Memorandum COSOR 90-20

Analysis of the asymmetric shortest queue problem with threshold jockeying

I.J.B.F. Adan
J. Wessels
W.H.M. Zijm
ANALYSIS OF THE ASYMMETRIC SHORTEST QUEUE PROBLEM
WITH THRESHOLD JOCKEYING

I.J.B.F. Adan *
J. Wessels *
W.H.M. Zijm *

University of Technology, Eindhoven

Abstract. In this paper we study a system consisting of two parallel servers with possibly different service rates. Jobs arrive according to a Poisson stream and generate an exponentially distributed workload. On arrival a job joins the shortest queue and in case both queues have equal lengths, he joins the first queue with probability $1 - a$ and the second one with probability $a$, where $a$ is an arbitrary number between 0 and 1. If the difference between the lengths of both queues reaches some threshold value $T$, then one job switches from the longest to the shortest queue. It is shown that the equilibrium probabilities of the queue lengths satisfy a product form for states where the longer queue exceeds the threshold value $T$. Furthermore, it is shown that for a sensible partitioning of the state space the matrix-geometric approach is essentially equivalent with our approach.

Keywords: difference equation, jockeying, matrix-geometric solution, product form, queues in parallel, stationary queue length distribution, shortest queue problem.

1. Introduction

Consider a queueing system consisting of two parallel servers with possibly different service rates. Jobs arrive according to a Poisson stream and generate an exponentially distributed workload. On arrival a job joins the shortest queue and in case both queues have equal lengths, he joins the first queue with probability $1 - a$ and the second one with probability $a$, where $a$ is an arbitrary number between 0 and 1. If the difference between the lengths of both queues reaches some threshold value $T$, then one job switches from the longest to the shortest queue. This model has already been studied by Gertsbakh [5]. He uses the matrix-geometric methodology developed by Neuts [8] to obtain a numerical solution. Koenigsberg [6] analyzes various
jockeying models. However, some of his results are in error, see [7]. So far, for the present model no exact analytic results are known.

In [1] we showed for the symmetric shortest queue problem without jockeying that the equilibrium probabilities of the queue lengths asymptotically satisfy a product form for states far away from the boundaries of the state space. That asymptotic product form does not satisfy the boundary equations. A compensation method is used to improve the initial asymptotic solution. That method consists of consistently adding on terms so as to compensate alternatingly for the error on one of the boundaries of the state space. This approach eventually leads to the exact determination of the equilibrium probabilities. In [2] we showed that the compensation idea also works for the asymmetric shortest queue problem. However, in that case the analysis is essentially more complicated. It seems natural to treat the present jockeying problem accordingly, i.e. by first investigating the asymptotic behaviour of the equilibrium probabilities and then investigating how the compensation method works out here. It is shown, however, that the equilibrium probabilities exactly satisfy a product form for states where the longer queue exceeds the threshold value \( T \). So no compensation at the boundaries is needed. That result is mainly due to the special boundary structure.

A simple transformation of the state variables makes the state space finite in one dimension, which suggests a matrix-geometric approach. Gertsbakh [5] studied the present problem by using the matrix-geometric approach. However, his solution has not much affinity with our approach and did not lead to the results discussed above. This is mainly due to the way Gertsbakh partitions the state space. In fact, our approach suggests another partitioning. Indeed, for this partitioning the matrix-geometric approach essentially leads to the same results as obtained by our approach.

The paper is organized as follows. In Section 2 we present the equilibrium equations. Section 3 investigates the asymptotic behaviour of the equilibrium probabilities of the queue lengths by means of a numerical experiment. Section 4 analyzes the jockeying problem. In that section it is proved that the equilibrium probabilities of the queue lengths satisfy a product form for states where the longer queue exceeds the threshold value \( T \). Section 5 investigates the relationship with the matrix-geometric solution. The final section is devoted to comments and extensions.

2. Equilibrium Equations

For simplicity of notation the servers have service rates \( \gamma_1 \) and \( \gamma_2 \) respectively with \( \gamma_1 > 0, \gamma_2 > 0 \) and \( \gamma_1 + \gamma_2 = 2 \), the Poisson arrival process has a rate \( 2p \) with \( 0 < p < 1 \) and on arrival each job generates an exponentially distributed workload with unit mean. The parameter \( a \) denotes the probability that an arriving job is sent to the second queue in case both queues have
equal lengths. This parallel queue system can be represented by a continuous time Markov process, with a state space consisting of the pairs \((m, n)\), \(m, n = 0, 1, \ldots\) where \(m\) and \(n\) are the lengths of the two queues. Because of the threshold jockeying, the state space is restricted to those pairs \((m, n)\) for which \(\mid m - n \mid \leq T\). The transition rates are illustrated in Figure 1.

\[\begin{align*}
\text{Figure 1: } (m, n) \text{ transition rate diagram}
\end{align*}\]

\([p_{m,n}]\) is the equilibrium distribution of the lengths of the two queues. The state space can be partitioned into levels \(l\), where each level \(l\) consists of those states for which the longest queue has length \(l\). The equilibrium or balance equations at the levels \(l > T\) state that:

\[
\begin{align*}
  p_{l-T,l} 2(p+1) &= p_{l-T+1,l} y_1 \\
  p_{l-T+1,l} 2(p+1) &= p_{l-T,l} 2p + p_{l-T+1,l+1} 2 + p_{l-T+2,l} y_1 \\
  p_{l-k,l} 2(p+1) &= p_{l-k-1,l} 2p + p_{l-k,l+1} y_2 + p_{l-k+1,l} y_1 \\
  p_{l-1,l} 2(p+1) &= p_{l-2,l} 2p + p_{l-1,l+1} y_2 + p_{l,l} y_1 + p_{l-1,l-1} 2a \rho \\
  p_{l,l} 2(p+1) &= p_{l-1,l} 2p + p_{l,l+1} y_2 + p_{l,l-1} 2p + p_{l+1,l} y_1 \\
  p_{l,l-1} 2(p+1) &= p_{l,l-2} 2p + p_{l+1,l-1} y_1 + p_{l,l} y_2 + p_{l-1,l-1} 2(1-a) \rho \\
  p_{l,l-k} 2(p+1) &= p_{l,l-k-1} 2p + p_{l+1,l-k} y_1 + p_{l,l-k+1} y_2 \\
  p_{l,l-T+1} 2(p+1) &= p_{l,l-T} 2p + p_{l+1,l-T+1} 2 + p_{l,l-T+2} y_2 \\
  p_{l,l-T} 2(p+1) &= p_{l,l-T+1} y_2 \\
\end{align*}
\]

The balance equations at the levels \(l \leq T\) are different at the vertical and horizontal boundary,
but their exact form is not important for the analysis that follows. In the next two sections we analyze the probabilities $p_{m,n}$ at the levels $l > T$. Particularly, we investigate whether the probabilities $p_{m,n}$ asymptotically satisfy a product form.

3. Experimental Analysis of the Equilibrium Probabilities $p_{m,n}$

In [1,2] we showed for the shortest queue problem without jockeying that the equilibrium probabilities of the queue lengths asymptotically satisfy a product form for states far away from the boundaries. Below we verify, by a numerical experiment, whether also for the present jockeying problem the probabilities $p_{m,n}$ asymptotically satisfy a product form. In the Figures 2 and 3 we list the ratios of the probabilities $p_{m,n}$ in various directions for the special case $\rho = 0.5$, $\gamma_1 = 1.2$, $a = 0.4$ and $T = 5$. Actually, we calculated approximations for the probabilities $p_{m,n}$ by solving a finite capacity system exactly, i.e. by means of a Markov chain analysis. For the example we calculated the probabilities $p_{m,n}$ for a system where each queue has a maximum capacity of 15 jobs, which approximates well the infinite capacity system in case $\rho = 0.5$ and $\gamma_1 = 1.2$. In Figure 2 we list the ratios of the probabilities $p_{m,n}$ in the direction of the diagonal.

$$\begin{array}{cccccccc}
\uparrow & 8 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
n & 7 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
6 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
5 & 0.13 & 0.22 & 0.24 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
4 & 0.18 & 0.23 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
3 & 0.18 & 0.23 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
2 & 0.19 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\
1 & 0.26 & 0.28 & 0.26 & 0.25 & 0.24 & 0.24 & 0.25 \\
0 & 0.44 & 0.31 & 0.22 & 0.21 & 0.21 & 0.18 \\
m \rightarrow & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8
\end{array}$$

Figure 2: $p_{m+1,n+1}/p_{m,n}$ for $\rho = 0.5$, $\gamma_1 = 1.2$, $a = 0.4$ and $T = 5$

We see that for all $m$ and $n$ for which $m > T(=5)$ or $n > T$,

$$p_{m+1,n+1} = 0.25 p_{m,n} .$$

To investigate the behaviour on a fixed level we list in Figure 3 the ratios of the probabilities $p_{m,n}$ in the upper wedge $m \leq n$ in the horizontal direction.
We see that for all $n > T$,

$$\begin{align*}
p_{m-1, n} &= 1.17 p_{m, n} \quad \text{if } m = n, \\
&= 0.5 p_{m, n} \quad \text{if } n-T+1 < m < n, \\
&= 0.4 p_{m, n} \quad \text{if } m = n-T+1.
\end{align*} \quad (11)$$

Repeatedly applying the relations (10) and (11) yields for all $n > T$,

$$\begin{align*}
p_{m, n} &= 0.25^{n-T} p_{T, T} \quad \text{if } m = n, \\
&= 0.25^{n-T} 0.5^{n-m-1} 1.17 p_{T, T} \quad \text{if } n-T < m < n, \\
&= 0.25^{n-T} 0.4 0.5^{T-2} 1.17 p_{T, T} \quad \text{if } m = n-T,
\end{align*}$$

or in terms of the state variables $k = n - m$ and $l = n$, we obtain for all $l > T$,

$$\begin{align*}
p_{l-k, l} &= 0.25^{l-T} 0.31^{k-1} 1.23 p_{T, T} \quad \text{if } k = 0, \\
&= 0.25^{l-T} 0.5^{k-1} 1.17 p_{T, T} \quad \text{if } 0 < k < T, \\
&= 0.25^{l-T} 0.4 0.5^{T-2} 1.17 p_{T, T} \quad \text{if } k = T.
\end{align*} \quad (12)$$

Accordingly, by evaluating the ratios of the probabilities in the lower wedge $m > n$ in the vertical direction, we obtain for all $l > T$,

$$\begin{align*}
p_{l, l-k} &= 0.25^{l-T} 0.31^{k-1} 1.23 p_{T, T} \quad \text{if } 0 < k < T, \\
&= 0.25^{l-T} 0.27 0.31^{T-2} 1.23 p_{T, T} \quad \text{if } k = T.
\end{align*} \quad (13)$$
The experiment reveals that the equilibrium probabilities satisfy a constant form on all levels higher than the threshold value $T$. That can be explained as follows. Define

$$Q = \text{The set of states at the levels } l > T, \text{ including state } (T, T),$$

and denote by $Q^*$ its complement with respect to the state space. Then consider the process restricted to the set $Q$, i.e. the occupation times of the process in the set $Q^*$ are not counted. Since the unrestricted process is irreducible, it follows that the $Q$ restricted process is also irreducible. Furthermore, since $Q^*$ is finite, it follows that the unrestricted process is ergodic if and only if the $Q$ restricted process is ergodic, and in that case the probabilities $p_{m,n}$ and $q_{m,n}$ of the $Q$ restricted process are related according to

$$p_{m,n} = q_{m,n} \left(1 - P(Q^*)\right) \quad \text{for } (m, n) \in Q,$$

where $P(Q^*)$ is the probability that the unrestricted process is in the set $Q^*$. For all states $(m, n), (m', n') \in Q$ holds that the transition rate of the $Q$ restricted process from $(m, n)$ to $(m', n')$ is identical to the one of the unrestricted process, except when it is possible to go from $(m, n)$ to $(m', n')$ via a path in the set $Q^*$. That is only the case for the transition rates from the states $(T-k+1, T+1)$ and $(T+1, T-k+1)$, $1 \leq k \leq T$, to state $(T, T)$. These transitions to $(T, T)$ exclusively come from excursions of the unrestricted process in the set $Q^*$. Any excursion of the unrestricted process to the set $Q^*$ always ends with a visit at state $(T, T)$ (and no other state in $Q$). Hence, the transition rate of the $Q$ restricted process from state $(T-k+1, T+1)$ to $(T, T)$ is $\gamma_2$ for $1 \leq k < T$ and 2 for $k = T$. Accordingly, the transition rate of the $Q$ restricted process from state $(T+1, T-k+1)$ to $(T, T)$ is $\gamma_1$ for $1 \leq k < T$ and 2 for $k = T$ (see Figure 4).

![Figure 4: The transition rates to state (T, T) for the Q restricted process](image)

For all other states $(m, n), (m', n') \in Q$, the transition rate of the $Q$ restricted process from $(m, n)$ to $(m', n')$ is identical to the one of the unrestricted process. Hence, the balance equations of the $Q$ restricted process are identical to the equations for the unrestricted process in the set $Q$ (i.e. the equations (1), (2), ..., (9)). The only exception is the balance equation in state $(T, T)$,
but that equation can be omitted, since exactly one equation is redundant. On all levels the balance equations of the $Q$ restricted process have the same form. Therefore, the probabilities $q_{m,n}$ will satisfy a constant form and by relation (14), that explains the constant form of the probabilities $p_{m,n}$ on all levels higher than the threshold value $T$. In the next section we try to find an explicit characterization of the product forms for $p_{m,n}$.

4. Analysis of the $Q$ Restricted Process

We first derive product forms for the probabilities $q_{m,n}$ of the $Q$ restricted process. Then the desired product forms for the probabilities $p_{m,n}$ can be obtained by virtue of relation (14) in Section 3. The experiment in Section 3 yields for a special case the product forms for $p_{m,n}$, or rather, for $q_{m,n}$ of the $Q$ restricted process. Below we try to characterize the product forms for $q_{m,n}$ in terms of $\rho$, $\gamma_1$ and $a$. That is, we seek parameters $\alpha$, $\beta$, $\delta$, $\phi_1$, $\phi_2$, $\psi_1$ and $\psi_2$, depending on $\rho$, $\gamma_1$ and $a$, such that for all $l > T$ (cf. (12) and (13)),

$$q_{i-k,l} = \alpha^{l-T} \phi_{T,T}, \quad \text{if } k = 0,$$
$$= \alpha^{l-T} \beta^{k-1} \phi_1 q_{T,T}, \quad \text{if } 0 < k < T,$$
$$= \alpha^{l-T} \psi_1 \beta^{T-2} \phi_1 q_{T,T}, \quad \text{if } k = T,$$

(15)

and

$$q_{i,l-k} = \alpha^{l-T} \delta^{k-1} \phi_2 q_{T,T}, \quad \text{if } 0 < k < T,$$
$$= \alpha^{l-T} \psi_2 \delta^{T-2} \phi_2 q_{T,T}, \quad \text{if } k = T.$$

As stated before, the equilibrium probabilities $q_{m,n}$ of the $Q$ restricted process are the unique solution of the equations (1), ..., (9), together with the normalization equation. Based on these equations, we try to derive explicit expressions for the parameters in the product forms (15). Specifically, we apply the general balance principle

Rate out of the set of states $A = \text{Rate into the set of states } A$,

to suitably chosen subsets $A$ of $Q$ in order to find the desired expressions for the parameters in the product forms (15). Afterwards we have to verify whether the balance equations are indeed satisfied for these expressions of the parameters. First, let us define for all $i$ the following subsets of $Q$:

$$A_i = \{(m,n) \in Q, n + m = i\},$$
$$B_i = \{(m,n) \in Q, n - m = i\}.$$

Then the parameter $\alpha$ can be obtained by applying the balance principle to the set $A = A_{M+1} + A_{M+2} + ...$ with $M \geq 2T$. Notice that it is only possible to leave set $A$ via state
\((m, n) \in A_{M+1}\) with rate 2. So the total rate out of set \(A\) equals \(P(A_{M+1})\) 2. Further, it is only possible to enter set \(A\) via state \((m, n) \in A_M\) with rate \(2p\). So the total rate into set \(A\) equals \(P(A_M)\) 2p. Equating the total rate out and the total rate into set \(A\) yields for all \(M \geq 2T\), that
\[P(A_{M+1}) = P(A_M) 2p.\]
Combining that equality for \(M = 2N\) and \(M = 2N+1\) leads to
\[P(A_{2N+2}) = P(A_{2N}) \rho^2.\] (16)
On the other hand, inserting the product forms (15) yields for \(N > T\),
\[P(A_{2N}) = \alpha^{N-T} \left[ 1 + \frac{1 - \beta^{T-1}}{1 - \beta} \Phi_1 + \Phi_2 \right],\]
so \(P(A_{2N})\) is of the form \(K \alpha^N\) with \(K\) is independent of \(N\). Then together with (16), we obtain that
\[\alpha = \rho^2.\] (17)
To determine \(\psi_1\) we apply the balance principle to the set \(B_T\). The rate out of set \(B_T\) equals \(P(B_T) 2(\rho + 1)\). Further, it is only possible to enter set \(B_T\) via state \((m, n) \in B_{T-1}\) with rate \(\gamma_1\). So the rate into set \(B_T\) equals \(P(B_{T-1}) \gamma_1\). Equating the rate out and the rate into set \(B_T\) yields
\[P(B_T) 2(\rho + 1) = P(B_{T-1}) \gamma_1.\]
Since by (15), \(P(B_T) = \psi_1 P(B_{T-1})\), we obtain that
\[\psi_1 = \frac{\gamma_1}{2(\rho + 1)},\] (18)
and by applying the balance principle to the set \(B_{T-1}\) we find
\[\psi_2 = \frac{\gamma_2}{2(\rho + 1)}.\]
The parameter \(\beta\) can be found by applying the balance principle to the set \(B_{T-1} + B_T\). That leads to the equality
\[P(B_{T-1}) (2\rho + \gamma_2) + q_{1,T+1} 2 = P(B_{T-2}) \gamma_1.\]
Substituting the product forms (15) into that equality yields for the left hand side,
\[\frac{\alpha}{1 - \alpha} \beta^{T-2} \Phi_1 q_{T,T} (2\rho + \gamma_2 + (1 - \alpha)\psi_1);\] (19)
and for the right hand side,
\[\frac{\alpha}{1 - \alpha} \beta^{T-3} \Phi_1 q_{T,T} \gamma_1.\] (20)
Inserting the expressions (17) and (18) into (19) and (20), and then equating (19) and (20) leads to

\[
\beta = \frac{\gamma_1}{2 + \rho \gamma_2}. \tag{21}
\]

Accordingly, the parameter $\delta$ is found by applying the balance principle to the set $B_{-T+1} + B_{-T}$. That leads to

\[
\delta = \frac{\gamma_2}{2 + \rho \gamma_1}. \tag{22}
\]

Finally, \( \phi_1 \) is obtained by applying the balance principle to the set \( B_1 + B_2 + \ldots + B_T \), which yields the equality

\[
P(B_1)(2\rho + \gamma_2) + q_{1,T+1} 2 + (q_{2,T+1} + \ldots + q_{T-1,T+1}) \gamma_2 = P(B_0) (\gamma_1 + 2\alpha \rho) - q_{T,T} \gamma_1. \tag{23}
\]

Inserting the product forms (15) into that equality yields for the left hand side,

\[
\frac{\alpha}{1 - \alpha} \phi_1 q_{T,T} \left[ 2\rho + \gamma_2 + (1 - \alpha) \beta^{T-2} \psi_{1,2} + (1 - \alpha) (1 - \beta^{T-2}) \frac{\beta}{1 - \beta} \gamma_2 \right], \tag{22}
\]

and for the right hand side,

\[
\frac{2\alpha \rho + \alpha \gamma_1}{1 - \alpha} q_{T,T}. \tag{23}
\]

Substituting the expressions (17), (18) and (21) into (22) and (23), and then equating (22) and (23) finally yields

\[
\phi_1 = \frac{(2\alpha + \rho \gamma_1)}{(2 + \rho \gamma_2)} \rho. \tag{24}
\]

Analogously, applying the balance principle to the set \( B_{-1} + B_{-2} + \ldots + B_{-T} \) leads to

\[
\phi_2 = \frac{(2(1 - \alpha) + \rho \gamma_2)}{(2 + \rho \gamma_1)} \rho. \tag{25}
\]

Hence, we derived explicit expressions for the parameters in the product forms (15). Now we have to verify that for these expressions all balance equations at the levels \( l > T \) are indeed satisfied. For example, substituting the product forms (15) into the left hand side of equation (3) yields

\[
\alpha^{l-T} \beta^{k-1} \phi_1 q_{T,T} 2(\rho + 1) \tag{26}
\]

and into the right hand side,

\[
\alpha^{l-T} \beta^k \phi_1 q_{T,T} 2\rho + \alpha^{l-T-1} \beta^k \phi_1 q_{T,T} \gamma_2 + \alpha^{l-T} \beta^{k-2} \phi_1 q_{T,T} \gamma_1. \tag{27}
\]
By dividing the left and right hand side by the common factor $\alpha^{l-T} \beta^{-k} \phi_1 \rho_{l,T}$ they simplify to

$$\beta 2(\rho + 1)$$

and

$$\beta^2 (2\rho + \alpha \gamma_2) + \gamma_1.$$  

By inserting the expressions (17), (21) and (24), both terms reduce to $\gamma_1 2(\rho + 1) / (2 + \rho \gamma_2)$. Accordingly, it can be verified that the other balance equations are satisfied.

Since $\alpha = \rho^2 < 1$, it follows that the sum of the product forms (15) over all levels $l > T$ absolutely converges. The balance equation in $(T, T)$ is also satisfied, for summing over all other equations in the set $Q$ and then interchanging summations exactly yields the equation in $(T, T)$. In other words, the product forms (15) are a nonnull and absolutely convergent solution of all balance equations. By a result of Foster ([4], Theorem 1) this proves that the $Q$ restricted process is ergodic. The probability $q_{l,T}$ finally follows from the normalization equation. However, we are actually interested in the unrestricted process. The ergodicity of the unrestricted process follows from the ergodicity of the $Q$ restricted process, and by relation (14) we obtain

**Theorem.**

For all $l > T$,

$$p_{l-k,l} = \rho^{2(l-T)} \rho_{l,T},$$

$$= \rho^{2(l-T)} \frac{\gamma_1}{2(\rho + 1)} \left( \frac{r_{l-k,l}}{2 + \rho \gamma_2} \right)^{k-1} \frac{(2a + \rho \gamma_1)}{(2 + \rho \gamma_2) \rho} \rho_{l,T}, \quad \text{if } k = 0,$$

$$= \rho^{2(l-T)} \frac{\gamma_1}{2(\rho + 1)} \left( \frac{r_{l-k,l}}{2 + \rho \gamma_2} \right)^{k-1} \frac{(2a + \rho \gamma_1)}{(2 + \rho \gamma_2) \rho} \rho_{l,T}, \quad \text{if } 0 < k < T,$$

and

$$p_{l,l-k} = \rho^{2(l-T)} \frac{\gamma_2}{2 + \rho \gamma_1} \left( \frac{r_{l,l-k}}{2 + \rho \gamma_2} \right)^{k-1} \frac{(2(1 - a) + \rho \gamma_2)}{(2 + \rho \gamma_1) \rho} \rho_{l,T}, \quad \text{if } 0 < k < T,$$

$$= \rho^{2(l-T)} \frac{\gamma_2}{2 + \rho \gamma_1} \left( \frac{r_{l,l-k}}{2 + \rho \gamma_2} \right)^{k-1} \frac{(2(1 - a) + \rho \gamma_2)}{(2 + \rho \gamma_1) \rho} \rho_{l,T}, \quad \text{if } k = T.$$  

**Remark 1.**

The product forms in the interior, i.e. for $0 \leq k < T$, are identical to the asymptotic solution of the asymmetric shortest queue problem without jockeying in [2]. The asymptotic
product form in [2] does not satisfy the boundary equations, and therefore, initializes a compensation method. That method consists of adding on terms so as to compensate alternatingly for errors on the boundaries of the state space. This approach eventually leads to the exact determination of the equilibrium probabilities. The above Theorem states that for the present jockeying problem the initial product form satisfies all balance equations, even those at the boundaries of the state space. So no compensation for errors at the boundaries is needed.

Remark 2.

From a computational point of view it is important to point out that the balance equations at the lower levels $l \leq T$ can be solved efficiently. In Conolly [3] and Adan, Wessels and Zijm [2] it is shown that the problem of solving the balance equations at the levels $l \leq T$ can be reduced to that of recursively solving the balance equations at the levels $T \rightarrow T-1 \rightarrow T-2 \rightarrow ... \rightarrow 1 \rightarrow 0$. Furthermore, in [2] it is shown that this procedure, if carried out in a sensible way, is numerically stable.

5. The Matrix-Geometric Approach

In this section we analyze the jockeying model by using the matrix-geometric approach and investigate the relationship with our approach. The matrix-geometric approach requires a partitioning of the vector $\overline{P}$ of equilibrium probabilities. Our approach suggests the following partitioning. Define $\overline{p}_l^T$ as the vector of equilibrium probabilities at level $l$, then the vector $\overline{P}$ can be partitioned into a $T^2$-vector $(\overline{p}_0, ..., \overline{p}_{T-1})$ of so-called boundary states and $2T+1$-vectors $\overline{p}_l$, $l \geq T$. Correspondingly, the generator $G$ is partitioned as

$$G = \begin{bmatrix}
B_{00} & B_{01} & 0 & 0 & 0 & \cdots \\
B_{10} & B_{11} & A_0 & 0 & 0 & \cdots \\
0 & A_2 & A_1 & A_0 & 0 & \cdots \\
0 & 0 & A_2 & A_1 & A_0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots
\end{bmatrix},$$

where the matrices $B_{00}$ and $B_{01}$ are of dimensions $T^2 \times T^2$ and $T^2 \times (2T+1)$ respectively. The matrix $B_{10}$ is of dimension $(2T+1) \times T^2$. The matrices $B_{11}, A_0, A_1$ and $A_2$ are square matrices of dimension $2T+1$, and
There are certainly other possibilities for the partitioning. In fact, Gertsbakh [5] uses a different partitioning (actually, he studies the symmetric problem, but his analysis can be easily extended to the asymmetric problem), but as will be shown at the end of this section our choice is definitely more useful.

In Section 4 we showed that the condition $\rho < 1$ is sufficient for ergodicity. Theorem 1.7.1 in Neuts' book [8] provides a necessary and sufficient condition for ergodicity. That Theorem
states that the Markov process is ergodic, if and only if
\[ \pi A_0 e < \pi A_2 e, \] (28)
where \( e \) is the column vector with all its components equal to one and \( \pi \) is given by
\[ \pi \left( A_0 + A_1 + A_2 \right) = 0, \quad \pi e = 1. \]

Straightforward calculations show that
\[ \pi_0 = C \frac{\gamma_1}{2(\rho + 1)} \left( \frac{\gamma_1}{2\rho + \gamma_2} \right)^{T-2} \frac{\gamma_1 + 2\alpha \rho}{2\rho + \gamma_2}, \]
\[ \pi_{T-k} = C \left( \frac{\gamma_1}{2\rho + \gamma_2} \right)^{k-1} \frac{\gamma_1 + 2\alpha \rho}{2\rho + \gamma_2}, \quad 0 < k < T, \]
\[ \pi_T = C, \]
\[ \pi_{T+k} = C \left( \frac{\gamma_2}{2\rho + \gamma_1} \right)^{k-1} \frac{\gamma_2 + 2(1 - \alpha)\rho}{2\rho + \gamma_1}, \quad 0 < k < T, \]
\[ \pi_{2T} = C \frac{\gamma_2}{2(\rho + 1)} \left( \frac{\gamma_2}{2\rho + \gamma_1} \right)^{T-2} \frac{\gamma_2 + 2(1 - \alpha)\rho}{2\rho + \gamma_1}, \]
where \( C \) follows from the normalization equation \( \pi e = 1 \). By inserting \( \pi \), inequality (28) reduces to \( \rho < 1 \). In other words, the condition \( \rho < 1 \) is not only sufficient for ergodicity, as shown in Section 4, but also necessary. Theorem 1.7.1 in [8] further states that if the Markov process is ergodic, then for all \( l > T \) the vectors \( p_l \) satisfy the matrix-geometric form
\[ \overline{p_l} = \overline{p_T} R^{l-T}, \]
where the matrix \( R \) is the minimal nonnegative solution of the matrix quadratic equation
\[ A_0 + R A_1 + R^2 A_2 = 0. \] (29)
In general that equation cannot be solved explicitly, but in our case more can be said. By Theorem 1.3.4 in [8], it follows that if \( A_0 \) has a zero row, the corresponding row in \( R \) is also zero. Hence, by (27), the matrix \( R \) is of the form
\[
R = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 \\
R_{T,0} & R_{T,1} & \cdots & R_{T,2T} \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 
\end{bmatrix}. \] (30)
This implies for all \( k \geq 0 \), that
\[
R^{k+1} = R^{T+1}
\]
and then the matrix-geometric solution simplifies to
\[
\vec{p}_l = \vec{p}_l R^{T-1} = p(T, 0, \ldots, R, 2T) R^{T-1}, \quad \text{if } l > T.
\]
Observe the similarities of that form with the product forms (15) in Section 4. By applying the same balance argument, which is used to obtain \( \alpha \) in the product forms (15), we obtain that
\[
R(T, 0) = \rho^2.
\]
It remains to solve the equations (29) for the rows \( R(T, 0), \ldots, R(T, T-1) \) and \( R(T, T+1), \ldots, R(T, 2T) \). By the relations (31) and (33), it follows that the equations (29) are not a set of quadratic equations, but a set of linear equations for the rows \( R(T, 0), \ldots, R(T, T-1) \) and \( R(T, T+1), \ldots, R(T, 2T) \). In fact, the equations (29) can be obtained by inserting the form (32) into the balance equations (1), (2), ..., (9) at level \( l > T \) and then dividing both sides by the common factor \( p(T, T) R^{T-2} \). Instead of solving the equations (29) algebraically, it is easier to use the form (32) as a starting point, together with the assumption that the rows \( R(T, 0), \ldots, R(T, T-1) \) and \( R(T, T+1), \ldots, R(T, 2T) \) satisfy a product form. Then we can mimic the analysis in Section 4 to obtain explicit expressions for the rows \( R(T, 0), \ldots, R(T, T-1) \) and \( R(T, T+1), \ldots, R(T, 2T) \). Specifically, we can apply the balance principle to exactly the same subsets \( A \) in order to obtain explicit expressions for the parameters in the product forms for the rows \( R(T, 0), \ldots, R(T, T-1) \) and \( R(T, T+1), \ldots, R(T, 2T) \). That results in the following expressions (cf. the Theorem in Section 4),
\[
R(T, T-k) = \rho^2 \left[ \frac{\gamma_1}{2 + \rho \gamma_2} \right]^{k-1} \left( \frac{2a + \rho \gamma_1}{2 + \rho \gamma_2} \right) \rho, \quad \text{if } 0 < k < T,
\]
\[
= \rho^2 \frac{\gamma_1}{2(\rho + 1)} \left[ \frac{\gamma_1}{2 + \rho \gamma_2} \right]^{T-2} \left( \frac{2a + \rho \gamma_1}{2 + \rho \gamma_2} \right) \rho, \quad \text{if } k = T,
\]
and
\[
R(T, T+k) = \rho^2 \left[ \frac{\gamma_2}{2 + \rho \gamma_1} \right]^{k-1} \left( \frac{2(1-a) + \rho \gamma_2}{2 + \rho \gamma_1} \right) \rho, \quad \text{if } 0 < k < T,
\]
\[
= \rho^2 \frac{\gamma_2}{2(\rho + 1)} \left[ \frac{\gamma_2}{2 + \rho \gamma_1} \right]^{T-2} \left( \frac{2(1-a) + \rho \gamma_2}{2 + \rho \gamma_1} \right) \rho, \quad \text{if } k = T.
\]

Remark 3.

The explicit solution of \( R \) is mainly due to the sparse matrix structure (30) of \( R \). That sparse structure is an immediate consequence of our choice for the partitioning of the state
space. The zero rows in $R$ correspond to zero rows in the matrix $A_0$, which contains the transition rates to the higher level, and our choice of the levels is such that the number of transition possibilities to the higher level is minimal (and thus the number of zero rows in $A_0$ maximal). Gertsbakh doesn’t exploit this property in his study. He partitions the vector $\tilde{p}$ into levels $l$, defined as the set of states for which the shortest queue has length $l$. Compared to our choice of the levels, an advantage of that choice is a less complex structure of the boundary states, but a disadvantage is that the sparse matrix structure of $R$ is lost (only the first and last row in $R$ are zero) and thereby all chance to find an explicit solution.

6. Conclusion and Extensions

In the preceding we studied the asymmetric shortest queue problem with threshold jockeying. We proved that the equilibrium probabilities $p_{m,n}$ of the queue lengths satisfy a product form at all levels higher than the threshold value $T$. In the remainder we discuss the extensions to different threshold values for each queue, one way jockeying, two nonidentical multi server groups and a perturbed model.

Different Threshold Values for Each Queue

The extension to a system where each queue uses a different threshold value, is straightforward. By choosing different threshold values one might take into account that the jobs in front of the slower server are more willing to join the other queue than the ones in front of the faster server.

One Way Jockeying

Another extension is a system allowing jockeying only in one direction, from the slower to the faster server say. The analysis of one way jockeying is a combination of the analysis in the present paper and the one in [1]. For this model compensation for errors on the boundaries is required. That compensation eventually leads to an infinite sum of product forms.

Two Nonidentical Multi Server Groups

Our approach also proceeds for a system with two nonidentical multi server groups, consisting of $N_1$ and $N_2$ servers say. In particular, the analysis of the probabilities $p_{m,n}$ at the levels $l$ higher than the maximum of $N_1+T$ and $N_2+T$ is identical to the analysis for two single server queues in Section 4 (with the service rates $\gamma_1$ and $\gamma_2$ replaced by $N_1\gamma_1$ and $N_2\gamma_2$).
A Perturbed Model

The product form structure of the probabilities $p_{m,n}$ at the levels $l > T$ is mainly due to the special boundary structure (cf. Remark 1). The question arises what happens if for all $l > T$ the rates at the boundary states $(l-T, l)$ and $(l, l-T)$ are perturbed. Let $\eta_1$ denote the new rate from $(l-T, l)$ to $(l-T+1, l)$ and $\varepsilon_1$ the new rate to $(l-T, l-1)$. The new rates at state $(l, l-T)$ are $\eta_2$ and $\varepsilon_2$ (see Figure 5).

![Figure 5: The rates in the perturbed model](image)

These perturbations possibly effect the ergodicity of the Markov process. Theorem 1.7.1 in [8] provides a necessary and sufficient ergodicity condition (cf. condition (28)). Furthermore, it can be shown that if the process is ergodic, then the equilibrium probabilities $p_{m,n}$ satisfy a sum of two product forms for states where the longer queue exceeds the threshold value $T$. In other words, for the perturbed model one extra product form is needed to satisfy the boundary equations.

References


<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>M 90-02</td>
<td>January</td>
<td>D.A. Overdijk</td>
<td>Meetkundige aspecten van de productie van kroonwielen</td>
</tr>
<tr>
<td>M 90-03</td>
<td>February</td>
<td>I.J.B.F. Adan, J. Wessels, W.H.M. Zijm</td>
<td>Analysis of the asymmetric shortest queue problem Part II: Numerical analysis</td>
</tr>
<tr>
<td>M 90-04</td>
<td>March</td>
<td>P. van der Laan, L.R. Verdooren</td>
<td>Statistical selection procedures for selecting the best variety</td>
</tr>
<tr>
<td>M 90-05</td>
<td>March</td>
<td>W.H.M. Zijm, E.H.L.B. Nelissen</td>
<td>Scheduling a flexible machining centre</td>
</tr>
<tr>
<td>M 90-06</td>
<td>March</td>
<td>G. Schuller, W.H.M. Zijm</td>
<td>The design of mechanizations: reliability, efficiency and flexibility</td>
</tr>
<tr>
<td>M 90-07</td>
<td>March</td>
<td>W.H.M. Zijm</td>
<td>Capacity analysis of automatic transport systems in an assembly factory</td>
</tr>
<tr>
<td>M 90-08</td>
<td>March</td>
<td>G.J. v. Houtum, W.H.M. Zijm</td>
<td>Computational procedures for stochastic multi-echelon production systems</td>
</tr>
<tr>
<td>Number</td>
<td>Month</td>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>M 90-09</td>
<td>March</td>
<td>P.J.M. van Laarhoven W.H.M. Zijm</td>
<td>Production preparation and numerical control in PCB assembly</td>
</tr>
<tr>
<td>M 90-10</td>
<td>March</td>
<td>F.A.W. Wester J. Wijngaard W.H.M. Zijm</td>
<td>A hierarchical planning system versus a schedule oriented planning system</td>
</tr>
<tr>
<td>M 90-11</td>
<td>April</td>
<td>A. Dekkers</td>
<td>Local Area Networks</td>
</tr>
<tr>
<td>M 90-12</td>
<td>April</td>
<td>P. v.d. Laan</td>
<td>On subset selection from Logistic populations</td>
</tr>
<tr>
<td>M 90-13</td>
<td>April</td>
<td>P. v.d. Laan</td>
<td>De Van Dantzig Prijs</td>
</tr>
<tr>
<td>M 90-14</td>
<td>June</td>
<td>P. v.d. Laan</td>
<td>Beslissen met statistische selectiemethoden</td>
</tr>
<tr>
<td>M 90-15</td>
<td>June</td>
<td>F.W. Steutel</td>
<td>Some recent characterizations of the exponential and geometric distributions</td>
</tr>
<tr>
<td>M 90-16</td>
<td>June</td>
<td>J. van Geldrop C. Withagen</td>
<td>Existence of general equilibria in infinite horizon economies with exhaustible resources. (the continuous time case)</td>
</tr>
<tr>
<td>M 90-17</td>
<td>June</td>
<td>P.C. Schuur</td>
<td>Simulated annealing as a tool to obtain new results in plane geometry</td>
</tr>
<tr>
<td>M 90-18</td>
<td>July</td>
<td>F.W. Steutel</td>
<td>Applications of probability in analysis</td>
</tr>
<tr>
<td>M 90-19</td>
<td>July</td>
<td>I.J.B.F. Adan J. Wessels W.H.M. Zijm</td>
<td>Analysis of the symmetric shortest queue problem</td>
</tr>
<tr>
<td>M 90-20</td>
<td>July</td>
<td>I.J.B.F. Adan J. Wessels W.H.M. Zijm</td>
<td>Analysis of the asymmetric shortest queue problem with threshold jockeying</td>
</tr>
</tbody>
</table>