

A NEW LOOK ON THE SHORTEST QUEUE SYSTEM WITH JOCKEYING

RACHEL RAVID 

ORT Braude College, Karmiel 2161002, Israel
E-mail: rachelr3@braude.ac.il

We introduce a Markov queueing system with Poisson arrivals, exponential services, and jockeying between two parallel and equivalent servers. An arriving customer admits to the shortest line. Every transition, of only the last customer in line, from the longer line to the shorter line may be accompanied by a certain fixed cost. Thus, a transition from the longer queue to the shorter queue occurs whenever the difference between the lines reaches a certain discrete threshold ($d = 2, 3, \dots$). In this study, we focus on the stochastic analysis of the number of transitions of an arbitrary customer.

Keywords: queueing theory

1. INTRODUCTION

In this paper, we consider a service facility where the rule according to which customers are assigned to the servers is the “join the shortest queue with threshold jockeying d ” ($JSQJ(d)$) rule. That is, a queueing system with $k \geq 2$ parallel servers, where an arriving customer joins the shortest queue or, in case of ties, one of the shortest queues according to some law. When the difference between the queue lengths reaches a threshold d , a customer from one of the longest queues switches to a shorter one. A Poisson input process with rate λ and independent exponential service times with rate μ are assumed. The goal of this paper is to study the distribution of the number of switches made by a customer during its waiting time. Most of the previous studies regarding the $JSQJ(d)$ model focused on the stationary distribution of the queue lengths. Adan, Wessels, and Zijm [1], Gertsbakh [7], and Haight [8], studied the model for $k = 2$. Tarabia [15] studied the case of finite capacity lines when a jockey occurs only to an empty line. This model was also studied by Jeganathan, Sumathi, and Mahalakshmi [11], assuming a perishable items inventory system. The case of $k > 2$ was analyzed by Adan, Wessels, and Zijm [2], Disney and Mitchell [5], Elsayed and Bastani [6], Kao and Lin [12], and Zhao and Grassmann [16]. Zaho and Grassmann [17] analyzed the model for an arbitrary renewal arrival process. The variance of the customer’s waiting time in the $JSQJ(2)$ system was analyzed by Haviv and Ritov [10]. Adan, Van Houtum, and Van der Wal [3] showed that the mean waiting time in the $JSQJ(d)$ forms a lower bound for the mean waiting time in the classical shortest queue model. It is also shown that the total number of jobs in the system is stochastically smaller than in the classical shortest

queue. Attention has also been given for jockeying policies by Dehghanian and Kharoufeh [4], Hassin and Haviv [9], and Stadje [14].

Motivation. The shortest queue model with jockeying is a very natural one and a switch between the lines may be accompanied by a cost. To our knowledge, the number of switches made by a customer during his sojourn time, has never been studied before. We believe this study yields valuable new insight into the model.

Contributions. Our main contributions are: (i) for $k = 2$ we derive the probability generating function of the number of switches made by a customer while waiting for service. This performance measure is useful for evaluating and optimizing service cost. (ii) We present some new methodological concepts which might be more broadly applicable; one example is the use of the “difference random walk.”

Organization of the paper. The model is introduced in Section 2. In Section 3, we derive recursive equations for the conditional generating function of the number of switches made by a customer, given its relative position in the system. The unconditional generating function is derived in Section 4.

2. MODEL DESCRIPTION AND PRELIMINARIES

We assume that the arrival process follows a homogeneous Poisson process with rate λ , and the service system consists of two parallel independent servers; each of them has a separate infinite capacity waiting line. The service times are independent and exponentially distributed with rate μ . A customer, upon arrival, joins the shorter queue. In case of a tie, each line will be chosen with probability 0.5. If the difference between the lengths of the lines reaches a threshold d , then the last customer in the longer line, joins the end of the shorter one. The system is considered to be in steady state. We study the case $d = 2$ in detail.

We are interested in probabilistic analysis of Y , the number of switches between lines, made by a customer while waiting for service. Let us first shift our attention to probabilities related to the queue lengths. Consider a tagged customer at his arrival instant. Since the system can be modeled as a continuous time Markov chain, then based on *PASTA* property, we know that the state distribution seen by this tagged customer is identical to the equilibrium distribution. Let N_1, N_2 be the random variables denoting the queue lengths in equilibrium. If we let $\rho/2 = \lambda/2\mu < 1$, then by Haight [8] we obtain

$$P(N_1 = n, N_2 = r) = \begin{cases} \frac{2 - \rho}{2 + \rho}, & n + r = 0, \\ \frac{\rho}{2} \cdot \frac{2 - \rho}{2 + \rho}, & n + r = 1, \\ 2 \frac{(2 - \rho)}{2 + \rho} \left(\frac{\rho}{2}\right)^{2(n+r)-1}, & |n - r| = 1, n + r \geq 2, \\ 2 \frac{(2 - \rho)}{2 + \rho} \left(\frac{\rho}{2}\right)^{2(n+r)}, & n = r, n + r \geq 2, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Additionally, from the tagged customer’s point of view, every customer standing in front of her, in her current line (including the served customer) will be called a “front” customer. All the other customers in system will be called “back” customers. The customer is considered to be in state (f, b) if she (or he) has f front customers and b back customers, $b \geq f \geq 1$

(excluding itself). Every decrease in the number of the front customers is defined to be a “step”. We distinguish between two types of steps: a “switch step” which is carried out when switching to the other line occurs before the next service completion in the customer’s current line, and a “non-switch step” which doesn’t involve a switch. A customer in state (f, b) will switch at most once before the number of front customers decreases by 1 to $f - 1$, and if there is a switch, the tagged customer will, at the time of the switch, have $f - 1$ front customers and f back customers. For the non-switch step case, we denote by $P_f(j | b)$ the conditional probability that when the customer next makes a step, its new state at that point is $(f - 1, j)$, given she (or he) is currently in state (f, b) , $1 \leq f \leq \min\{b, j\}$. Let $Y_{f,b}$ be the number of switch steps made by a customer since entering state (f, b) until entering service. Recursive relation for its generating function will be obtained in the next section.

3. THE CONDITIONAL NUMBER OF SWITCHES

The main goal of this section is to formulate recursive equations for $\Phi_{f,b}(s) = E(s^{Y_{f,b}})$, $0 \leq s \leq 1$. This requires that we first analyze the conditional probabilities $P_f(j | b)$, $1 \leq f \leq \min\{b, j\}$. Let us now state two results which will play a crucial role in the analysis. The first is an extension of the ballot problem, and the second relates to generating function theory.

LEMMA 3.1 [13]: Assume that candidate A begins the race $m - 1$ votes ahead of candidate B ($m = 1, 2, 3, \dots$), and collects $k \geq 0$ more votes for a total of $k + m - 1$ to B’s l votes ($l = 0, 1, 2, \dots$). The probability that candidate A will be ahead throughout the entire counting process, is denoted by $Q_m(k, l)$, where

$$Q_m(k, l) = \begin{cases} 1, & 0 \leq l < m, \\ 1 - \frac{\binom{l+k}{l+m}}{\binom{l+k}{l}}, & m \leq l \leq k + m - 1. \end{cases} \tag{2}$$

PROPOSITION 3.1 [18]: For $0 < x < 0.25$:

$$\sum_{n=0}^{\infty} \binom{2n+k}{n} x^n = (1 - 4x)^{-0.5} (2x)^{-k} [1 - \sqrt{1 - 4x}]^k. \tag{3}$$

Now, the conditional probabilities $P_f(j | b)$ can be determined. This enables us to investigate the non-switch step case. Without loss of generality we assume that the tagged customer is currently in line 1.

LEMMA 3.2: For $1 \leq f \leq b$

$$P_f(j | b) = \begin{cases} \Gamma(\mu)^{b-j+1} [1 - (\rho\Gamma^2(\mu))^{j-f+1}] \cdot [1 - \rho\Gamma^2(\mu)]^{-1}, & f \leq j \leq b - 1, \\ \Gamma(\mu)(\rho\Gamma(\mu))^{j-b} [1 - (\rho\Gamma^2(\mu))^{b-f+1}] \cdot [1 - \rho\Gamma^2(\mu)]^{-1}, & j \geq b, \end{cases} \tag{4}$$

where

$$\Gamma(\mu) = \frac{\rho + 2 - \sqrt{\rho^2 + 4}}{2\rho}.$$

PROOF: Consider a tagged customer who enters state (f, b) . In order that at the next front customer departure instant, it will still be in line 1 and enter state $(f - 1, j)$, $f \leq j \leq b - 1$,

two conditions must be satisfied regarding the current service in line 1: (i) the back customers always exceed $f - 1$ until that instant and (ii) the difference between the numbers of back customer departures and arrivals at that instant equals $b - j$. After substituting $m = b - (f - 1)$, and $l = k + (b - j)$ in(2), and using the total probability formula we obtain:

$$\begin{aligned}
 P_f(j | b) &= \int_{t=0}^{\infty} \sum_{k=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{k+b-j}}{(k+b-j)!} Q_{b-(f-1)}(k, k+b-j) \mu e^{-\mu t} dt \\
 &= \int_{t=0}^{\infty} \sum_{k=0}^{j-f} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{k+b-j}}{(k+b-j)!} \mu e^{-\mu t} \cdot 1 dt \\
 &\quad + \int_{t=0}^{\infty} \sum_{k=j-f+1}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{k+b-j}}{(k+b-j)!} \mu e^{-\mu t} \\
 &\quad \cdot \left[1 - \frac{k!(k+b-j)!}{(k+b-f+1)!(k-j+f-1)!} \right] dt \\
 &= \int_{t=0}^{\infty} \sum_{k=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{k+b-j}}{(k+b-j)!} \mu e^{-\mu t} dt \\
 &\quad - \int_{t=0}^{\infty} \sum_{k=j-f+1}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \frac{e^{-\mu t} (\mu t)^{k+b-j}}{(k+b-j)!} \mu e^{-\mu t} \\
 &\quad \cdot \frac{k!(k+b-j)!}{(k+b-f+1)!(k-j+f-1)!} dt.
 \end{aligned}$$

Substituting $h = k - (b - f + 1)$ in the last summand, using (3) and the identity:

$$\frac{1 - \sqrt{1 - \frac{4\lambda\mu}{(\lambda+2\mu)^2}}}{\frac{2\lambda\mu}{(\lambda+2\mu)^2}} = (\rho + 2)\Gamma(\mu),$$

yields the desired result. Similar considerations have been used for the case $j \geq b$. ■

From (4) it follows that the tagged customer's next step will be a switch step with probability

$$1 - \sum_{j=f}^{\infty} P_f(j | b) = [\Gamma(\mu)]^{b-f+1}. \tag{5}$$

Remark 3.1: Note that $\Gamma(\mu)$ is the Laplace transform for an $M/M/1$ busy period evaluated at the point μ .

We are now ready to derive recursive relation for $\Phi_{f,b}(s)$.

LEMMA 3.3: For $1 \leq f \leq b$,

$$\begin{aligned} \Phi_{f,b}(s) &= s\Phi_{f-1,f}(s) \cdot \Gamma(\mu)[(\Gamma(\mu))^{b-f} - (1 - \rho(\Gamma(\mu))^2)^{-1} \\ &\quad \cdot (\rho\Gamma(\mu))^{f-b} - \rho(\Gamma(\mu))^{b-f+2}] \\ &\quad + \left\{ (1 - \rho(\Gamma(\mu))^2)^{-1} \sum_{j=f}^{b-1} \Phi_{f-1,j}(s)[(\Gamma(\mu))^{b-j+1} - (\Gamma(\mu))^{j-b+1}] \right. \\ &\quad \left. + (\rho\Gamma(\mu))^f \Phi_{f,f}(s) \right\}. \end{aligned} \tag{6}$$

PROOF: Consider a customer who has f front customers and b back customers. As it is explained in the previous analysis, we distinguish between two cases according to the type of the customer’s next step. (i) If the next step is a switch step (from line 1 to line 2), the customer moves to state $(f - 1, f)$ hence $\Phi_{f,b}(s) = s\Phi_{f-1,f}(s)$. (ii) In the non-switch step case, the customer remains in line 1 and moves to state $(f - 1, j)$, $f \leq \min\{j, b\}$, so $\Phi_{f-1,j}(s) = \Phi_{f,b}(s)$. By assigning the probabilities according to (4) and (5) we get

$$\begin{aligned} \Phi_{f,b}(s) &= s\Phi_{f-1,f}(s) \cdot (\Gamma(\mu))^{b-f+1} \\ &\quad + \left\{ \sum_{j=f}^{b-1} \Phi_{f-1,j}(s) \cdot (\Gamma(\mu))^{b-j}[1 - (\rho\Gamma^2(\mu))^{j-f+1}] \right. \\ &\quad \left. + \sum_{j=b}^{\infty} \Phi_{f-1,j}(s) \cdot (\rho\Gamma(\mu))^{j-b}[1 - (\rho\Gamma^2(\mu))^{b-f+1}] \right\} \\ &\quad \cdot \Gamma(\mu)[1 - (\rho\Gamma^2(\mu))]^{-1}. \end{aligned} \tag{7}$$

If we let $f = b \geq 1$ in (7) then

$$\Phi_{f,f}(s) = s\Phi_{f-1,f}(s) \cdot \Gamma(\mu) + \Gamma(\mu) \cdot \sum_{j=f}^{\infty} \Phi_{f-1,j}(s)(\rho\Gamma(\mu))^{j-f}. \tag{8}$$

Substituting (8) into (7) yields the result after simple algebraic manipulations. ■

Remark 3.2: Since a customer does not switch during its own service, $\Phi_{0,b}(s) = 1$. Also, one can easily verify that $\Phi_{1,b}(s) = (s - 1)\Gamma^b(\mu) + 1$.

4. THE UNCONDITIONAL NUMBER OF SWITCHES

Return now to Y , the number of switch steps made by a customer during waiting time. By the total probability formula we get from (1)

$$\Phi_Y(s) = \frac{2 - \rho}{2 + \rho} \left[\frac{2 + \rho}{2} + 2 \sum_{n=1}^{\infty} \Phi_{n,n}(s) \left(\frac{\rho}{2}\right)^{2n} + \frac{4}{\rho} \sum_{n=1}^{\infty} \Phi_{n,n+1}(s) \left(\frac{\rho}{2}\right)^{2n} \right]. \tag{9}$$

In order to simplify the forthcoming notations, let

$$\begin{aligned}
 B_0(s) &= \sum_{n=1}^{\infty} \Phi_{n,n}(s) \left(\frac{\rho^2}{4}\right)^n, \\
 B_1(s) &= \sum_{n=1}^{\infty} \Phi_{n,n+1}(s) \left(\frac{\rho^2}{4}\right)^n.
 \end{aligned}
 \tag{10}$$

The main purpose in this section is to derive accessible expressions for $B_0(s)$ and $B_1(s)$. The analysis will be carried out by formulating two linear equations involving the desired expressions, and will consists of the following steps: in Subsection 4.1 we study probabilistic results regarding the number of front customer departures during a tagged customer waiting time. This yields a linear equation in $B_0(s)$ and $B_1(s)$ which will be derived in Subsection 4.2. Then, additional equation will be constructed by conditioning on the system’s next step (arrival or departure).

4.1. The Number of Front Customer Departures During Waiting Time

For a tagged customer, we look at the difference between the numbers of back and front customers at state-change epochs during waiting time. The differences at these discrete time points form a random process in the nonnegative integers. The process takes steps as follows: (i) A departure of a front customer or an arrival occurs with probability $(\mu + \lambda)/(\lambda + 2\mu)$ and increases the difference by one. It will be called an “up step.” (ii) A departure of a back customer occurs with probability $\mu/(\lambda + 2\mu)$ and decreases the difference by one. It will be called a “down step.” We call this process the “difference random walk” (*DRW*).

Since switching between the lines is possible only immediately after the customer leaves state (n, n) , $n \geq 1$, it is worthwhile to analyze the *DRW* hitting times at 0. From the model description in Section 2, it appears that the *DRW* reaches 1 at least once. Let $T_{1,0}$ be the time elapsed until the *DRW* first hits 0, given it starts in 1. Note that the *DRW* terminates when the tagged customer enters service and it has a positive drift. Each of these implies that $P(T_{1,0} = \infty) > 0$. We denote by K the number of front customer departures during $T_{1,0}$, and by $\Psi_K(\theta)$ its probability generating function (K is a defective random variable on the nonnegative integers).

LEMMA 4.1:

$$\Psi_K(\theta) = \frac{\lambda + 2\mu - \sqrt{(\lambda + 2\mu)^2 - 4(\mu\theta + \lambda)\mu}}{2(\mu\theta + \lambda)}, \quad 0 \leq \theta < 1.
 \tag{11}$$

PROOF: Denote by \tilde{D} the number of up steps during $T_{1,0}$, and by $\Psi_{\tilde{D}}(\theta)$ its probability generating function. From the simple random walk theory it follows that

$$\Psi_{\tilde{D}}(\theta) = \frac{\lambda + 2\mu - \sqrt{(\lambda + 2\mu)^2 - 4\theta\mu(\lambda + \mu)}}{2(\lambda + \mu)\theta}, \quad 0 < \theta < 1.$$

The up steps are considered as independent Bernoulli trials. A success is defined to be a front customer departure. $\Psi_K(\theta)$ now follows, since given $\tilde{D} = d < \infty$, we have that $K|\tilde{D} = d \sim \text{Binomial}(d, \mu/(\mu + \lambda))$. Hence

$$\Psi_K(\theta) = \Psi_{\tilde{D}} \left(\frac{\mu\theta}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} \right).
 \tag{12}$$

Rewriting (12) completes the proof. ■

4.2. Determination of $B_0(s)$ and $B_1(s)$

Now we are ready to state the following lemmas, in which linear equations containing $B_0(s)$ and $B_1(s)$ will be constructed.

LEMMA 4.2:

$$B_1(s) = B_0(s) \frac{2}{\rho + 4} + \sum_{n=1}^{\infty} \left(\frac{\rho^2}{4}\right)^n P(K \geq n). \tag{13}$$

PROOF: Assume that the tagged customer is in state $(n, n + 1)$, $n \geq 1$, this implies that the *DRW* takes the value 1. With probability $P(K = k)$, k front customers will depart the system during $T_{1,0}$. Hence, the *DRW* first hits 0 when the customer enters state $(n - k, n - k)$, $0 \leq k < n$. If $K \geq n$, then the *DRW* does not hit 0. Hence, the tagged customer takes only non-switch steps since she (or he) enters state $(n, n + 1)$ until her service begins. From the total probability formula we get

$$\Phi_{n,n+1}(s) = \sum_{k=0}^{n-1} P(K = k) \Phi_{n-k,n-k}(s) + [P(K \geq n)] s^0. \tag{14}$$

(13) is obtained by multiplying both sides of (14) by $(\rho^2/4)^n$ and summing over $n \geq 1$. Note that $\Psi_K(\rho^2/4) = \frac{2}{\rho+4}$. ■

Until now we have derived one linear equation containing $B_0(s)$ and $B_1(s)$, so one more equation is needed. It will be derived in the following lemma.

LEMMA 4.3:

$$B_0(s) = \left[\frac{(s+1)\rho^2}{\rho+2} + \frac{\rho}{\rho+2} \right] B_1(s) + \frac{(s+1)\rho^2}{\rho+2}. \tag{15}$$

PROOF: We start with a recursive relation in terms of $\Phi_{n,n}(s)$ and $\Phi_{n,n+1}(s)$. Consider a customer who is in state (n, n) , $n \geq 1$. The next event can be an arrival with probability $\lambda/(\lambda + 2\mu)$, a front or a back customer departure occurs, each with probability $\mu/(\lambda + 2\mu)$. By the total probability formula we obtain

$$\Phi_{n,n}(s) = \frac{\mu(s+1)}{\lambda + 2\mu} \Phi_{n-1,n}(s) + \frac{\lambda}{\lambda + 2\mu} \Phi_{n,n+1}(s). \tag{16}$$

The result follows after multiplying both sides of (16) by $(\rho^2/4)^n$ and summing over $n \geq 1$. ■

From (9), (13), and (15), it now follows

$$\Phi_Y(s) = \frac{2-\rho}{2} + \frac{2-\rho}{2+\rho} B_0(s) \left[2 + \frac{8}{\rho(4+\rho)} \right] + \frac{2-\rho}{2+\rho} \cdot \frac{4}{\rho} \sum_{n=1}^{\infty} \left(\frac{\rho^2}{4}\right)^n P(K \geq n), \tag{17}$$

where

$$B_0(s) = \left\{ \left[\sum_{n=1}^{\infty} \left(\frac{\rho^2}{4}\right)^n P(K \geq n) \right] \cdot [(s+1)\rho^2 + 4\rho] + (s+1)\rho^2 \right\} \cdot 0.5 \cdot (4+\rho) \cdot [2(\rho^2 + 4\rho + 8) - (s+1)\rho^2]^{-1}. \tag{18}$$

Acknowledgments

I would like to thank Prof. David Perry for his useful assistance and guidance with this paper.

References

1. Adan, I.J.B.F., Wessels, J., & Zijm, W.H.M. (1991). Analysis of the asymmetric shortest queue problem with threshold jockeying. *Communications in Statistics. Stochastic Models* 7(4): 615–627.
2. Adan, I.J.B.F., Wessels, J., & Zijm, W.H.M. (1993). Matrix-geometric analysis of the shortest queue problem with threshold jockeying. *Operations Research Letters* 13: 107–112.
3. Adan, I.J.B.F., Van Houtum, J.G., & Van der Wal, J. (1994). Upper and lower bounds for the waiting time in the symmetric shortest queue system. *Annals of Operations Research* 48: 197–217.
4. Dehghanian, A. & Kharoufeh, J.P. (2016). Strategic dynamic jockeying between two parallel queues. *Probability in the Engineering and Informational Sciences* 30(1): 41–60.
5. Disney, R.L. & Mitchell, W.E. (1970). A solution for queues with instantaneous jockeying and other customer selection rules. *Naval Research Logistics Quarterly* 17(3): 315–325.
6. Elsayed, E.A. & Bastani, A. (1985). General solutions of the jockeying problem. *European Journal of Operational Research* 22(3): 387–396.
7. Gertsbakh, I. (1984). The shorter queue problem: A numerical study using the matrix-geometric solution. *European Journal of Operational Research* 15(3): 374–381.
8. Haight, F.A. (1958). Two queues in parallel. *Biometrika* 45: 401–410.
9. Hassin, R. & Haviv, M. (1994). Equilibrium strategies and the value of information in a two line queueing system with threshold jockeying. *Communications in Statistics. Stochastic Models* 10(2): 415–435.
10. Haviv, M. & Ritov, Y. (1987). The variance of the waiting time in a queueing system with jockeying. *Communications in Statistics. Stochastic Models* 4(1): 161–182.
11. Jeganathan, K., Sumathi, J., & Mahalakshmi, G. (2016). Markovian inventory model with two parallel queues, jockeying and impatient customers. *Yugoslav Journal of Operations Research* 26(4): 467–506.
12. Kao, E.P.C. & Lin, C. (1990). A matrix-geometric solution of the jockeying problem. *European Journal of Operational Research* 44(1): 67–74.
13. Reuveni, S. (2014). Catalan's trapezoids. *Probability in the Engineering and Informational Sciences* 28: 353–361.
14. Stajje, W. (2009). A queueing system with two parallel lines, cost-conscious customers, and jockeying. *Communications in Statistics. Theory and Methods* 38(16–17): 3158–3169.
15. Tarabia, A.M.K. (2008). Analysis of two queues in parallel with jockeying and restricted capacities. *Applied Mathematical Modelling* 32(5): 802–810.
16. Zhao, Y. & Grassmann, W.K. (1990). The shortest queue model with jockeying. *Naval Research Logistics* 37(5): 773–787.
17. Zhao, Y. & Grassmann, W.K. (1995). Queueing analysis of a jockeying model. *Operations Research* 43(3): 520–529.
18. Wilf, H.S. (2006). *Generating functionology*. 3rd ed. Wellesley, Mass: A K Peters, Ltd.