wiley and Sons, Inc., M.Y., 1967.

2) Walpole, R. E., Myers, R. H., Probability and Statistics for Engineers and Scientists, Macmillan Publishing Co., N.Y., 1972.

3) Duncan, A. J., Quality Control and Industrial Statistics, Richard D. Irwin, Inc., Homewood, Ill., 1974.

4) Grant, E. L., Leavenworth, R. S., Statistical Quality Control, McGraw Hill Book Co., N.Y., 1974.

5) Phillips, D. T., Ravindrau, A. and Solberg, J. J., Operations Research Principles and Practice, John Wiley, 1977.

6) Schmidt, J. W. and Davis, R. P., Foundations of Analysis in Operations Research, Academic Press, 1980.