MANAGEMENT SCIENCE
Vol. 12, No. 5, January, 1966
Printed in U.S.A.

ON JOCKEYING IN QUEUES*

ERNEST KOENIGSBERG†

Matson Research Corporation, San Francisco

There are numerous queueing situations in which those awaiting service may be allowed to make choices which affect the time spent in the service system. In such cases, it should be possible to formulate a "strategy" for customers to follow in order to optimize a given parameter. A whole class of queue problems which involve "jockeying" can be regarded in this way, but has so far received little attention. Jockeying can be described as the movement of of a waiting customer from one queue to another (of shorter length or which appears to be moving faster, etc.) in anticipation of a shorter delay. We have so far considered steady state solutions for just a few of the various possible jockeying disciplines in two-server systems with heterogeneous "exponential" servers and Poisson inputs. As a basis for comparison, we first treat the heterogeneous server problems where arriving customers join the shorter of two independent waiting lines and are not permitted to jockey. The same problem is then considered allowing instantaneous jockeying from the longer to the shorter line when the difference in line lengths exceeds one. The results, in terms of expected line lengths and delays, are identical to those obtained by Gumbel [6] for heterogeneous servers fed from a single queue. When the same problem, with customer preferences for a specific line, is considered the results are identical to those of Krishnamoorthi [8]. It is important to note that in these two problems the slower server has a larger throughput than might be expected from the classical theory. In other words, the slower server acts as a trapping state. In the last problems treated here, customers join a preferred line and may jockey at a rate proportional to the line lengths or proportional to the difference in line lengths.

Introduction

In recent years, attention has been focused on queue systems which allow variation in the strategy of the servers; such variations include multiple servers, priority service systems, service rate variations and variations in the number of servers as functions of the line length or delay time, etc. [11]‡ There has been very little work reported on the effects of different strategies open to the customer. The only strategy studied in any detail is that of the "impatient customer" who leaves the waiting system [1, 2, 3] never to return.

In this paper, we consider one customer strategy, jockeying, in detail. In the literature of Queue Theory, "jockeying" refers to the movements of customers who have the option of switching from one line to another when several servers, each having a separate and distinct waiting line, are available. Jockeying is not to be confused with the phenomenon of "interference" between queues. This

* Received May 1964.

† This work was completed while the author was at CEIR, Inc., San Francisco, California.

‡ While many references could be cited, the reader is referred to [11], which contains a description of the problems cited and an extensive bibliography.

latter phenomenon, which has been treated by Tanner [13], concerns competition between servers for a common facility. Glazer [5] has treated the jockeying problem by considering the presence of other queues as an external field which interacts with the individual queue system and produces transitions between adjacent states. He then uses the reaction rate equations of chemical kinetics to describe transition mechanisms for such multi-queue systems. The interactions discussed by Glazer differ from those of Camp [4] in that Camp deals with interactions between systems in series rather than in parallel. Saaty [11, pp. 284–285] has outlined the structure of one problem of this type, but does not provide the solution.[1]

Jockeying is quite common in many queue systems. Most of us have been jockeying for years—switching lines in auto license offices and supermarkets, changing lanes on highways, changing routes in rush-hour traffic and changing suppliers (servers) when confronted by long queues. At times (e.g., in supermarkets) we jockey with full knowledge of the state of both the old and the new line. At other times (e.g., changing routes), we jockey without information on the state of the new line. In some cases, we join a preferred line and do not jockey until we have suffered some delay. In all cases, we jockey in anticipation of a reduction in the total time spent in the service system.

Jockeying can only occur when the customer has one or more alternate waiting lines. Thus, it can only apply to systems with multiple servers *and* multiple waiting lines. We shall here be concerned with two server systems, each of which has its distinct waiting line. Such situations are quite prevalent, although it is well known that delays are shorter if all servers are fed from a single line (monitored, perhaps, by a "Maitre d' Hotel"). Space limitations often preclude the single line and, to most customers, the sheer size of the queue is more meaningful than the "velocity" of the queue. The results for two server queues, in most cases, indicate the form of solution for n server queues.

We shall be concerned here with rather realistic situations and strategies of a simple form which one finds quite frequently. There is no doubt that better strategies can be found [8], but they generally require more information about the system than most customers have, or are willing to take the trouble to obtain. The customer intent on "beating the system" will always find a way to do so.

The paper is limited to systems defined by:

1) Arrival rate is Poisson; at a rate λ when customers have no server preference, at rates λ_1 and λ_2 for each server when preference is expressed ($\lambda = \lambda_1 + \lambda_2$ or $\lambda_i = \pi_i\lambda$, $\sum \pi_i = 1$);
2) Service rate is exponential at each server with service rates μ_1 and μ_2 for the respective servers;
3) Service is first-come first-served in any queue (but a jockeying customer joins the end of the new line); and
4) Various strategies, to be defined in each instance.

[1] Saaty's problem is a special case of our Problem 3, the solutions of which are given here.

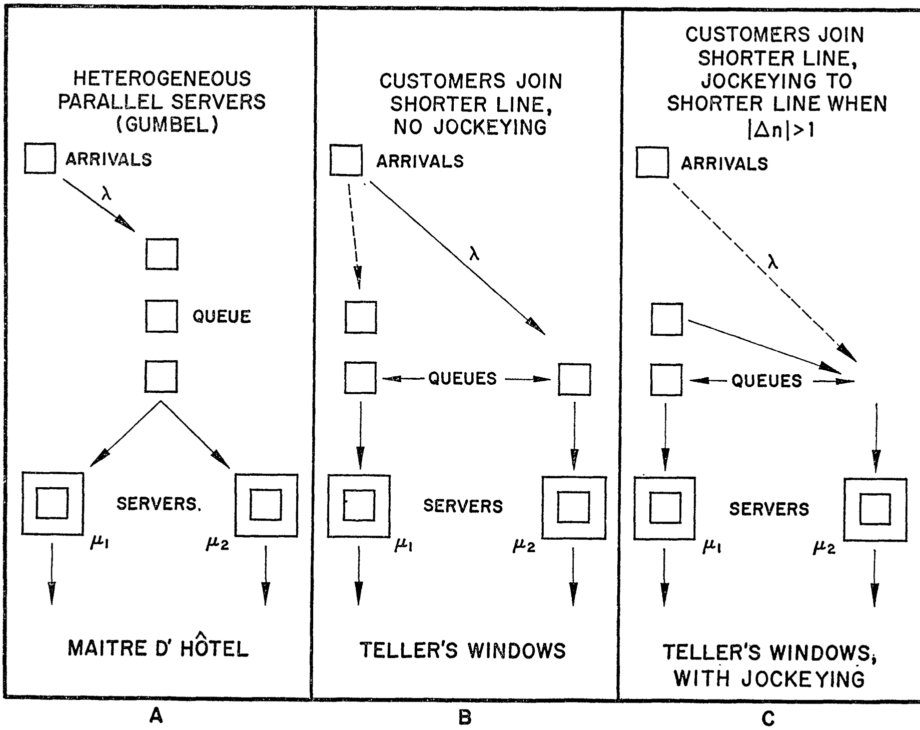

FIGURE 1

For purposes of comparison with the jockeying situations, we consider three "standard" queue situations, two of which are already well reported in the literature. The various strategies will be referred to by descriptive names to avoid confusion in the discussion. The strategies for which new results are obtained are indicated by an asterisk.

The results presented here are for the equilibrium or steady state condition and do not take into account the dynamics of jockeying. Thus we have omitted, in this paper, any consideration of the detailed jockeying history of a specific customer, nor can we calculate the delays suffered by a customer who jockeys exactly n times before being serviced. The steady state equations do allow the calculation of the mean line length and its variance, from which the average delay can be calculated.

The figure of merit of service to the customer used is the average waiting line length. In the absence of jockeying the delay suffered by a customer is proportional to the line length if the service discipline is first-come first-served. For other disciplines the mean waiting time is the same as for the first-come first-served discipline, but the variance of waiting time is greater. The figure of merit used for the server is the probability that the server is idle.

1. Maitre d' Hotel (*Figure 1* (*a*))

This is the well known multiple server model [11], which has been well reported in the literature.

Customer Strategy

A unit arrives to find

a) Both servers engaged: It waits in a single queue in order-of-arrival sequence; the unit at the top of the queue occupies the server that becomes vacant first.

b) Only one server free: It occupies the free channel.

c) Both channels free: It chooses either channel with equal probability.

A modification of this strategy which we shall call "Maitre d' Hotel with Preferences" has been studied by Krishnamoorthi [8]. It differs from Gumbel's case only in that the last statement describing the customer strategy is changed to read

c′) Both servers free: It chooses Server 1 with Probability π_1, and Server 2 with Probability π_2, where $\pi_1 + \pi_2 = 1$.

The results for this case are well known for homogeneous servers as "the multiple channel queue system" [11]. We include Gumbel's results for heterogeneous servers [6], converted to our notation for two-channel systems.

$$Q_{00} = \frac{2\mu_1 \mu_2 \rho(1 - \rho)}{\lambda^2 + 2\mu_1 \mu_2 \rho(1 - \rho)}, \tag{1.1}$$

$$P_0^1 = \frac{\mu_1 \rho(2\mu_2 + \lambda)(1 - \rho)}{\lambda^2 + 2\mu_1 \mu_2(1 - \rho)\rho}, \tag{1.2}$$

$$P_0^2 = \frac{\mu_2 \rho(2\mu_1 + \lambda)(1 - \rho)}{\lambda^2 + 2\mu_1 \mu_2(1 - \rho)\rho}, \tag{1.3}$$

$$\bar{n} = \frac{\lambda^2}{(1 - \rho)(\lambda^2 + 2\mu_1 \mu_2(1 - \rho)\rho)}, \tag{1.4}$$

$$\bar{w} = \frac{\lambda^2 \rho^2}{(1 - \rho)(\lambda^2 + 2\mu_1 \mu_2(1 - \rho)\rho}, \tag{1.5}$$

$$r_i = \frac{\mu_i}{\lambda}(1 - P_0^i) = \frac{\mu_i(\lambda - \mu_i \rho(1 - \rho))}{\lambda^2 + 2\mu_1 \mu_2 \rho(1 - \rho)}$$

The addition of preferences changes the situation. Some of the results have been reported by Krishnamoorthi [8], others will appear in a later section.

$$Q_{00} = \frac{\mu_1 \mu_2(1 - \rho)(1 + 2\rho)}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)}, \tag{1.6}$$

$$P_0^1 = \frac{\mu_1(1 - \rho)[\mu_2(1 + 2\rho) + \lambda(\rho + \pi_2)]}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)}, \tag{1.7}$$

$$P_0^2 = \frac{\mu_2(1 - \rho)[\mu_1(1 + 2\rho) + \lambda(\rho + \pi_1)]}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)}. \tag{1.8}$$

Krishnamoorthi has shown that the average waiting line is minimized when $\pi_i = 1$ if i is the faster channel; i.e., a customer who enters the system when both channels are free goes to the faster server. Krishnamoorthi's contention that because of this result "his" discipline is superior to Gumbel's is correct if the customer knows which server is indeed the faster one. The point of fact to be raised is that a customer arriving at an empty service system has no measure of the relative speeds of the servers, and his optimum strategy, in the absence of such information, would be to select the server by a toss of a coin. Thus, the statistical strategy in the absence of information would be Gumbel's discipline.

2. Tellers' Windows (Figure 1 (b))*

In this system, each server has a unique waiting line to service. Once a customer has selected a line, he remains in that line. The customer's strategy is to join the shorter line.

Customer Strategy

A unit arrives to find

a) Both servers engaged; it joins the shorter line. If both lines are of equal length, it chooses either channel with equal probability.

b) Only one server free; it occupies the free channel.

c) Both servers free; it chooses either channel with equal probability.

A modification of this strategy can be called "Tellers' Windows with Preference." The customer strategy is changed, so that when the customer has a choice (either both servers free or both waiting lines of the same length), he selects Server 1 with probability π_1, and Server 2 with probability π_2, where $\pi_1 + \pi_2 = 1$.

Following the queue discipline and customers' strategies, we have the transition equations:

$$
\begin{aligned}
\frac{d}{dt} Q_{00} &= -\lambda Q_{00} + \lambda_1 Q_{10} + \mu_2 Q_{01} \\
\frac{d}{dt} Q_{10} &= -(\lambda + \mu_1)Q_{10} + \frac{\lambda}{2} Q_{00} + \mu_1 Q_{20} + \mu_2 Q_{11} \\
\frac{d}{dt} Q_{01} &= -(\lambda + \mu_2)Q_{01} + \frac{\lambda}{2} Q_{00} + \mu_1 Q_{11} + \mu_2 Q_{02} \qquad (2.1)\\
\frac{d}{dt} Q_{n_1, n_2} &= -(\lambda + \mu_1 + \mu_2)Q_{n_1, n_2} + \Delta_1 Q_{n_1-1, n_2} + \Delta_2 Q_{n_1, n_2} \\
&\quad + \mu_1 Q_{n_1+1, n_2} + \mu_2 Q_{n_1, n_2+1} (n_1, n_2 \geqq 1)
\end{aligned}
$$

where

$$
\Delta_1 = \begin{cases} 0; & n_1 - 1 > n_2 \\ \lambda/2; & n_1 - 1 = n_2 \\ \lambda; & n_1 - 1 < n_2 \end{cases} \qquad \Delta_2 = \begin{cases} 0; & n_2 - 1 > n_1 \\ \lambda/2; & n_2 - 1 = n_1 \\ \lambda; & n_2 - 1 < n_1 \end{cases} \qquad (2.2)
$$

Writing the equations for Q_{nr} for all n and fixed r, and multiply the n^{th} equation by ξ^n, and writing the generating function $g_r(\xi)$ as

$$
g_r(\xi) = \textstyle\sum_{n=0}^{\infty} \xi^n Q_{nr} \qquad (2.3)
$$

we obtain the following equations:

$$\frac{d}{dt}\sum \xi^n Q_{n0} = g_0(\xi)\left[\frac{\mu_1}{\xi} - \mu_1 - \lambda\right] + g_0(0)\left[\mu_1 - \frac{\mu_1}{\xi}\right] + \mu_2\, g_1(\xi) + \frac{\lambda}{2}\,\xi Q_{00},$$

$$\frac{d}{dt}\sum \xi^n Q_{n1} = g_1(\xi)\left[\frac{\mu_1}{\xi} - \mu_1 - \lambda - \mu_2\right] + g_1(0)\left[\mu_1 - \frac{\mu_1}{\xi}\right] + \mu_2\, g_2(\xi) + \lambda g_0(\xi) - \frac{\lambda}{2}\, Q_{00} + \lambda\xi Q_{01} + \frac{\lambda}{2}\,\xi^2 Q_{11},$$

(2.4)
$$\frac{d}{dt}\sum \xi^n Q_{n2} = g_2(\xi)\left[\frac{\mu_1}{\xi} - \mu_1 - \lambda - \mu_2\right] + g_2(0)\left[\mu_1 - \frac{\mu_1}{\xi}\right] + \mu_2\, g_3(\xi) + \lambda g_1(\xi) - \lambda Q_{01} - \frac{\lambda\xi}{2}\, Q_{11} + \lambda\xi Q_{02} + \lambda\xi^2 Q_{12} + \frac{\lambda}{2}\,\xi^3 Q_{22},$$

$$\frac{d}{dt}\sum \xi^n Q_{n3} = g_3(\xi)\left[\frac{\mu_1}{\xi} - \mu_1 - \lambda - \mu_2\right] + g_3(0)\left[\mu_1 - \frac{\mu_1}{\xi}\right] + \mu_2\, g_4(\xi) + \lambda g_2(\xi) - \lambda Q_{02} - \lambda\xi Q_{12} - \frac{\lambda\xi^2}{2}\, Q_{22} + \lambda\xi Q_{03} + \lambda\xi^2 Q_{13} + \lambda\xi^3 Q_{23} + \frac{\lambda\xi^4}{2}\, Q_{33},$$

$$\cdots .$$

Now, multiplying the r^{th} equation of Eqs. (2.4) by ξ^r and summing over all values of r, we have

(2.5)
$$\sum_{r=0}^{\infty} \xi^r \frac{d}{dt}\sum_{n=0}^{\infty} \xi^n Q_{nr} = \left[\frac{\mu_1}{\xi} - \mu_1 - \lambda\right]\sum_{r=0}^{\infty} \xi^r g_r(\xi) - \mu_2 \sum_{r=1}^{\infty} \xi^r g_r(\xi) + \left[\mu_1 - \frac{\mu_1}{\xi}\right]\sum_{r=0}^{\infty} \xi^r g_r(0) + \mu_2 \sum_{r=1}^{\infty} \xi^{r-1} g_r(\xi) + \lambda\xi \sum_{r=0}^{\infty} \xi^r g_r(\xi).$$

Since we are only interested in the steady state solutions, the left hand side is zero. After reorganizing terms, we have

(2.6)
$$0 = [(\mu_1 + \mu_2)/\xi - \lambda - \mu_1 - \mu_2 + \lambda\xi]\sum_{r=0}^{\infty} \xi^r g_r(\xi) + [\mu_1 - \mu_1/\xi]\sum_{r=0}^{\infty} \xi^r g_r(0) + [\mu_2 - \mu_2/\xi] g_0(\xi).$$

We let

(2.7)
$$g_r(\xi) = a_r(1 + \rho\xi + (\rho\xi)^2 + (\rho\xi)^3 + \cdots) = a_r/(1 - \rho\xi),$$

where

$$\rho = \lambda/(\mu_1 + \mu_2).$$

Substituting Eq. (2.7) in Eq. (2.6), we find

$$a_r = \rho^r a, \tag{2.8}$$

and

$$g_r(\xi) = \rho^r a_0/(1 - \rho\xi). \tag{2.9}$$

Now, we know that

$$g_r(1) = \sum_{n=0}^{\infty} Q_{nr} = P_r^2$$

and, therefore

$$\sum_{r=0}^{\infty} g_r(1) = \sum_{r=0}^{\infty} P_r^2 = 1.$$

Thus, we have

$$a_0 = (1 - \rho)^2$$

and

$$g_r(1) = P_r^2 = \rho^r(1 - \rho),$$
$$g_0(1) = P_0^2 = (1 - \rho) = P_0^1.$$

Similarly,

$$g_r(0) = Q_{0r} = \rho^r a_0 = \rho^r(1 - \rho)^2,$$

and

$$g_0(0) = Q_{00} = (1 - \rho)^2.$$

By suitable manipulation, we find the other measures of interest

(2.10) $\bar{n}_i = \bar{n}_1 = \bar{n}_2 = \rho a_0/(1 - \rho)^3 = \rho/(1 - \rho)$

= mean number at queue i,

(2.11) $\bar{w}_i = \bar{w}_1 = \bar{w}_2 = \bar{n}_i - \rho = \rho^2/(1 - \rho)$

= mean length of waiting line i,

(2.12a) $r_1 = (\mu_1/\lambda)(1 - P_0^1) = \mu_1\rho/\lambda$

= fraction of customers served in queue 1,

(2.12b) $r_2 = (\mu_2/\lambda)(1 - P_0^2) = \mu_2\rho/\lambda$

= fraction of customers served in queue 2.

For the more general case, "Tellers' Windows with Preferences," $\lambda/2$ in Eqs. 2.1 is replaced by $\lambda\pi_1$, and in carrying out the summations to obtain Eqs. 2.6, these terms again disappear. Thus, the general case yields the same results as the case in which $\pi_1 = \pi_2 = \frac{1}{2}$; customers preferences do not affect the standards of service in queue situations which follow the general rule, "choose shortest line and stay in it."

In both cases, the system yields a steady state condition in which the mean waiting lines have the same average length independent of the differences in the service rates. The mean waiting time in a line, however, does depend on the service rate of the line. If the service rates are vastly different, the customer who arrives at a time in which the slower server has a shorter line suffers much longer delays than one who finds the other line free. Further, the proportion of customers served by line i is

$$(\mu_i/(\mu_1 + \mu_2)) \qquad (i = 1, 2).$$

3. Tellers' Windows with Jockeying (Figure 1 (c))*

In this system, customers join the shorter line but may jockey when the other line has more than one unit less than the line they are in. We distinguish two specific jockeying strategies:

A. Probabilistic Jockeying

B. Instantaneous Jockeying

Customer Strategy

The customer strategy is the same as that given above for the "Tellers' Windows" system, with one additional rule:

A. Probabilistic Jockeying

d_P) If line i is longer than line $j (n_i > n_j)$, the customers leave line i at a rate $k(w_i - w_j)$, where w_i and w_j are the lengths of the respective waiting lines.

Here

$$w_i = 0 \qquad \text{if } n_i = 0$$

$$w_i = n_i - 1 \qquad \text{if } n_i \geqq 1.$$

B. Instantaneous Jockeying

d_I) If, at any time, the difference in line lengths exceeds one, the last customer in the longer line "jockeys" to the end of the shorter line. Jockeying is instantaneous.

The effect of rule d_I), which implies that customers always jockey to the shorter line, is to restrict the queue occupancy states to only three types: (n, n), $(n + 1, n)$ and $(n, n + 1)$. Rule d_P) on the other hand, implies that jockeying is an option available to the customer if he so chooses. The factor k is measure of his likelihood to do so. A modification of this strategy can be called "Tellers' Windows, with Preference, with Jockeying." In this case, the strategy is defined in the same way as for "Tellers' Windows with Preference," with the addition of rule d_P) or d_I) above.

A. In the probabilistic case, jockeying is at a rate proportional to the difference in length of the waiting lines and from the longer to the shorter line. Thus, we have the transition equations:

$$\frac{d}{dt}Q_{00} = -\lambda Q_{00} \qquad\qquad + \mu_1 Q_{10} + \mu_2 Q_{01}$$

$$\frac{d}{dt}Q_{n_1,0} = -(\lambda + \mu_1 + w_1 k_1)Q_{00} + \lambda\beta_1 Q_{n_1-1,0} + \mu_1 Q_{n_1+1,0} + \mu_2 Q_{n_1,1} \tag{3.1}$$

$$\frac{d}{dt}Q_{n_1,n_2} = -(\lambda + \mu_1 + \mu_2 + \alpha|w_1 - w_2|) + \lambda\beta_1 Q_{n_1-1,n_2}$$
$$+ \lambda\beta_2 Q_{n_1,n_2-1} + \mu_1 Q_{n_1+1,n_2-1} + \mu_2 Q_{n_1,n_2+1}$$
$$+ k_1\gamma_1(w_1 - w_2)Q_{n_1+1,n_2-1} + k_2\gamma_2(w_2 - w_1)Q_{n_1-1,n_2+1},$$

where

$$w_i = \begin{cases} n_i - 1, & n_i \geqq 1 \\ 0, & n_i = 0, \end{cases} \tag{3.2a}$$

and

$$\alpha = \begin{cases} 0, & w_1 = w_2 \\ k_1, & w_1 > w_2, \\ k_2, & w_1 < w_2 \end{cases} \tag{3.2b}$$

and

$$\beta_i = \begin{cases} \pi_i, & w_1 = w_2 \\ 1, & w_i < w_j, \\ 0, & w_i > w_j \end{cases} \tag{3.2c}$$

and

$$\gamma_i = \begin{cases} 0, & w_i \geqq w_j \\ 1, & w_i < w_j. \end{cases} \tag{3.2d}$$

Then, writing the generating function $g_r(\xi)$ as in Eq. (2.3) and carrying the summations as in the previous section, we find that the resulting equation is exactly Eq. (2.5), or for the steady state solution, Eq. (2.6). Thus, if jockeying is probabilistic, the jockeying action is to no avail, the system operates as if customers joined the shorter line and remained in the selected line. The results are independent of the values of k_i, the jockeying rates. The same result is obtained if we inhibit jockeying when $\Delta w = 1$ (i.e., if the difference in waiting line lengths is 1 or 0 no jockeying can take place). Since the identical result is obtained in another jockeying problem, we postpone a discussion of its significance until later.

B. In the instantaneous case, jockeying takes place instantaneously any time the difference in line lengths exceeds one. Thus, we cannot have states in which $\Delta n = | n_1 - n_2 |$ exceeds one.

The transition equations for this case can be written as follows, since we can only have states of the form (n, n), $(n + 1, n)$ and $(n, n + 1)$:

$$
\begin{aligned}
&\frac{d}{dt} Q_{00} = -\lambda Q_{00} && + \mu_1 Q_{10} + \mu_2 Q_{01} \\
(3.3) \qquad &\frac{d}{dt} Q_{10} = -(\lambda + \mu_1)\, Q_{10} + \frac{\lambda}{2} Q_{00} && + \mu_2 Q_{11} \\
&\frac{d}{dt} Q_{01} = -(\lambda + \mu_2)\, Q_{01} \qquad + \frac{\lambda}{2} Q_{00} && + \mu_1 Q_{11}
\end{aligned}
$$

and, when $n_1, n_2 \geqq 1$,

$$
\begin{aligned}
&\frac{d}{dt} Q_{n,n} = -(\lambda + \mu_1 + \mu_2) Q_{n,n} + \lambda Q_{n-1,n} + \lambda Q_{n,n-1} \\
&\qquad\qquad + (\mu_1 + \mu_2)\,[Q_{n+1,n} + Q_{n,n+1}] \\
(3.4) \qquad &\frac{d}{dt} Q_{n+1,n} = -(\lambda + \mu_1 + \mu_2) Q_{n+1,n} + \frac{\lambda}{2} Q_{n,n} + \mu_2 Q_{n+1,n+1} \\
&\frac{d}{dt} Q_{n,n+1} = -(\lambda + \mu_1 + \mu_2) Q_{n,n+1} \qquad + \frac{\lambda}{2} Q_{n,n} + \mu_1 Q_{n+1,n+1}\,.
\end{aligned}
$$

We recognize that we can arrange the equations in groups of three and solve the simultaneous equations in terms of previous state equations (with Q_{00} as an unknown in Eqs. (3.3)). Thus, we have

$$
(3.5) \qquad Q_{10} = \frac{\lambda}{2\mu_1} Q_{00}\,, \qquad Q_{01} = \frac{\lambda}{2\mu_2} Q_{00}\,, \qquad Q_{11} = \frac{\lambda^2}{2\mu_1 \mu_2} Q_{00}
$$

from Eqs. (3.3) and

$$
\begin{aligned}
&Q_{n,m} \equiv 0, \qquad\qquad |n - m| > 1 \\
&Q_{n,n} = \frac{\lambda^2 \rho^{2(n-1)}}{2\mu_1 \mu_2} Q_{00}\,, \\
(3.6) \qquad &Q_{n+1,n} = \frac{\lambda(\lambda + 2\mu_2 \rho^2)\rho^{2n}}{4\mu_1 \mu_2 \rho(1 + \rho)} Q_{00}\,, \\
&Q_{n,n+1} = \frac{\lambda(\lambda + 2\mu_1 \rho^2)\rho^{2n}}{4\mu_1 \mu_2 \rho(1 + \rho)} Q_{00}\,.
\end{aligned}
$$

Using the definition

$$
\textstyle\sum_{n_1, n_2 = 0}^{\infty} Q_{n_1, n_2} = 1
$$

we can solve for Q_{00}. We find

$$
Q_{00} = 2\mu_1\mu_2\rho(1 - \rho)/(\lambda^2 + 2\mu_1\mu_2\rho(1 - \rho)),
$$

and

$$
\begin{aligned}
(3.7) \qquad &P_0^1 = Q_{00} + Q_{01} = \mu_1\rho(1 - \rho)(2\mu_2 + \lambda)/(\lambda^2 + 2\mu_1\mu_2\rho(1 - \rho)), \\
&P_0^2 = Q_{00} + Q_{10} = \mu_2\rho(1 - \rho)(2\mu_1 + \lambda)/(\lambda^2 + 2\mu_1\mu_2\rho(1 - \rho)).
\end{aligned}
$$

We note that these equations are identical to those for the "Maitre d' Hotel" problem described in the previous section.

Similarly, using the definition

$$\bar{n}_1 = \sum_{n_1, n_2=0}^{\infty} n_1 Q_{n_1, n_2}$$

we find,

$$\text{(3.8)} \quad \begin{aligned} \bar{n}_1 &= \frac{\lambda\,[\mu_2(\rho^3 - \rho^2 + 2\rho + 2) + \mu_1\,\rho(2 + 3\rho - \rho^2)]}{4\mu_1\,\mu_2\,(1 - \rho^2)^2}\, Q_{00}\,, \\ \bar{n}_2 &= \frac{\lambda\,[\mu_1(\rho^3 - \rho^2 + 2\rho + 2) + \mu_2\,\rho(2 + 3\rho - \rho^2)]}{4\mu_1\,\mu_2(\,1 - \rho^2)^2}\, Q_{00}\,. \end{aligned}$$

The mean waiting line lengths are also found in the usual way. They are

$$\text{(3.9)} \quad \begin{aligned} \bar{w}_1 &= \frac{\lambda\rho\,[\lambda(1 + \rho) + 2\rho^2(\mu_2 + \mu_1\,\rho)]}{4\mu_1\,\mu_2(1 - \rho^2)^2}\, Q_{00}\,, \\ \bar{w}_2 &= \frac{\lambda\rho\,[\lambda(1 + \rho) + 2\rho^2(\mu_1 + \mu_2\,\rho)]}{4\mu_1\,\mu_2(1 - \rho^2)^2}\, Q_{00}\,. \end{aligned}$$

The portion of customers served at station i is given by r_i, where

$$r_i = (\mu_i/\lambda)(1 - P_0^{\,i})$$

since the server is working, except where the station is empty. Thus, we have

$$\text{(3.10)} \qquad r_i = \mu_i(\lambda - \mu_i\rho(1 - \rho))/(\lambda_2 + 2\mu_1\mu_2\rho(1 - \rho)).$$

For the more general case, "Tellers' Windows with Preference, Instantaneous Jockeying," we have the transition equations

$$\text{(3.11)} \quad \begin{aligned} \frac{d}{dt} &= -\lambda Q_{00} && + \mu_1\,Q_{10} + \mu_2\,Q_{01}\,, \\ \frac{d}{dt}\,Q_{10} &= -(\lambda + \mu_1)Q_{10} + \pi_1\,\lambda Q_{00} && + \mu_2\,Q_{11}\,, \\ \frac{d}{dt}\,Q_{01} &= -(\lambda + \mu_2)Q_{01} \qquad + \pi_2\,\lambda Q_{00} && + \mu_1\,Q_{11}\,. \end{aligned}$$

and

$$\text{(3.12)} \quad \begin{aligned} \frac{d}{dt}\,Q_{nn} &= -(\lambda + \mu_1 + \mu_2)Q_{nn} + \lambda Q_{n-1,n} + \lambda Q_{n,n-1} + (\mu_1 + \mu_2) \\ &\qquad \cdot [Q_{n+1,n} + Q_{n,n+1}], \\ \frac{d}{dt}\,Q_{n+1,n} &= -(\lambda + \mu_1 + \mu_2)Q_{n+1,n} + \pi_1\,\lambda Q_{n,n} + \mu_2\,Q_{n+1,n+1}\,, \\ \frac{d}{dt}\,Q_{n,n+1} &= -(\lambda + \mu_1 + \mu_2)Q_{n,n+1} \qquad + \pi_2\,\lambda Q_{n,n} + \mu_1\,Q_{n+1,n+1}\,. \end{aligned}$$

Again, we recognize the "triplets" of equations and for the steady state we can solve the series of simultaneous equations. We obtain

$$Q_{10} = \frac{\lambda}{\mu_1}\left[\frac{\rho + \pi_1}{1 + 2\rho}\right] Q_{00}, \qquad Q_{01} = \frac{\lambda}{\mu_2}\left[\frac{\rho + \pi_2}{1 + 2\rho}\right] Q_{00},$$

$$(3.13) \qquad Q_{11} = \frac{\lambda_2}{2\mu_1 \mu_2}\left[\frac{\rho + \dfrac{\pi_1 \mu_2 + \pi_2 \mu_1}{\mu_1 + \mu_2}}{1 + 2\rho}\right] Q_{00},$$

$$Q_{n,m} \equiv 0, \mid n - m \mid > 1$$

and

$$Q_{n,n} = \rho^{2n-2} Q_{11},$$

$$(3.14) \qquad Q_{n+1,n} = \left(\pi_1 + \frac{\rho\mu_2}{\mu_1 + \mu_2}\right)\frac{\rho^{2n-1}}{1 + \rho} Q_{11},$$

$$Q_{n,n+1} = \left(\pi_2 + \frac{\rho\mu_1}{\mu_1 + \mu_2}\right)\frac{\rho^{2n-1}}{1 + \rho} Q_{11}.$$

Solving for Q_{00}, we have

$$(3.15) \qquad Q_{00} = \frac{\mu_1 \mu_2(1 - \rho)(1 + 2\rho)}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)}.$$

The results for average waiting line length and average number in the system are identical to those of Krishnamoorthi [8]. Further, we find

$$(3.16) \qquad P_0^1 = \frac{\mu_1(1 - \rho)[\mu_2(1 + 2\rho) + \lambda(\rho + \pi_2)]}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)},$$

$$P_0^2 = \frac{\mu_2(1 - \rho)[\mu_1(1 + 2\rho) + \lambda(\rho + \pi_1)]}{\mu_1 \mu_2(1 - \rho)(1 + 2\rho) + \lambda(\lambda + \mu_1 \pi_2 + \mu_2 \pi_1)},$$

and, as before

$$(3.17) \qquad r_i = (\mu_1/\lambda)(1 - P_0^i).$$

4. Fixed Lanes (Figure 2 (a))

This is the system in which we have independent single server queues; each with its own input, at a rate λ_i, and its own server, who serves at a rate μ_i. There is no interaction between queues.

Customer Strategy

A unit of type i arrives and joins queue i ($i = 1, 2$). This is the well known single server queue system which has been studied exhaustively [11].

This system reduces to a group of single server queues, each with its own input and its own server. The single server queue has been studied in great detail, for all forms of service time distributions and the results are well known [11]. For a two server system, we have,

$$Q_{00} = (1 - \rho_1)(1 - \rho_2)$$

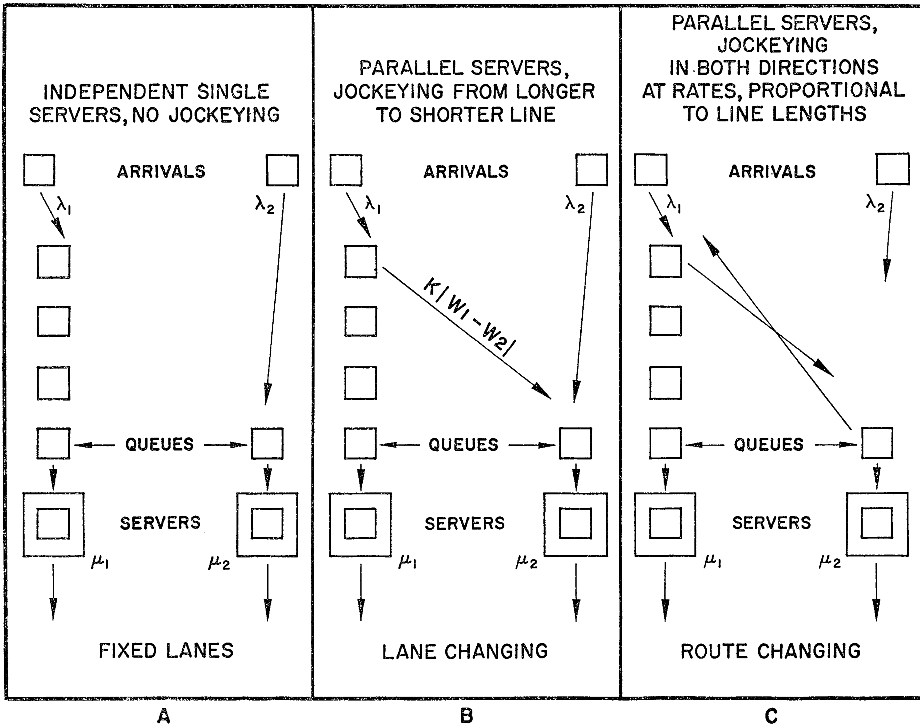

FIGURE 2

$$P_0^1 = (1 - \rho_1) \qquad P_0^2 = (1 - \rho_2), \tag{4.1}$$

$$\bar{n}_1 = \rho_1/(1 - \rho_1) \qquad \bar{n}_2 = \rho_2/(1 - \rho_2), \tag{4.2}$$

$$\bar{w}_1 = \rho_1^2/(1 - \rho_1) \qquad \bar{w}_2 = \rho_2^2/(1 - \rho_2), \tag{4.3}$$

$$r_1 = \gamma_1/(\gamma_1 + \gamma_2) \qquad r_2 = \gamma_2/(\gamma_1 + \gamma_2). \tag{4.4}$$

It is also well known that for homogeneous servers, the mean waiting line and the mean delay is reduced if such a system were replaced by a Maitre d' Hotel system.

5. Lane Changing (Figure 2 (b))*

Here we consider a system in which arrivals choose a line initially, but jockey from the longer to the shorter line at a rate proportional to difference in the length of the two waiting lines. Jockeying in this case is not guaranteed; the customer may or may not change lines when he is in a longer line. It should be noted in this case and the "Tellers' Windows" cases, the customer has information on the state of the other line.

Customer Strategy

a) A unit of type i arrives and joins queue i ($i = 1, 2$).

b) If line i is longer than line $j (n_i > n_j)$, then customers leave line i at a rate

$k_i(w_i - w_j)$, where w_i, w_j are the lengths of the respective waiting lines. A transition rate $k_j(w_j - w_i)$ applies when $n_j > n_i$.

The "Lane Changing" strategy is basically one in which a customer in the longer line has the option of jockeying to the shorter line or remaining in the line he selected originally. The probability of jockeying from longer line i to shorter line j in a time interval t is

$$1 - \exp(-k_i(w_i - w_j)t)$$

where $w_i = n_i - 1$ when $n_i \geqq 1$, $w_j = n_j - 1$ when $n_j \geqq 1$ and zero otherwise. Thus, we have "exponential" jockeying from the longer to the shorter line.

In this situation each line has its own input λ_i and $\lambda = \lambda_1 + \lambda_2$ is the total input. We let n_i indicate the number of customers at the i^{th} queue, either waiting or being served.

The transition equations are

$$\frac{d}{dt} Q_{00} = -\lambda Q_{00} + \mu_1 Q_{10} + \mu_2 Q_{01}$$

$$\frac{d}{dt} Q_{10} = -(\lambda + \mu_1)Q_{10} + \lambda_1 Q_{00} + \mu_1 Q_{20} + \mu_2 Q_{11}$$

$$\tag{5.1} \begin{aligned} \frac{d}{dt} Q_{n_1,0} &= -(\lambda + \mu_1 + k_1(n_1 - 1))Q_{n_1,0} + \lambda_1 Q_{n_1-1,0} \\ &\quad + \mu_1 Q_{n_1+1,0} + \mu_2 Q_{n_1,1}; \qquad n_1 > 1 \\ \frac{d}{dt} Q_{0,n_2} &= -(\lambda + \mu_2 + k_2(n_2 - 1))Q_{0,n_2} + \lambda_2 Q_{0,n_2-1} \\ &\quad + \mu_1 Q_{1,n_2} + \mu_1 Q_{0,n_2} + \mu_2 Q_{0,n_2+1}; \\ &\qquad n_2 > 1 \end{aligned}$$

$$\frac{d}{dt} Q_{11} = -(\lambda + \mu_1 + \mu_2)Q_{11} + \lambda_1 Q_{01} + \lambda_2 Q_{10} + \mu_1 Q_{21} + \mu_2 Q_{12},$$

and

$$\tag{5.2} \begin{aligned} \frac{d}{dt} Q_{n_1,n_2} &= -(\lambda + \mu_1 + \mu_2 + k_i \,|\, w_1 - w_2 \,|)Q_{n_1,n_2} + \lambda_1 Q_{n_1-1,n_2} \\ &\quad + \lambda_2 Q_{n_1,n_2-1} + \mu_1 Q_{n_1+1,n_2} + \mu_2 Q_{n_1,n_2+1} \\ &\quad + k_1[n_1 - n_2 + 2]Q_{n_1+1,n_2-1} + k_2[n_2 - n_1 + 2]Q_{n_1-1,n_2+1}, \end{aligned}$$

when $n_1, n_2 > 1$ and where

$$\tag{5.3} \begin{aligned} [n_2 - n_1 + 2] &= \begin{cases} 0, & \text{if } (n_2 - n_1 + 2) \leqq 0 \\ n_2 - n_1 + 2, & \text{if } (n_2 - n_1 + 2) > 0, \end{cases} \\ k_i &= \begin{cases} k_1, & \text{if } w_1 > w_2 \\ 0, & \text{if } -w_1 = w_2 \\ k_2, & \text{if } w_1 < w_2. \end{cases} \end{aligned}$$

Using the notation of Section 3, we again write

$$g_r(\xi) = \sum_{r=0}^{\infty} \xi^n Q_{nr}, \tag{5.4}$$

organize our equations by n for fixed r, and multiply the n^{th} equation by ξ^n. Then, having obtained the equations for $\sum_{n=0}^{\infty} \xi^n d/dtQ_{nr}$, we multiply the r^{th} equation by ξ^r. The result is

$$\begin{aligned}\sum_{r=0}^{\infty} \xi^r \sum_{n=0}^{\infty} \xi^n (d/dt) Q_{nr} = ((\mu_1 + \mu_2)/\xi - \lambda_1 - \lambda_2 - \mu_1 - \mu_2 \\ + \xi(\lambda_1 + \lambda_2)) \sum_{r=0}^{\infty} \xi^r g_r(\xi) + (\mu_2 - \mu_2/\xi) g_0(\xi) \\ + (\mu_1 - \mu_1/\xi) \sum_{r=0}^{\infty} \xi^r g_r(0).\end{aligned} \tag{5.5}$$

For the steady state solution, the left hand side vanishes and, since $\lambda_1 + \lambda_2 = \lambda$, Eq. (5.5) is identical to Eq. (2.5), and the results are identical to those for "Tellers' Windows" and for "Tellers' Windows with Probabilistic Jockeying."

We note that the steady state solution for exponential jockeying to the shorter line is independent of the "jockeying rate," k, and independent of the manner in which customers select a line; i.e., it depends on λ alone, rather than on the λ_i.

The three cases which yield Eq. (5.5) (or its equivalent Eq. (2.5)) can be distinguished by the fact that they allow the existence of the states $(n_1, 0)$ and $(0, n_2)$ for all values of n_i. Thus, available service time at one of the servers is lost and cannot be recaptured. When we consider the effects of jockeying, we recognize that it can only improve system performance if it increases the utilization of the servers. If it does not, jockeying is merely a game played by people in the waiting room to amuse themselves. The only significant jockeying is that which prevents the server from being idle.

6. *Route Changing* (Figure 2 (c))*

Here we consider a system in which arriving customers choose a line initially, but jockeying in both directions is allowed. The probability of jockeying depends only on the size of the waiting line that a customer sees in his own line. We can consider this as a situation in which the customer waiting in line i has no information on the state of line j. Thus, it is equivalent to the problem of changing routes when one knows only that the route being used is congested. It is also equivalent to changing queues for theatre tickets at a different price when lines are on opposite sides of the theatre.

Customer Strategy

a) A unit of type i arrives and joins queue i $(i = 1, 2)$.

b) A "dissatisfied" waiting customer in line i will jockey to line j at a rate $k_i(n_i - 1)$, when $n_i > 0$; similarly for line j.

This differs from the previous case in that jockeying can take place in both directions, from longer line to shorter line and vice-versa. Such transitions occur when the lines are "screened" from each other; i.e., a customer in line i has no information on the status of line j. The probability of a customer shifting from line i to line j $(i \neq j, i, j = 1, 2)$, in an interval t is

$$(6.1)\qquad \begin{matrix} 0, & n_i = 0 \\ 1 - e^{-k_i(n_i-1)t}; & n_i \geqq 1, \end{matrix}$$

where n_i is the number of units at the i^{th} server.

In this case, the transition equations are defined by the following:

$$\frac{d}{dt} Q_{00} = -\lambda Q_{00} \qquad + \mu_1 Q_{10} + \lambda_2 Q_{01}$$

$$\frac{d}{dt} Q_{n_1,0} = -(\lambda + \mu_1 + k_1(n_1 - 1))Q_{n_1,0} + \lambda_1 Q_{n_1-1,0}$$
$$+ \mu Q_{n_1+1,0} + \mu_2 Q_{n_1,1}$$

$$(6.2)\qquad \frac{d}{dt} Q_{0,n_2} = -(\lambda + \mu_2 + k_2(n_2 - 1)Q_{0,n_2} \qquad + \lambda_2 Q_{0,n_2-1}$$
$$+ \mu_1 Q_{1,n_2} + \mu_2 Q_{0,n_2+^1}$$

$$\frac{d}{dt} Q_{n_1,n_2} = -(\lambda + \mu_1 + \mu_2 + k_1(n_1 - 1) + k_2(n_2 - 1))Q_{n_1,n_2}$$
$$+ \lambda_1 Q_{n_1-1,n_2} + \lambda_2 Q_{n_1,n_2-1} + \mu_1 Q_{n_1+1,n_2}$$
$$+ \mu_2 Q_{n_1,n_2+1} + n_1 k_1 Q_{n_1+1,n_2} + n_2 k_2 Q_{n_1,n_2+1}$$

and

$$\lambda = \lambda_1 + \lambda_2 .$$

Writing the generating function $g_r(\xi)$ as before,

$$(6.3)\qquad g_r(\xi) = \sum_{n=0}^{\infty} \xi^n Q_{nr}$$

and multiplying the n^{th} equation for fixed r by ξ^n as in Sections 3 and 4, we have:

$$(6.4)\qquad \begin{aligned} d/dt \sum_{n=0}^{\infty} \xi^n Q_{n0} = {} & (\mu_1/\xi - \lambda - \mu_1 + k_1 + \lambda_1 \xi) g_0(\xi) - k_1\xi(d/d\xi) g_0(\xi) \\ & + \mu_2 g_1(\xi) + (\mu_1 - \mu_1/\xi - k_1) g_0(0); \quad r = 0, \end{aligned}$$

and

$$\begin{aligned} d/dt \sum_{n=0}^{\infty} \xi^n Q_{nr} = {} & (\mu_1/\xi - \lambda - \mu_1 - \mu_2 + k_1 - (r-1)k_2 + \lambda_1\xi) g_r(\xi) \\ & - k_1\xi(d/d\xi) g_r(\xi) + \mu_2 g_{r+1}(\xi) + (\mu_1 - \mu_1/\xi - k_1) g_r(0) + k_1(d/d\xi) g_{r-1}(\xi) \\ & + rk_2\xi g_{r+1}(\xi) + (\lambda_2 - k_1/\xi) g_{r-1}(\xi) + (k_1/\xi) g_{r-1}(0); \quad r \geqq 1. \end{aligned}$$

Now, if we introduce $G(\xi)$ given by

$$(6.5)\qquad G(\xi) = \sum_{r=0}^{\infty} g_r(\xi),$$

we have

$$(6.6)\qquad \begin{aligned} G(1) &= 1 \\ G(0) &= \sum_{r=0}^{\infty} g_r(0) = \sum_{r=0}^{\infty} Q_{0r} = P_0^1. \end{aligned}$$

Performing the summation indicated in Eq. (6.5) and re-assembling terms, we have:

$$\begin{aligned}\sum_{r=0}^{\infty} d/dt \sum_{h=0}^{\infty} \xi^n Q_{nr} &= (\mu_1/\xi - k_1/\xi + k_1 - \lambda_1 - \mu_1 + \lambda_1\xi)G(\xi) \\ &+ k_1(1-\xi)(dG(\xi)/d\xi) + (\mu_1 - \mu_1/\xi - k_1 + k_1/\xi)G(0) \\ &+ k_2(\xi - 1)\sum_{r=1}^{\infty}(r-1)g_r(\xi).\end{aligned} \tag{6.7}$$

In the steady state, the left hand side of Eq. (6.7) vanishes. Now, we set

$$G(\xi) = \sum_{r=0}^{\infty} \xi^r a_r , \tag{6.8}$$

and we have:

$$\begin{aligned}0 = (\mu_1/\xi - k_1/\xi + k_1 - \lambda_1 - \mu_1 + \lambda_1\xi)\sum_{r=0}^{\infty} \xi^r a_r + k_1(1-\xi)\sum_{r=0}^{\infty} r\xi^{r-1}a_r \\ + a_0(\mu_1 - \mu_1/\xi - k_1 + k_1/\xi) + k_2(\xi-1)\sum_{r=1}^{\infty}(r-1)\xi^r a_r .\end{aligned} \tag{6.9}$$

We note that Eqs. (6.7) and (6.9) are independent of μ_2 and λ_2. The interaction depends only on k_2. We can obtain similar equations for the other queue by writing

$$H(\xi) = \sum_{n=0}^{\infty} h_n(\xi) = \sum_{n=0}^{\infty}\sum_{r=0}^{\infty} \xi^r Q_{nr} = \sum_{n=0}^{\infty} \xi^n a_n'. \tag{6.10}$$

Collecting terms in ξ^r in Eq. (6.9), we solve for a_r and hence for $g_r(\xi)$. We introduce the following notation:

$$\begin{aligned}\rho_1 &= \lambda_1/\mu_1 , & y_1 &= \lambda_1/k_1 \\ b_1 &= \mu_1/k_1 , & \beta_1 &= k_2/\mu_1\end{aligned} \tag{6.11}$$

The solutions are:

$$\begin{aligned}g_0(\xi) &= a_0 , \\ g_1(\xi) &= \frac{y_1}{b_1}\xi a_0 = \rho_1 \xi a_0 , \\ g_2(\xi) &= \frac{y_1^2\xi^2}{b_1(b_1+1)} a_0 , \\ g_3(\xi) &= \frac{y_1^3\xi^3}{b_1(b_1+1)(b_1+2)}\left[1 + \frac{\beta_1}{\rho_1}\right] a_0 , \\ g_4(\xi) &= \frac{y_1^4\xi^4}{b_1(b_1+1)(b_1+2)(b_1+3)}\left[1 + \frac{3\beta_1}{\rho_1} + 2\frac{\beta_1^2}{\rho_1^2}\right] a_0 ,\end{aligned} \tag{6.12}$$

and the r^{th} term for $r > 1$ is

$$\begin{aligned}g_r(\xi) = \frac{y_1^r\xi^r}{b_1(b_1+1)\cdots(b_1+r-1)} \\ \cdot\left(1 + \frac{\beta_1}{\rho_1}\right)\left(1 + \frac{2\beta_1}{\rho_1}\right)\cdots\left(1 + (r-2)\frac{\beta_1}{\rho_1}\right) a_0 .\end{aligned} \tag{6.12a}$$

A. The General Case

From Eqs. (6.12) we have, after a slight manipulation,

$$g_r(\xi) = \frac{y_1^2\xi^2(\xi x_1)^{r-2}}{b_1(b_1+1)\cdots(b_1+r-1)} \cdot \left(\frac{\rho_1}{\beta_1}+1\right)\left(\frac{\rho_1}{\beta_1}+2\right)\cdots\left(\frac{\rho_1}{\beta_1}+r-2\right)a_0, \tag{6.13}$$

where

$$x_1 = \frac{y_1\,\beta_1}{\rho_1} = \frac{k_2}{k_1}. \tag{6.14}$$

Then $G(\xi)$ becomes (from Eq. (6.5))

$$G(\xi) = a_0\left[1+\rho_1\,\xi+\frac{y_1^2\xi^2}{b_1(b_1+1)}\sum_{r=0}^{\infty} \cdot \frac{\left(\frac{\rho_1}{\beta_1}+1\right)\left(\frac{\rho_1}{\beta_1}+2\right)\cdots\left(\frac{\rho_1}{\beta_1}+r\right)}{(b_1+2)(b_1+3)\cdots(b_1+r-1)}(\xi x_1)^r\right], \tag{6.15}$$

$$G(\xi) = a_0\left[1+\rho_1\,\xi+\frac{\rho_1\,y_1\,\xi^2}{(b_1+1)}\,{}_2F_1\left(1,\frac{\rho_1}{\beta_1}+1, b_1+2; x_1\,\xi\right)\right],$$

where ${}_2F_1(\alpha, \beta, \gamma; x)$ is the Hypergeometric Function [12], defined by

$${}_2F_1(\alpha,\beta,\gamma;x) = \sum_{r=0}^{\infty} \frac{\Gamma(\alpha+r)\Gamma(\beta+r)\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma+r)}\frac{x^r}{r!}, \tag{6.16}$$

and $\Gamma(z)$ is the Gamma Function given by

$$\Gamma(z+1) = \int_0^{\infty} u^z e^{-u}\,du = z\Gamma(z). \tag{6.17}$$

When z is an integer,

$$\Gamma(z+1) = z!.$$

The Hypergeometric Function converges absolutely when $|x| < 1$ and diverges when $|x| > 1$. When $x = 1$, the function converges absolutely when $\gamma > \alpha + \beta$, and when $x = -1$, it is convergent when $\gamma > \alpha + \beta - 1$.

Using Eqs. (6.6), we find

$$P_0^1 = a_0 = [1+\rho_1+(\rho_1 y_1/(b_1+1))\,{}_2F_1(1, \rho_1/\beta_1+1, b_1+2; x_1)]^{-1}, \tag{6.18}$$

and the mean number in the first queue is

$$\bar{n}_1 = a_0[\rho_1 + (\rho_1 y_1/(b_1+1))[{}_2F_1(2, \rho_1/\beta_1+1, b_1+2; x_1) + {}_2F_1(1, \rho_1/\beta_1+1, b_1+2; x_1)]. \tag{6.19}$$

The results for the second queue are similar: the subscript 2 replacing the subscript 1 in Eqs. (6.18) and (6.19).

B. Jockeying out of one queue only $(k_1 > 0, k_2 = 0)$
When $k_2 = 0$, then $\beta_1 = 0$, and we have

$$g_r(\xi) = (\xi^r y_1^r / b_1(b_1 + 1) \cdots (b_1 + r - 1))a_0, \tag{6.20}$$

and

$$G(\xi) = \textstyle\sum_{r=0}^{\infty} (\xi^r y_1^r / b_1(b_1 + 1) \cdots (b_1 + r - 1))a_0 = a_0\, {}_1F_1(1, b_1 ; \xi y_1), \tag{6.21}$$

where ${}_1F_1(a, b; y)$ is the Confluent Hypergeometric Function [12], defined by

$${}_1F_1(a, b; y) = \textstyle\sum_{r=0}^{\infty} \dfrac{\Gamma(a + r)\Gamma(b)}{\Gamma(a)\Gamma(b + r)} \dfrac{y^r}{r!}. \tag{6.22}$$

Thus, when $k_2 = 0$ $(\beta_1 = 0)$, we have, from Eq. (6.6),

$$a_0 = [{}_1F_1(1, b_1 ; y_1)]^{-1} = P_0^1 \tag{6.23}$$

and

$$\bar{n}_1 = a_0 \frac{y_1}{b_1}\, {}_1F_1(2, b_1 + 1; y_1) = \rho_1 \frac{{}_1F_1(2, b_1 + 1; y_1)}{{}_1F_1(1, b_1 ; y_1)}. \tag{6.24}$$

We also have

$$\bar{w}_1 = \bar{n}_1 - 1 + P_0^1, \tag{6.25}$$

also expressable in terms of the same functions.

The results above are exactly those obtained by Palm [9, 10] and later by Anker and Gafarian [1], for "Reneging" in single server queues.

For the second queue we have, for $n > 1$

$$h_n(\xi) = \frac{y_2^n \xi^n}{b_2(b_2 + 1) \cdots (b_2 + n - 1)} \cdot \left(1 + \frac{\beta_2}{\rho_2}\right)\left(1 + \frac{2\beta_2}{\rho_2}\right) \cdots \left(1 + (n - 2) \frac{\beta_2}{\rho_2}\right) a_0',$$

which, in the limits $b_2 \to \infty$, $y_2 \to \infty$ becomes

$$h_n(\xi) = a_0' \rho_2^n \xi^n (1 + \beta_2/\rho_2)(1 + 2\beta_2/\rho_2) \cdots (1 + (n - 2)\beta_2/\rho_2). \tag{6.26}$$

Using Eq. (6.10), we obtain

$$\begin{aligned} H(\xi) = a_0'[1 + \rho_2\xi + \rho_2^2\xi^2 \textstyle\sum_{n=0}^{\infty} (1 + \beta_2/\rho_2) \\ \cdot (1 + 2\beta_2/\rho_2) \cdots (1 + n\beta_2/\rho_2)\rho_2^n \xi^n]. \end{aligned} \tag{6.27}$$

From the $H(\xi)$ equivalent to Eqs. (6.6) we have

$$P_0^2 = a_0' = \left[1 + \rho_2 + \rho_2^2 \textstyle\sum_{n=0}^{\infty} \beta_2^n \dfrac{\Gamma(\rho_2/\beta_2 + n + 1)}{\Gamma(\rho_2/\beta_2 + 1)}\right]^{-1}. \tag{6.28}$$

Following the same procedures, we have for the mean number at the second line

$$\bar{n}_2 = a_0^1 \left[\rho_2 + \rho_2^2 \textstyle\sum_{n=0}^{\infty} (n + 2)\beta^n \dfrac{\Gamma(\rho_2/\beta_2 + n + 1)}{\Gamma(\rho_2/\beta_2 + 1)}\right]. \tag{6.29}$$

We have not been able to express the summations in Eqs. (6.28) and (6.29) in closed form, but it is easy to show that when $\beta_2 = 0$ (i.e., $k_1 = 0$), a_0' and $\bar{n}_2$ are given by Eqs. (4.1) and (4.2), i.e., the same as for fixed lanes since $k_1 = k_2 = 0$.

In special cases (such as $\beta_2/\rho_2 \ll 1$) we can obtain solutions easily by using only the first terms of the expansion. However, we can still find P_0^2 using information already obtained and by using this value obtain values for $\bar{n}_2$ and $\bar{w}_2$ assuming a single server queue with Poisson input.

For the steady state solution, we know that all the input to the system is serviced. Thus, we have

$$\lambda = \mu_1(1 - P_0^1) + \mu_2(1 - P_0^2) \tag{6.30}$$

or

$$1 = (\mu_1/\lambda)(1 - P_0^1) + (\mu_2/\lambda)(1 - P_0^2),$$

$$1 = r_1 + r_2$$

where

$$r = (\mu_i/\lambda)(1 - P_0^i)$$

is the fraction of customers served by the i^{th} server. Thus, since we know (by using Eq. (6.23)),

$$r_1 = \frac{\mu_1}{\lambda}\left(1 - \frac{1}{{}_1F_1(1, b_1\,;\, y_1)}\right),$$

we find

$$P_0^2 = \frac{\mu_1 + \mu_2 - \lambda}{\mu_2} - \frac{\mu_1}{\mu_2}\frac{1}{{}_1F_1(1, b_1\,;\, y)}.$$

Since, for a single server queue, with no jockeying

$$P_0 = 1 - \rho$$

we can find a ρ_2' which yields the value of P_0^2 just obtained; and then, using Eqs. (4.2) and (4.3), calculate $\bar{n}_2$ and $\bar{w}_2$. Thus, we assume that the customers who leave queue 1 act as Poisson arrivals at queue 2. The assumption is a reasonable first approximation, but is not exact since the jockeying customers tend to bunch rather than enter queue 2 independently. In any event, the values of r_2 and P_0^2 are precise.

Table III lists a result for $\lambda_1 = \lambda_2 = 1$, $\mu_1 = \mu_2 = \frac{3}{2}$ for $k_1 = \frac{3}{2}$. By examining the values of r_1 and r_2, we can calculate the fraction of customers lost by queue 1 (and gained by queue 2). In the case illustrated, the original input is shared equally by the two queues; queue 1 handles only .3651 of the input and queue 2 handles .6349 of the input. Hence, a fraction 0.1349/0.5000 = .27 of the input to queue 1 is transferred to queue 2.

C. $\beta_1/\rho_1 = k_2/\lambda_1 \ll 1$, $k_1 > 0$

When β_1/ρ_1 is small, we can consider, for the first order approximation, only the terms in β_1/ρ_1 in Eq. (6.12). Then, we have

$$(6.31)\quad G(\xi) = a_0\left[{}_1F_1(1, b_1; \xi y_1) + \frac{\beta_1\,\xi^3 y_1^{\,3}}{\rho_1\, b_1(b_1+1)(b_1+2)}\,{}_1F_1(3, b_1+3; \xi y_1)\right].$$

Then, by Eq. (6.6), we have

$$(6.32)\quad P_0^{\,1} = a_0 = \left[{}_1F_1(1, b_1; y_1) + \frac{k_2\,\rho_1\, y_1}{k_1(b_1+1)(b_1+2)}\,{}_1F_1(3, b_1+3; y_1)\right]^{-1}.$$

The mean number at the first queue is

$$(6.33)\quad \bar{n}_1 = a_0\left[\rho_1\,{}_1F_1(2, b_1+1; y_1) + \frac{3k_2\,\rho_1\, y_1}{k_1(b_1+1)(b_1+2)}\,{}_1F_1(4, b_1+3; y_1)\right]$$

D. $\beta_1/\rho_1 = k_2/\lambda_1 = 1,\ k_1 > 0$

When β_1/ρ_1 is equal to one, then

$$(6.34)\qquad g_r(\xi) = a_0(\xi^r y_1^{\,r}/b_1(b_1+1)\cdots(b_1+r-1))(r-1)!,$$

and we have

$$(6.35)\qquad G(\xi) = a_0\left[1 + \frac{\xi y_1}{b_1}\sum_{r=0}^{\infty}\frac{\Gamma(b_1+1)\Gamma(r+1)}{\Gamma(b+r+1)\Gamma(1)}\,(\xi y_1)^r\right].$$

Now, we know [7]

$$(6.36)\quad \sum_{r=0}^{\infty}\frac{\Gamma(a+r)\Gamma(b)}{\Gamma(a)\Gamma(b+r)}\,x^r = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1\frac{t^{a-1}(1-t)^{b-a-1}}{(1-xt)}\,dt,$$

when $b > a > 0$. Substituting Eq. (6.36) in Eq. (6.35) and setting $\xi = 1$, we obtain ($a = 1,\ b = b_1 + 1$),

$$(6.37)\qquad a_0 = \left[1 + \rho_1\frac{\Gamma(b_1+1)}{\Gamma(1)\Gamma(b_1)}\int_0^1\frac{(1-t)^{b_1-1}}{(1-y_1\,t)}\,dt\right]^{-1},$$

and for the mean number at the first server, we have

$$(6.38)\qquad \bar{n}_1 = a_0\left[\rho_1\frac{\Gamma(b_1+1)}{\Gamma(2)\Gamma(b_1-1)}\int_0^1\frac{t(1-t)^{b_1-2}}{(1-y_1\,t)}\,dt\right].$$

If we multiply and divide the r^{th} term in the sum on the left side of Eq. (6.63) by $(r!)$, then we have

$$(6.36a)\qquad \sum_{r=0}^{\infty}\frac{\Gamma(r+1)\Gamma(a+r)\Gamma(b)}{\Gamma(a)\Gamma(b+r)}\,\frac{x^r}{r!} = {}_2F_1(1, a, b; x).$$

Thus, Eq. (6.37) can be written as

$$(6.37a)\qquad a_0 = [1 + \rho_1\,{}_2F_1(1, 1, b_1+1; y_1)]^{-1},$$

and Eq. (6.38) becomes

$$(6.38a)\qquad \bar{n}_1 = a_0\rho_1\,{}_2F_1(1, 2, b_1+1; y_1).$$

E. $\beta_1/\rho_1 = k_2/\lambda_1 \gg 1$, $k_1 > 0$

When β_1/ρ_1 is very large, we can omit all terms in $(\beta_1/\rho_1)^n$, except the largest in the value of $g_r(\xi)$. Then, we have

$$G(\xi) = a_0\left[1 + \frac{y_1\,\xi}{b_1} + \frac{y_1^2\,\xi^2}{b_1(b_1+1)}\left(\sum_{r=0}^{\infty} \frac{r!\,\Gamma(b_1+2)}{\Gamma(b_1+r+2)}\,\xi^r\left(\frac{k_2}{k_1}\right)^r\right)\right]. \tag{6.39}$$

Now, solving for $G(1)$, and writing as before $k_2/k_1 = x_1$, we have

$$G(1) = a_0\left[1 + \rho_1 + \frac{\rho_1\,y_1}{b_1+1}\left(\sum_{r=0}^{\infty} \frac{\Gamma(r+1)\Gamma(b_1+2)}{\Gamma(b_1+r+2)\Gamma(1)}\,x_1^r\right)\right]. \tag{6.40}$$

Using Eq. (6.6), we find

$$P_0^1 = a_0 = \left[1 + \rho_1 + \frac{\rho_1\,y_1}{b_1+1}\left(\frac{\Gamma(b_1+2)}{\Gamma(1)\Gamma(b_1+1)}\int_0^1 \frac{(1-t)^{b_1}}{(1-x_1\,t)}\,dt\right)\right]^{-1}, \tag{6.41}$$

and the mean number at the first server is

$$\begin{aligned}\bar{n}_1 = a_0\Bigg[\rho_1 &+ \frac{\rho_1\,y_1}{b_1+1}\left(\frac{\Gamma(b_1+2)}{\Gamma(1)\Gamma(b_1+1)}\right)\\ &\cdot\left(b_1\int_0^1 \frac{t(1-t)^{b_1}}{(1-x_1\,t)^2}\,dt + \int_0^1 \frac{(1-t)^{b_1}}{(1-x_1\,t)}\,dt\right)\Bigg].\end{aligned} \tag{6.42}$$

Some Numerical Results

Some sample results for the various forms of jockeying are included in Tables I, II and III. Table I compares results for Tellers' Windows with Instantaneous Jockeying (which also hold for Maitre d' Hotel systems) for various values of ρ_i and π_i, and $\rho = \lambda/\sum\mu_i = \frac{2}{3}$. In the first four cases, $\pi_1 = \pi_2 = \frac{1}{2}$, and we note that as μ_1 increases Q_{00} decreases, while $\bar{n}$ increases. The unbalance of the two servers has the effect of causing the slower server to act as a trapping state; i.e., the unwary customer who is served by the slower server is trapped. We also note that the slower server has an r_i (fraction of customers served by line i) which is larger than $\mu_i/\sum\mu_i$. Thus, it is idle far less than the faster server (for example, when $\mu_1 = \frac{11}{4}$, $\mu_2 = \frac{1}{4}$, $P_0^1 = .355$, $P_0^2 = .097$), for while the faster server clears a line of some size the slower server still holds the same customer. When arrivals reappear, both servers become busy.

In cases e) through h) of Table I, the effects of varying π_i are demonstrated. Cases e) and f) represent the selection of the faster line and the slower line, respectively, by a customer arriving when both lines are of the same length. The proper selection (i.e., $\pi_i = 1$) increases the portion of customers served by the faster line (which is desirable) and reduces the size of the waiting line as compared with the other alternative. Case g), in which the choice is proportional to the relative service rates, yields an intermediate result. Case h), in which customers arriving when both lines are of equal length always choose line 1 (which is no faster than line 2), is interesting in that the resulting total number waiting $\bar{n}$ is identical to case a). When the service rates are homogeneous, the

TABLE I

Results for "Maitre d'Hotel" and "Tellers Windows with Jockeying"

Measure	a	b	c	d	e	f	g	h
	$\pi_1 = \pi_2 = \frac{1}{2}$				$\pi_1 = 1$, $\pi_2 = 0$	$\pi_1 = 0$, $\pi_2 = 1$	$\pi_1 = \frac{2}{3}$, $\pi_2 = \frac{1}{3}$	$\pi_1 = 1$, $\pi_2 = 0$
λ	2	2	2	2	2	2	2	2
μ_1	$\frac{3}{2}$	2	$\frac{5}{2}$	$\frac{11}{4}$	2	2	2	$\frac{3}{2}$
μ_2	$\frac{3}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	1	1	1	$\frac{3}{2}$
ρ	.6667	.6667	.6667	.6667	.6667	.6667	.6667	.6667
Q_{00}	.2000	.1818	.12195	.07097	.2059	.1628	.1892	.2000
P_0^1	.3333	.3636	.36585	.35490	.3236	.3954	.3513	.2761
P_0^2	.3333	.2727	.17073	.09696	.3529	.2093	.2972	.3905
$\bar{n}$	2.4000	2.4545	2.6352	2.7871	2.3470	2.4864	2.3981	2.400
$\bar{w}$	1.0667	1.0908	1.1718	1.2388	1.0235	1.0911	1.0466	1.0667
r_1	.5000	.6364	.7927	.8871	.6764	.6046	.6487	.5429
r_2	.5000	.3636	.2073	.1129	.3236	.3954	.3513	.4571
$\bar{n}_1$*	1.2000	1.1672	1.1888	1.2231	1.2430	1.0799	1.1850	1.3042
$\bar{n}_2$*	1.2000	1.2873	1.4464	1.5640	1.1040	1.4065	1.2131	1.0958
$\bar{w}_1$*	.5333	.5308	.5547	.5780	.5666	.4753	.5363	.5803
$\bar{w}_2$*	.5333	.5600	.6171	.6611	.4569	.6158	.5103	.4864

* Measures marked with an asterisk hold only for "Tellers' Windows, Jockeying Guaranteed".

TABLE II

Results for "Tellers' Windows, No Jockeying"

Measure	a	b
λ	2	2
μ_1	$\mu_1, \mu_2 > 0$	3
μ_2	$\mu_1 + \mu_2 = 3$	0
ρ	.6667	.6667
Q_{00}	.1111	.3333
P_0^1	.3333	.3333
P_0^2	.3333	—
$\bar{n}_1$	2.0000	2.0000
$\bar{n}_2$	2.0000	—
$\bar{n}$	4.0000	2.0000
$\bar{w}_1$	1.3333	1.3333
$\bar{w}_2$	1.3333	—
$\bar{w}$	2.6667	1.3333
$\bar{r}_1$	$\mu_1/(\mu_1 + \mu_2)$	1.0000
$\bar{r}_2$	$\mu_2/(\mu_1 + \mu_2)$	0.00

customer's choice has no influence on the total number waiting, although it influences the individual lines and the total number served at each queue.

Table II lists the measures of interest for "Tellers' Windows, No Jockeying." Here, the results (case a)) are independent of the values of μ_i and depend only on the sum of the service rates, if the μ_i are greater than zero. If one of the

TABLE III

Results for "Route Changing"

$\lambda_1 = \lambda_2 = 1 \quad \mu_1 = \mu_2 = \frac{3}{2} \quad \rho_1 = \rho_2 = \frac{2}{3}$

Measure	a $\beta_1 = \beta_2 = 0$	b $\beta_1 = 0$	c $\beta_1/\rho_1 \ll 1$	d $\beta_1/\rho_1 = 1$
k_1	0	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$
k_2	0	0	$\frac{1}{10}$	1
β_1	0	0	$\frac{1}{15}$	$\frac{2}{3}$
β_2	0	1	1	1
P_0^1	.3333	.5133	.5111	.4765
P_0^2	.3333	.1536	.1556	.1902
$\bar{n}_1$	2.000	.6667	.6901	.7519
$\bar{n}_2$	2.000	5.5105		
$\bar{w}_1$	1.333	.1800	.2012	.2284
$\bar{w}_2$	1.333	4.6641		
r_1	.5000	.3651	.3667	.3927
r_2	.5000	.6349	.6333	.6073

service rates is zero, we have the usual single server queue with the results given in case b). In case a), we find that each server handles a share of traffic proportional to $\mu_i/(\mu_1 + \mu_2)$. The policy of joining the shorter line serves to make $\bar{n}_1 = \bar{n}_2$ independent of the precise values of the service rates. The results of the strategy of joining the shorter line are not as good (in terms of $\bar{n}$ and $\bar{w}$) as are obtained with jockeying, as can be seen by comparing case a) of Table II with the results given in Table I.

Some results for "Route Changing" are given in Table III. Case a) which represents two independent single server queues without jockeying is included for comparison. In case b), $\beta_1 = 0$ (i.e., $k_2 = 0, k_1 > 0$), the effects of jockeying from line 1 to line 2 with no counter-jockeying can be seen. The third column, case c), indicates the effects of a small counter-current; i.e., line 1 carries a slightly larger share of the traffic. Case d), in which the counter-current is large, results in a larger share of traffic returning to the first server.

Acknowledgment

The author thanks James E. Enstrom of Harvey Mudd College, who reviewed the original manuscript, for bringing to his attention a number of inaccuracies and for suggesting the more satisfying forms for Eq. (3.9) and Eq. (6.12).

Glossary of Terms

$\lambda_i = \lambda\pi_i =$ Arrival rate of i^{th} server.

$\lambda = \sum_i \lambda_i =$ Total arrival rate for the system.

$\mu_i =$ Service rate offered by i^{th} server.

$\rho_i = \lambda_i/\mu_i$

$\rho = \lambda/\sum \mu_i$

$Q_{nm} =$ Probability that n units are at the first queue and m units are at the second queue.

$P_0^1 = \sum_{m=0}^{\infty} Q_{0m}$ = Probability first server is idle.
$P_0^2 = \sum_{n=0}^{\infty} Q_{n0}$ = Probability second server is idle.
$\bar{n}_i$ = Number of units at i^{th} server (including customer being served).
$\bar{w}_i$ = Number of units waiting for i^{th} server.
r_i = Fraction of customers serviced by i^{th} server.

References

1. ANCKER, C. J., JR. AND GAFARIAN, A. V., "Some Queuing Problems with Balking and Reneging-I," *Operations Research*, 11, 88 (1963), Part II, *Operations Research*, 11, 928 (1963).
2. BARRER, D. Y., "Queuing With Impatient Customers and Ordered Service," *Operations Research*, 5, 650 (1957).
3. ——, "Queuing With Impatient Customers and Indifferent Clerks," *Operations Research*, 5, 644 (1957).
4. CAMP, G. D., "Bounding the Solutions of Practical Queuing Problems by Analytic Methods" in *Operations Research for Management, Vol. II*, McCloskey, J. F. and Coppinger, J. M. (eds.), Johns Hopkins University Press, Baltimore, pp. 307–339 (1956).
5. GLAZER, H., "Jockeying in Queues," *Operations Research*, 6, 145 (1958).
6. GUMBEL, H., "Waiting Lines With Heterogeneous Servers," *Operations Research*, 8, 504 (1960).
7. JOLLEY, L. B. W., *Summation of Series*, Dover, New York (1961) Series 1014, pp. 188–189.
8. KRISHNAMOORTHI, B., "On a Poisson Queue With Two Heterogeneous Servers," *Operations Research*, 11, 321 (1963).
9. PALM, C., "Inhomogeneous Telephone Traffic in Full Availability Groups," *Ericsson Technics*, 5, 3 (1937).
10. ——, "Research on Telephone Traffic Carried by Full Availability Group," *Tele*, No. 1, 1 (1957) (English Edition).
11. SAATY, T. L., *Elements of Queuing Theory*, McGraw-Hill, New York, 1961.
12. SNEDDON, I. N., *Special Functions of Mathematical Physics and Chemistry*, Oliver and Boyd, London, 1961, Chap. II, pp. 19–49.

See also for Tables:

(a) SLATER, L. J., *Confluent Hypergeometric Functions*, Cambridge University Press, Cambridge, 1960; 8 place tables, ${}_1F_1(\alpha, \beta; X)$ $\alpha = -1.0(.01)1.0$, $\beta = 0.1(0.1)1.0$, $X = 0.1(.01)10.0$.
(b) British Association for the Advancement of Science, Mathematical Tables Committee, Section A, Oxford 1926 Leeds 1927; 5 to 6 place tables ${}_1F_1(\alpha, \beta; X)$, $\alpha = -4(.5)4$, $\beta = -1.5(.5)4.0$, $0 < X < 8$. (Jahnke, E., and Emde, F., Tables of Functions, Dover, New York, p. 275, (1945) gives curves for above range).

13. TANNER, J. C., "A Problem of Interference Between Two Queues," *Biometrika*, 40, 58 (1953).