The shorter queue polling model

Ivo Adan, Onno Boxma, Stella Kapodistria, Vidyadhar Kulkarni

The shorter queue polling model

Ivo J.B.F. Adan[*]   Onno J. Boxma [†]   Stella Kapodistria[‡]   Vidyadhar G. Kulkarni [§]

June 3, 2013

Abstract

We consider a two-queue polling model in which customers upon arrival join the shorter of two queues. Customers arrive according to a Poisson process and the service times in both queues are independent and identically distributed random variables having the exponential distribution. The two-dimensional process of the numbers of customers at the queue where the server is and at the other queue is a two-dimensional Markov process. We derive its equilibrium distribution using two methodologies: the compensation approach and a reduction to a boundary value problem.

Keywords: polling models, join the shorter queue, compensation approach, boundary value problem

AMS 2000 *Subject Classification*: 60K25, 90B22

1   Introduction

Multi-queue systems appear everywhere, from supermarkets, airport check-in points and highway tolls, to complex wireless data exchange and telecommunications systems. A common characteristic behind all these systems is the decision which queue to join. The answer is not always simple or obvious, and is based on optimality criteria and the structure of the underlying system. There exists an extensive literature on dispatching policies and their optimality [9, 26, 27, 35, 36, 37, 42, 46]. Among the dispatching policies, the join-the-shortest-queue (JSQ) policy has received considerable attention [5, 6, 10, 19, 20, 23, 24, 28, 30, 44, 45]. The JSQ policy in some scenarios has been proven to be the optimal policy; on the one hand it minimizes the customers mean waiting time [25] and on the other hand it stochastically maximizes the number of customers served by time t, $t > 0$ [43]. Indeed, when i) the service stations are symmetrical, ii) any one of them can process incoming work, and iii) the work requirements for the incoming jobs are identically distributed, it is intuitive that JSQ optimally balances the workload and minimizes congestion. Proving the optimality of JSQ, and doing a performance analysis of the ensuing model, become more difficult when there exist structural dependencies between the queues. Menich in [31] considers a general Markovian system with arrival, service and holding costs that depend on the entire vector of queue lengths, and proves in this context that JSQ minimizes discounted and average costs over an infinite horizon. In [32] the authors show that, under certain symmetry and monotonicity conditions, assigning the customers to the shortest queue and devoting auxiliary capacity to serve the longest queue is optimal for a wide class of Markovian systems: the queue lengths will be stochastically smaller in the

---

[1]Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: `i.j.b.f.adan@tue.nl`

[2]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: `o.j.boxma@tue.nl`

[3]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: `s.kapodistria@tue.nl`

[4]Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599, USA, E-mail: `vkulkarn@email.unc.edu`

weak submajorization ordering than the queue lengths under any other policy.

In [3], Adan et al. view a similar question – *which is the optimal dispatching policy for a system with correlated queues* – under a different perspective, that of polling models. A typical polling model consists of multiple queues, attended by a single server who visits the queues in some order to render service to the customers waiting at the queues. Moving from one queue to another, the server incurs a (possibly zero) switchover time. Triggered by a wide variety of applications, polling models have been extensively studied in the literature; see [38], [40], [41] for a series of comprehensive surveys and [11], [29], [39] for extensive overviews of the applicability of polling systems. In the polling literature much attention has been given to the determination of the probability generating function (PGF) of the joint queue length distribution under stationarity and at various epochs. In particular, a wide range of service disciplines has been considered, like *exhaustive service* (per visit to a queue, the server continues to serve all customers until it empties) and *gated service* (per visit to a queue, the server serves only those customers which are already present at the start of the visit). In [33], Resing shows that the joint queue length PGF of polling systems in which the service discipline satisfies a so-called branching property equals the (known) PGF of a multi-type branching process with immigration. Service disciplines which satisfy the branching property include the exhaustive and gated disciplines. Polling systems with disciplines which do not satisfy the branching property usually defy an exact analysis.

The authors in [3] consider a typical polling system consisting of two queues. More explicitly, they assume that customers arrive at this system according to a Poisson process. There is a single server serving both queues in a cyclic fashion with exhaustive service. The service times at each queue are independent and identically distributed random variables having an exponential distribution and both queues have identical service rates. Moreover, the switch-over times are assumed to be zero. The authors in [3] identify the optimal dispatching policy for this system given a structure of cost-reward functions that depend on the mean waiting time. They distinguish three levels of information that the customers receive: i) no information (the customers upon arrival receive no information at all, i.e., they do not know the lengths of the two queues nor the position of the server), ii) partial information (the customers are informed about the position of the server, but not about the queue lengths) and iii) full information (the customers upon arrival can fully observe the system, i.e., they know the queue lengths and the position of the server).

For the case of no information, Adan et al. show in [3] that the optimal dispatching policy is to toss a fair coin and with probability 1/2 join queue 1 and otherwise join queue 2. In the case of partial information, the authors [3] show that the optimal strategy is to follow the server, i.e., to join the queue that is currently being served. For the study of the fully observable case Adan et al. embed a reward-cost structure in the two-queue polling model so as to determine the optimal policy of dispatching arrivals in the two queues (busy and idle) that minimizes the long run average cost per customer per time unit. They assume that each customer in the system pays a cost c per time unit he spends waiting in the queue that is currently being served (referred to as the busy queue) and a cost C per time unit he spends in the service area, while there is no cost to wait in the queue that is not being served (referred to as the idle queue). Furthermore, they assume that there is a fixed reward for completing service. Customers that enter the system are not allowed to renege. Given full information (position of the server and queue lengths), the individual optimal (Nash equilibrium) policy as proven in [3] is to assign the new customer to the shorter of the two queues and in the case of a tie to dispatch the customer to the queue currently receiving service. Moreover, the authors consider a fluid model approximation to derive the socially optimal costs and linear switching curve policies.

We are motivated on the one hand by the analysis performed in [3] and on the other hand by the lack of analytic results on the stationary distribution of polling models under various dispatching policies. For this reason, we present

a methodological approach for the derivation of the stationary distribution of the two-queue polling model under the optimal strategy (JSQ policy and, in the case of a tie, follow the server). In particular, we use two different techniques to obtain the steady-state distribution of the numbers of customers in the busy and idle queue: the compensation approach and the boundary value method. Both methods have been explored for a class of two-dimensional random walks, but have considerable restrictions. Our paper makes contributions regarding the applicability and accessibility of these methods. Moreover, we believe that the approaches described in this manuscript can be easily extended to other dispatching policies [27, 35, 36, 37]. Below we globally describe both methods.

The compensation approach is developed by Adan et al. in a series of papers [1, 2, 5, 6, 7] and aims at a direct solution for the sub-class of two-dimensional random walks on the lattice of the first quadrant that obey the following conditions:

i) Step size: Only transitions to neighboring states.

ii) Forbidden steps: No transitions from interior states to the North, North-East, and East.

iii) Homogeneity: The same transitions occur according to the same rates for all interior points, and similarly for all points on the horizontal boundary, and for all points on the vertical boundary.

It exploits the fact that the balance equations in the interior of the quarter plane are satisfied by linear (finite or infinite) combinations of product-forms, the parameters of which satisfy a kernel equation, and that need to be chosen such that the equilibrium equations on the boundaries are satisfied as well. As it turns out, this can be done by alternatingly compensating for the errors on the two boundaries, which eventually leads to an infinite series of product-forms. The JSQ polling system under consideration is somewhat more complicated than the models in [1, 2, 5, 6, 7], but we show that its solution can in this case be obtained as a *tree of product-forms*.

We use the compensation approach to directly solve the equilibrium equations of the system and obtain the equilibrium distribution. Subsequently, we demonstrate that a generating function approach for the equilibrium distribution of the joint queue lengths yields as expected the same result. We show that the analysis of the shorter queue polling model can be reduced to that of a boundary value problem.

The boundary value method is an analytic method which is applicable to some two-dimensional random walks restricted to the first quadrant. The bivariate probability generating function, say $P(x, y)$, of the position of a homogeneous small step (nearest neighbor) random walk satisfies a functional equation of the form

$$K(x, y)P(x, y) + A(x, y)P(x, 0) + B(x, y)P(0, y) \tag{1.1}$$
$$+ C(x, y)P(0, 0) + D(x, y) = 0.$$

The method consists of the following steps:

i) First, define the zero tuples (x, y) such that $K(x, y) = 0$, $|x|, |y| < 1$.

ii) Then, along the curve $K(x, y) = 0$ (and provided that $P(x, y)$ is defined on this curve), Equation (1.1) reads

$$A(x, y)P(x, 0) + B(x, y)P(0, y) \tag{1.2}$$
$$+ C(x, y)P(0, 0) + D(x, y) = 0.$$

iii) Finally, Equation (1.2) can be solved as a Riemann-Hilbert boundary value problem.

3

Malyshev pioneered this approach of transforming the functional equation to a boundary value problem in the 1970's. The idea to reduce the functional equation for the generating function to a standard Riemann-Hilbert boundary value problem stems from the work of Fayolle and Iasnogorodski [16] on two parallel M/M/1 queues with coupled processors (the service speed of a server depends on whether or not the other server is busy). Extensive treatments of the boundary value technique for functional equations can be found in Cohen and Boxma [14] and Fayolle, Iasnogorodski and Malyshev [17].

Contribution of the paper. In this paper we provide an exact analysis of a model that unifies two fundamental queueing models: the JSQ model and the polling model. As stated above, the main contribution of the paper is methodological. We extend the applicability of the compensation approach and the boundary value method, applying them successfully to a two-dimensional random walk on the lattice of the first quadrant with arbitrary size jumps on the boundaries and transitions to the North (lower part of the first quadrant) or the East (upper part of the quadrant), cf. Figure 1. We would also like to note that our analysis can be extended to the case that there are different service rates at the two queues. Another contribution of our paper is that we present an alternative derivation of the equilibrium distribution that is very efficient for a wide class of two-dimensional queueing problems that exhibit different behavior in two different parts (upper and lower) of the quadrant and the underlying bivariate generating function of which is meromorphic.

Organization of the paper. In Section 2 we present the model in detail, and we derive the stability condition and the Kolmogorov equations for the equilibrium joint queue length distribution. In Sections 3 and 4 we study the equilibrium distribution using the compensation approach and the boundary value method, respectively. Section 5 contains several numerical results.

2 The Model

We consider a two-queue polling system with exhaustive service discipline and cyclic polling schedule. The service times are independent and identically distributed random variables having an exponential distribution with parameter μ, at both stations. Customers arrive to this system according to a Poisson process with rate λ and join the shorter queue (including the customer in service, if any). When both queues are equal, they join the queue where the server is. For the sake of concreteness, we assume that the server stays at the current queue if the system is completely empty after a service completion. Let $X(t)$ be the number of customers in the queue where the server is and $Y(t)$ be the number of customers in the other queue, at time t. Then the state of the system at time t, $t \geq 0$, is given by $(X(t), Y(t))$. The two-dimensional stochastic process $\{(X(t), Y(t)), t \geq 0\}$ is a CTMC on state space $S_{X,Y} = \{(0,0)\} \cup \{(i,j) : i \geq 1, j \geq 0\}$. The transition rates read as follows:

$$q\{(i,j),(i+1,j)\} = \lambda, \;\; 0 \leq i \leq j,$$
$$q\{(i,j),(i,j+1)\} = \lambda, \;\; i > j \geq 0,$$
$$q\{(i,j),(i-1,j)\} = \mu, \;\; i \geq 2, \; j \geq 0,$$
$$q\{(1,j),(j,0)\} = \mu, \;\; j \geq 0.$$

The rate transition diagram of the model is depicted in Figure 1.

We note that the total number of customers, $X(t) + Y(t)$, behaves as in a standard $M/M/1$ model with arrival rate λ and service rate μ. Based on this remark it is evident that the stability condition for the polling model at hand is
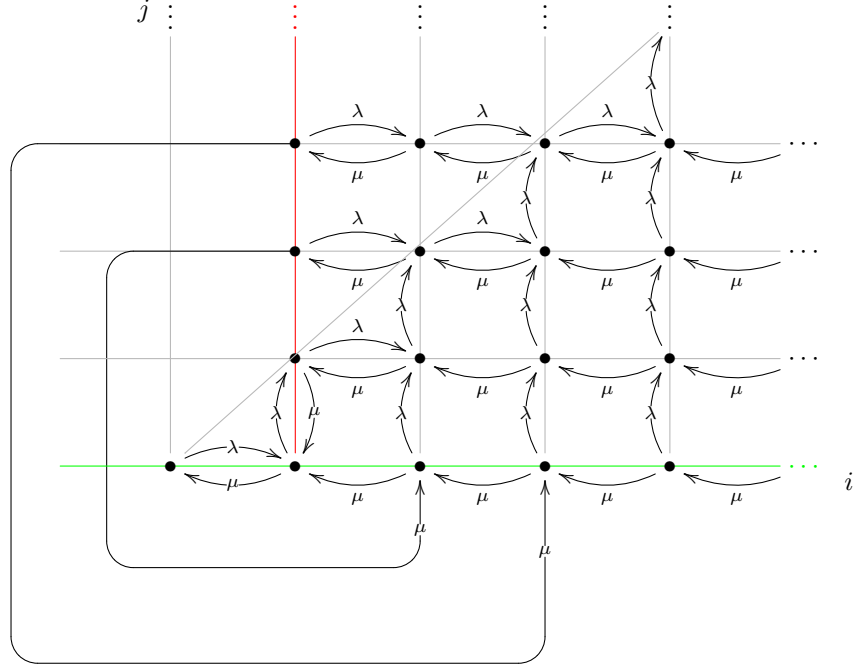
4

Figure 1: Transition rate diagram $\{X, Y\}$.

given by

$$\rho = \lambda/\mu < 1, \tag{2.1}$$

and under this condition the stationary distribution of the total number of customers has the geometric distribution with parameter ρ,

$$\mathbb{P}[X + Y = k] = (1 - \rho)\rho^k, \quad k = 0, 1, 2, \dots .$$

Under the stability condition (2.1) the limiting probabilities

$$\pi(i, j) = \lim_{t \to \infty} P(X(t) = i, Y(t) = j), \quad (i, j) \in S_{X,Y},$$

exist and are derived as the unique solution to the normalizing condition $\sum_{(i,j) \in S_{X,Y}} \pi(i, j) = 1$ plus the following balance equations:

$$(\lambda + \mu)\pi(i, j) = \lambda\pi(i - 1, j)\mathbb{1}_{\{i \leq j+1, i \geq 2\}} + \lambda\pi(i, j - 1)\mathbb{1}_{\{i > j-1\}} \tag{2.2}$$
$$+ \mu\pi(i + 1, j), \quad i \geq 1, j \geq 1,$$

$$(\lambda + \mu)\pi(i, 0) = \lambda\pi(0, 0)\mathbb{1}_{\{i=1\}} + \mu\pi(i + 1, 0) \tag{2.3}$$
$$+ \mu\pi(1, i), \quad i \geq 1, j = 0,$$

$$\lambda\pi(0, 0) = \mu\pi(1, 0), \tag{2.4}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes Kronecker's delta function taking value one whenever the event inside the curly brackets is satisfied and value zero otherwise.

5

3   Compensation approach

For the first technique we find it more convenient to work with the following random variables:

$$M(t) = \min(X(t), Y(t)), \quad N(t) = Y(t) - X(t).$$

Clearly the state of the system at time t, $t \geq 0$, is equally described by $(M(t), N(t))$. The two-dimensional stochastic process $\{(M(t), N(t)), t \geq 0\}$ is a CTMC on the state space

$$S_{M,N} = \{(m, n) : m \geq 1, n \geq 1\} \cup \{(m, -n) : m \geq 0, n \geq 0\}.$$

The transition rates are

$$q\{(m, n), (m + 1, n - 1)\} = \lambda, \quad m \geq 1, n \geq 1,$$
$$q\{(m, n), (m - 1, n + 1)\} = \mu, \quad m \geq 2, n \geq 0,$$
$$q\{(1, n), (0, -n - 1)\} = \mu, \quad n \geq 0,$$
$$q\{(m, 0), (m, -1)\} = \lambda, \quad m \geq 0,$$
$$q\{(m, -n), (m, -n + 1)\} = \mu, \quad m \geq 0, n \geq 1,$$
$$q\{(m, -n), (m + 1, -n + 1)\} = \lambda, \quad m \geq 0, n \geq 1.$$

Under the stability condition $\rho = \lambda/\mu < 1$ the limiting probabilities

$$p(m, n) = \lim_{t \to \infty} P(M(t) = m, N(t) = n), \quad (m, n) \in S_{M,N},$$

exist and are derived as the unique solution to the following balance equations:

$$p(m, -1)(\lambda + \mu) = p(m, 0)\lambda + p(m - 1, -2)\lambda \tag{3.1}$$
$$+ p(m, -2)\mu, \quad m \geq 1, \tag{3.2}$$
$$p(m, 0)(\lambda + \mu) = p(m - 1, 1)\lambda + p(m, -1)\mu \tag{3.3}$$
$$+ p(m - 1, -1)\lambda, \quad m \geq 2, \tag{3.4}$$
$$p(m, n)(\lambda + \mu) = p(m - 1, n + 1)\lambda \tag{3.5}$$
$$+ p(m + 1, n - 1)\mu, \quad m \geq 2, n \geq 1, \tag{3.6}$$
$$p(1, n)(\lambda + \mu) = p(2, n - 1)\mu, \quad n \geq 1, \tag{3.7}$$
$$p(0, -n)(\lambda + \mu) = p(0, -n - 1)\mu + p(1, n - 1)\mu, \quad n \geq 2, \tag{3.8}$$
$$p(m, -n)(\lambda + \mu) = p(m, -n - 1)\mu \tag{3.9}$$
$$+ p(m - 1, -n - 1)\lambda, \quad m \geq 1, n \geq 2, \tag{3.10}$$
$$p(0, 0)\lambda = p(0, -1)\mu, \tag{3.11}$$
$$p(0, -1)(\lambda + \mu) = p(0, 0)\lambda + p(1, 0)\mu + p(0, -2)\mu, \tag{3.12}$$
$$p(1, 0)(\lambda + \mu) = p(0, -1)\lambda + p(1, -1)\mu, \tag{3.13}$$

plus the normalizing condition $\sum_{(m,n) \in S_{M,N}} p(m, n) = 1$. Once we have the limiting probabilities $p(m, n)$, $(m, n) \in S_{M,N}$, we can compute $\pi(i, j)$, $(i, j) \in S_{X,Y}$, as follows:

$$\pi(i, j) = p(\min(i, j), j - i).$$

In this section we obtain the equilibrium distribution of the queue lengths using the compensation approach. This approach yields an explicit expression, without transforms, for the equilibrium distribution. Surprisingly, it is successful for this model, although one-step transitions are allowed to the North, North-East and East and arbitrarily
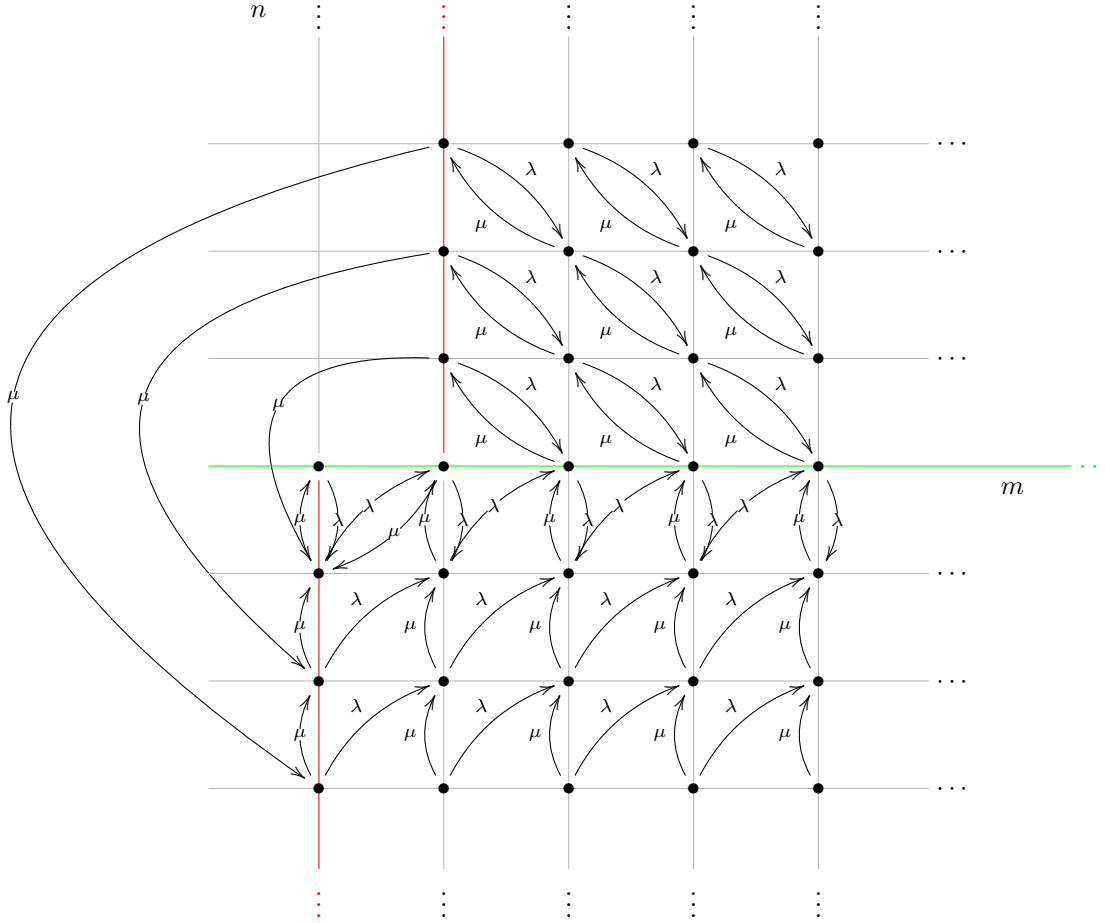
Figure 2: Transition rate diagram $\{M, N\}$.

big transitions are allowed on the boundaries, see Figure 1. In [1, 2, 5, 6, 7] the compensation approach was developed for nearest-neighbour two-dimensional random walks under the explicit condition that such transitions are not allowed. The successful application of the compensation approach for our model indicates that the conditions (i)-(iii) for applying the compensation approach, see Section 1 page 3, are sufficient but not necessary. For other exceptional applications of the compensation approach "violating" the conditions (i)-(iii) see [4, 8, 34].

The compensation approach as has been briefly discussed in the introduction attempts to solve the equilibrium equations by a linear combination of product-forms. This is achieved by first characterizing a sufficiently rich basis of product-form solutions satisfying the equilibrium equations in the interior of the state space. Subsequently this basis is used to construct a linear combination that also satisfies the equations for the boundary states. Note that the basis contains uncountably many elements. Therefore a procedure is needed to select the appropriate elements. This procedure is based on a compensation argument (which explains the name of the method): after introducing the first term, countably many terms are subsequently added so as to alternatingly compensate for the error on one of the two boundaries. The main steps in the analysis are briefly outlined below:

Step 1 Characterize the set of product-forms

$$\alpha^m \beta^n \text{ and } \alpha^m \beta^{-n}$$

satisfying the equilibrium equations in the interior of the state space for the first and the fourth quadrant, i.e., Equations (3.5) and (3.9), respectively. Substitution of the product-form $\alpha^m \beta^n$ into (3.5) and division by common powers yields a quadratic equation in α and β:

$$\alpha\beta(\lambda + \mu) = \beta^2 \lambda + \alpha^2 \mu. \tag{3.14}$$

Similarly, substituting the product-form $\alpha^m \beta^{-n}$ into (3.9) yields after some manipulations:

$$\alpha(\lambda + \mu) = \beta\lambda + \alpha\beta\mu. \tag{3.15}$$

The solutions to Equations (3.14) and (3.15) form the basis. In particular the points on the curves (3.14) and (3.15) inside the region $0 < |\alpha|, |\beta| < 1$ characterize a continuum of product-forms satisfying the inner equations.

Step 2 Construct a linear combination of elements in this rich basis, which is a formal solution to the equilibrium equations. Here the word formal is used to indicate that (at this stage) we do not bother about the convergence of the solution. This aspect is treated in Step 3. The construction of a linear combination starts with a suitable initial term that satisfies the interior of the state space and also the equilibrium equations on the horizontal boundary $n = 0$, corresponding to Equations (3.1), (3.3), (3.5) and (3.9). This term violates the equilibrium equations (3.7) and (3.8) on the vertical boundary. To compensate for this error we add a new product-form term coming from the basis, such that the sum of the two terms satisfies the equilibrium equations in all states on the vertical boundary. Adding the new term violates the equilibrium equations on the horizontal boundary, hence we again need to add a product-form solution such that the sum of the three terms satisfies the equilibrium equations in all states on the horizontal boundary. We continue in this manner until we construct the entire formal series.

Step 3 Prove that the formal solution converges. This is split up into two parts: i) we first show in Proposition 3.1 that the sequences of α's and β's converge to zero exponentially fast and ii) we show in Theorem 3.2 that the formal solution converges absolutely in all states.

Step 4 Determine the normalization constant.

Performing the steps described above for the compensation approach leads to the following main result for the equilibrium distribution.

Theorem 3.1. *For all states* $(m, n) \in S$,

$$p(m, n) = C^{-1}(1 - \rho^m) \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \alpha_{i,k}^{m+n}, \ m \geq 1, n \geq 0, \tag{3.16}$$

$$p(m, -n) = C^{-1}(1 - \rho)\rho^m \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}^{m+n}}{(1 + \rho - \alpha_{i,k})^{m+1}}, \tag{3.17}$$

$$m \geq 0, n \geq 2,$$

$$p(m, -1) = C^{-1} \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} d_{i,k} \alpha_{i,k}^{m}, \ m \geq 1, \tag{3.18}$$

$$p(0, -1) = (1 - \rho)\rho, \tag{3.19}$$

$$p(0, 0) = 1 - \rho, \tag{3.20}$$

where

$$C = \frac{1+\rho}{\rho^2} \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1+\rho - \alpha_{i,k}}, \tag{3.21}$$

and the α's, c's and d's are obtained recursively as follows:

$$\alpha_{i+1,2k-1} = \rho \alpha_{i,k} \ \text{and} \ \alpha_{i+1,2k} = \frac{\rho \alpha_{i,k}}{1+\rho - \alpha_{i,k}}, \tag{3.22}$$

$$c_{i+1,2k-1} = \frac{\rho^2(1-\alpha_{i+1,2k-1})}{\rho^2 - \alpha_{i+1,2k-1}} c_{i,k} \ \text{and} \ c_{i+1,2k} = -\frac{(1-\rho^2)\alpha_{i+1,2k}(\rho+\alpha_{i+1,2k})}{\rho(\rho^2 - \alpha_{i+1,2k})} c_{i,k}, \tag{3.23}$$

$$d_{i+1,2k-1} = \frac{(1-\rho)\alpha_{i+1,2k-1}}{\rho(1-\alpha_{i+1,2k-1})} c_{i+1,2k-1} \ \text{and} \ d_{i+1,2k} = \frac{\alpha_{i+1,2k}}{\rho+\alpha_{i+1,2k}} c_{i+1,2k}, \tag{3.24}$$

for $i = 0, 1, \ldots, \ k = 1, 2, 3, \ldots, 2^i$, with initial conditions

$$\alpha_{0,1} = \rho^2, \ c_{0,1} = 1, \ d_{0,1} = \frac{\rho}{1+\rho}.$$

Clearly, from this result we can derive similar expressions for other performance characteristics such as the mean queue lengths, the correlation between the queue lengths, the mean waiting time, etc. These results are discussed further in the last section containing the numerical results.

3.1 Step 1: Construction of the basis of product-forms

The first step in applying the compensation method is to determine a basis of product-form solutions satisfying the interior of the state space. We distinguish two types of product-form solutions: one corresponding to the positive (first) quadrant and a second one corresponding to the negative (fourth) quadrant. So we seek non-zero solutions of the form

$$p(m,n) \vDash \alpha^m \beta^n$$

satisfying Equation (3.5), where the double turnstile symbol ($\vDash$) is used to signify that $p(m,n)$ semantically entails the form $\alpha^m \beta^n$. Substitution of this product-form into (3.5) yields after some manipulations Equation (3.14). Note that the quadratic equation in α and in β, (3.14), has two solutions, namely

$$\alpha = \beta, \ \text{or} \ \alpha = \rho\beta. \tag{3.25}$$

Similarly, a solution of the form

$$p(m,-n) \vDash \alpha^m \beta^n$$

satisfies the equilibrium equations for the negative interior given in Equation (3.9) if

$$\alpha = \frac{\rho\beta}{1+\rho - \beta}. \tag{3.26}$$

The solutions of Equations (3.25) and (3.26) inside the region $|\alpha|, |\beta| < 1$ form the basis. We use these solutions to build the complete solution to all the balance equations.

3.2 Step 2: Construction of the formal solution

Having established a very rich basis of product-form solutions that satisfy the interior of the state space we now need to develop a methodology to select the product-forms that will construct the formal solution to the equilibrium distribution.

3.2.1 Initial solution

We start with a single product-form coming from the basis that simultaneously satisfies the horizontal boundary $n = 0$ and the interior of the state space. Let the product-form $c_0 \alpha_0^m \beta_0^n$ satisfy the positive interior, i.e. Equation (3.5). Hence, by Equation (3.25) either $\alpha_0 = \beta_0$ or $\alpha_0 = \rho \beta_0$. Subsequently, we want the product-form $c_0 \alpha_0^m \beta_0^n$ to also satisfy the horizontal boundary, i.e. Equation (3.3). Due to the terms $p(m, -1)$ and $p(m - 1, -1)$ appearing in Equation (3.3) we need to further assume that the product-form satisfying the negative interior for $n = -1$ shares the same α as the one in the positive interior, hence $p(m, -1) = d_0 \alpha_0^m$. Finally, setting $p(m, -n) = 0$ for $n \leq 2$ yields the unique starting solution.

Lemma 3.1. *The product-form*

$$
p(m, n) \vDash \begin{cases} c_0 \alpha_0^m \beta_0^n, & m \geq 1, n \geq 0, \\ d_0 \alpha_0^m, & m \geq 1, n = -1, \\ 0, & m \geq 0, n \leq -2, \end{cases}
$$

satisfies Equations (3.1), (3.3) *and* (3.5) *if:*

$$
\alpha_0 = \beta_0 = \rho^2, \ c_0 = 1, \ d_0 = \frac{\alpha_0}{\alpha_0 + \rho} c_0.
$$

3.2.2 Compensation on the vertical boundary

The initial term presented in Lemma 3.1 satisfies the equilibrium equations (3.1), (3.3) and (3.5) for the horizontal boundary and the interior of the state space. However, it violates Equation (3.7) and (3.8) for the vertical boundary, i.e. $\{(1, n) : n \geq 1\} \cup \{(0, -n) : n \geq 0\}$. Because we need a solution which satisfies all equilibrium equations, we add a compensation term to the initial term, such that the sum of these two terms satisfies the equations for the vertical boundary. Note that we need to add one product-term that compensates for the error on the positive part and another term that compensates for the error on the negative part.

Let us consider the initial term $c_0 \alpha_0^m \beta_0^n$, $m \geq 1, n \geq 0$, with $\alpha_0 = \beta_0$, and show how to compensate for the error of this term on the vertical boundary. We add a new term $c_1 \alpha_1^m \beta_1^n$, such that $c_0 \alpha_0^m \beta_0^n + c_1 \alpha_1^m \beta_1^n$ satisfies (3.7). Simultaneously, we add a new term $c_2 \alpha_2^m \beta_2^{-n}$ that satisfies (3.8).

Lemma 3.2. *The sum of products*

$$
p(m, n) \vDash \begin{cases} c_0 \alpha_0^m \beta_0^n + c_1 \alpha_1^m \beta_1^n, & m \geq 1, n \geq 0, \\ d_0 \alpha_0^m, & m \geq 1, n = -1, \\ c_2 \alpha_2^m \beta_2^{-n}, & m \geq 0, n \leq -2, \end{cases}
$$

satisfies Equation (3.5), (3.7), (3.8) *and* (3.9) *if:*

$$
\beta_0 = \beta_1 = \beta_2,
$$

α_0 *and* α_1 *are the companion roots of the quadratic equation* (3.14) *with* $\beta = \beta_0$, *so* $\alpha_0 = \beta_0$ *and* $\alpha_1 = \rho \beta_0$, α_2 *is the root of Equation* (3.15) *with* $\beta = \beta_0$, *so*

$$
\alpha_2 = \frac{\rho \beta_0}{1 + \rho - \beta_0},
$$

and

$$
c_1 = -c_0, \ c_2 = \frac{1 - \rho}{1 + \rho - \beta_0} c_0.
$$

The newly added compensation terms generate new errors on the horizontal boundary, so more compensation terms have to be added. To show the details of the construction of the compensation terms, we give an extensive description of the compensation step on the horizontal boundary in the following subsection. All other compensation terms are constructed in the same way.

3.2.3   Compensation on the horizontal boundary

The products $c_1\alpha_1^m\beta_1^n$, $m \geq 1, n \geq 0$, and $c_2\alpha_2^m\beta_2^{-n}$, $m \geq 0, n \leq -2$, with $\alpha_1 = \rho\beta_1$ and $\alpha_2 = \rho\beta_2/(1+\rho-\beta_2)$, can be easily seen not to satisfy the Equations (3.1) and (3.3). To compensate for the error we create on the horizontal boundary we add a new product-form expression. The compensation on the horizontal boundary is conducted in two steps. First we compensate for the error created by a term 'living' on the first quadrant. Afterwards, we compensate for the error created by a term 'living' on the fourth quadrant.

i) Compensation on the horizontal boundary of the product $c_1\alpha_1^m\beta_1^n$, $m > 0, n \geq 0$, with $\alpha_1 = \rho\beta_1$:
The sum of products

$$p(m,n) \vDash \begin{cases} c_0\alpha_0^m\beta_0^n + c_1\alpha_1^m\beta_1^n + c_3\alpha_3^m\beta_3^n, & m > 0, n \geq 0, \\ d_0\alpha_0^m + d_1\alpha_1^m, & m > 0, n = -1, \\ 0, & m \geq 0, n < -1, \end{cases}$$

satisfies (3.1), (3.3), (3.5) and (3.9) if:

$$\alpha_3 = \alpha_1,$$

where β_1 and β_3 are companion roots of the quadratic equation (3.14) with $\alpha = \alpha_1 = \beta_3$, so

$$c_3 = \frac{\rho^2(1-\beta_3)}{\beta_3 - \rho^2}c_1, \ d_1 = \frac{(1-\rho)\beta_3}{\rho(1-\beta_3)}c_3.$$

ii) Compensation on the horizontal boundary of the product $c_2\alpha_2^m\beta_2^{-n}$, $m \geq 0, n < -1$, with $\alpha_2 = \frac{\rho\beta_2}{1+\rho-\rho\beta_2}$ (or, equivalently, $\beta_2 = \frac{(1+\rho)\alpha_2}{\alpha_2+\rho}$):
The sum of products

$$p(m,n) \vDash \begin{cases} c_0\alpha_0^m\beta_0^n + c_1\alpha_1^m\beta_1^n + c_3\alpha_3^m\beta_3^n + c_4\alpha_4^m\beta_4^n, & m > 0, n \geq 0, \\ d_0\alpha_0^m + d_1\alpha_1^m + d_2\alpha_2^m, & m > 0, n = -1, \\ c_2\alpha_2^m\beta_2^{-n}, & m \geq 0, n < -1, \end{cases}$$

satisfies (3.1), (3.3), (3.5) and (3.9) if:

$$\alpha_4 = \alpha_2, \ \beta_4 = \alpha_4,$$

and

$$c_4 = \frac{(1+\rho)^2\beta_4}{\beta_4 - \rho^2} \, c_2, \ d_2 = \frac{\beta_4}{\beta_4+\rho}c_4.$$

Combining these two results we write the following.

Lemma 3.3. *The sum of products*

$$p(m,n) \vDash \begin{cases} c_0\alpha_0^m\beta_0^n + c_1\alpha_1^m\beta_1^n + c_3\alpha_3^m\beta_3^n + c_4\alpha_4^m\beta_4^n, & m \geq 1, n \geq 0, \\ d_0\alpha_0^m + d_1\alpha_1^m + d_2\alpha_2^m, & m \geq 1, n = -1, \\ c_2\alpha_2^m\beta_2^{-n}, & m \geq 0, n \leq -2, \end{cases}$$

11

satisfies Equations (3.1), (3.3), (3.5) *and* (3.9) *if:*

$$\alpha_3 = \beta_3 = \alpha_1, \ \alpha_4 = \beta_4 = \alpha_2,$$

where β_1 and β_3 are companion roots of the quadratic equation (3.14) *with $\alpha = \alpha_1$, so*

$$c_3 = \frac{\rho^2(1-\beta_3)}{\beta_3 - \rho^2} \ c_1, \quad d_1 = \frac{(1-\rho)\beta_3}{\rho(1-\beta_3)} c_3$$

and

$$c_4 = \frac{\beta_4(1+\rho)^2}{\beta_4 - \rho^2} \ c_2, \quad d_2 = \frac{\beta_4}{\beta_4 + \rho} c_4.$$

According to the procedure we have described, a tree of terms is generated as depicted in Figure 3. All products $c_i\alpha_i^m\beta_i^n$ 'live' in the positive (first) quadrant, while the products $c_i\alpha_i^m\beta_i^{-n}$ 'live' in the negative (fourth) quadrant.
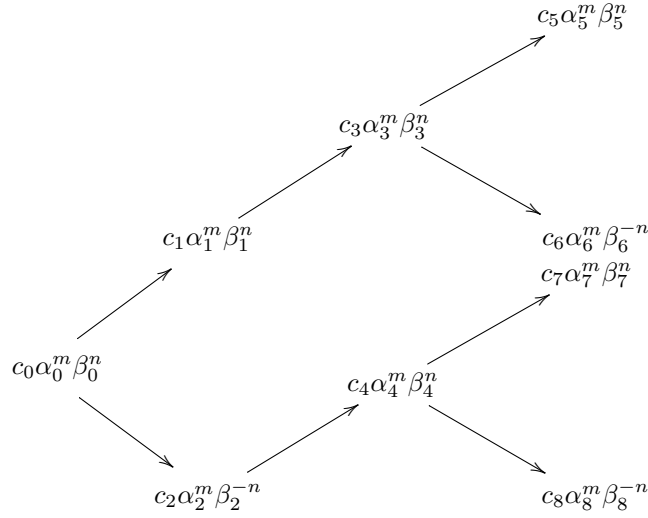

Figure 3: Tree evolution of product-forms.

3.3 Formal construction of the solution

In order to formally construct the tree of terms we observe that the terms that 'live' in the first quadrant form a binary tree as can be seen in Figure 4. Moreover, for the term $c_i\alpha_i^m\beta_i^n$ let $c_{u(i)}\alpha_{u(i)}^m\beta_{u(i)}^n$ denote its unique descendant in the upper quadrant. Then, we observe that $c_i\alpha_i^m\beta_i^n + c_{u(i)}\alpha_{u(i)}^m\beta_{u(i)}^{-n} = c_i(1-\rho^m)\alpha_i^m\beta_i^n$, since $\alpha_{u(i)} = \rho\alpha_i$, $\beta_{u(i)} = \beta_i$ and $c_{u(i)} = -c_i$.

For this reason at each step we use the following notation $c_{i,k}(1-\rho^m)\alpha_{i,k}^m\beta_{i,k}^n$, with $i = 0, 1, \ldots$ and $k = 1, 2, 3, \ldots, 2^i$. Note that the parameter i denotes the level on the tree and the parameter k stands for the specific branch. A schematic representation of the notation used is depicted in Figure 5.

Using this notation we can write the formal solution of the equilibrium equations. More explicitly, we have that the solution at the positive quadrant is up to a multiplicative constant equal to:

$$p(m,n) \propto (1-\rho^m)\sum_{i=0}^{\infty}\sum_{k=1}^{2^i} c_{i,k}\alpha_{i,k}^m\beta_{i,k}^n, \quad m \geq 1, n \geq 0. \tag{3.27}$$
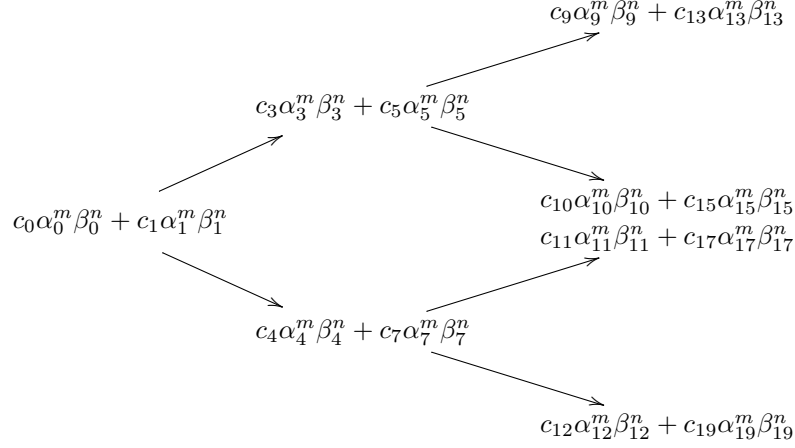
12

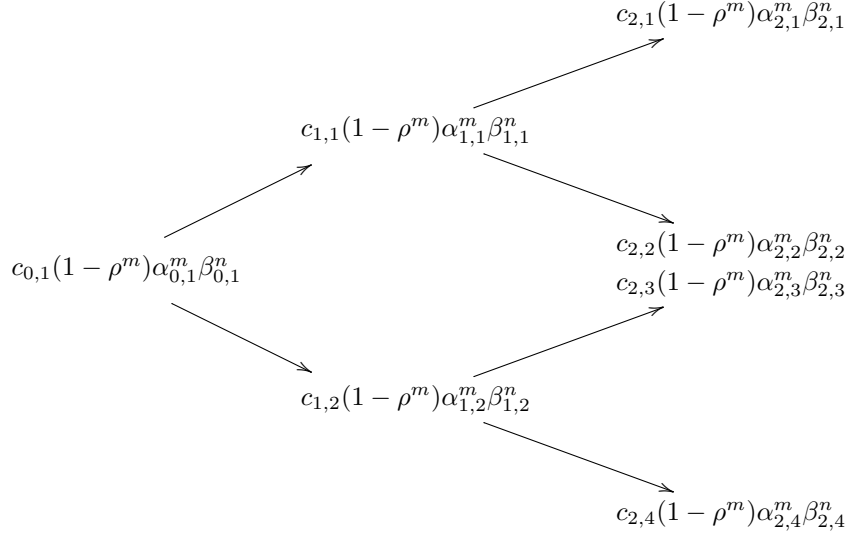Figure 4: Tree of product-forms for the positive quadrant.

Figure 5: Notation and schematic representation of product-form terms evolution at the positive quadrant.

Note that the product-form $c_{i,k}(1 - \rho^m)\alpha_{i,k}^m\beta_{i,k}^n$ has two children $c_{i+1,2k-1}(1 - \rho^m)\alpha_{i+1,2k-1}^m\beta_{i+1,2k-1}^n$ and $c_{i+1,2k}(1 - \rho^m)\alpha_{i+1,2k}^m\beta_{i+1,2k}^n$, that compensate for the error of the parent on the horizontal boundary. According to Lemma 3.3 we have that, for $i = 0, 1, \ldots,\ k = 1, 2, 3, \ldots, 2^i$,

$$\alpha_{i+1,2k-1} = \beta_{i+1,2k-1} = \rho\,\alpha_{i,k}, \tag{3.28}$$

$$\alpha_{i+1,2k} = \beta_{i+1,2k} = \frac{\rho\,\alpha_{i,k}}{1 + \rho - \alpha_{i,k}}, \tag{3.29}$$

and

$$c_{i+1,2k-1} = \frac{\rho^2(1 - \alpha_{i+1,2k-1})}{\rho^2 - \alpha_{i+1,2k-1}}c_{i,k}, \tag{3.30}$$

$$c_{i+1,2k} = -\frac{(1 - \rho^2)\alpha_{i+1,2k}(\rho + \alpha_{i+1,2k})}{\rho(\rho^2 - \alpha_{i+1,2k})}c_{i,k}, \tag{3.31}$$

13

with initial conditions

$$\alpha_{0,1} = \beta_{0,1} = \rho^2, \ c_{0,1} = 1.$$

Hence, Equation (3.27) can be simplified into

$$p(m,n) \propto (1 - \rho^m) \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \alpha_{i,k}^{m+n}, \ \ m \geq 1, n \geq 0. \tag{3.32}$$

Similarly, the solution at the negative quadrant is up to a multiplicative constant equal to:

$$p(m,-n) \propto \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} \frac{1 - \rho}{1 + \rho - \beta_{i,k}} c_{i,k} \alpha_{i+1,2k}^{m} \beta_{i,k}^{n}, \ \ m \geq 0, n \geq 2, \tag{3.33}$$

$$p(m,-1) \propto \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} d_{i,k} \alpha_{i,k}^{m}, \ \ m \geq 1. \tag{3.34}$$

Note that the solution (3.33) for the negative quadrant is immediate in light of Lemma 3.2. Moreover, the solution (3.34) follows from Lemma 3.3 which also yields, for $i = 0, 1, \ldots, \ k = 1, 2, 3, \ldots, 2^i$,

$$d_{i+1,2k-1} = \frac{(1-\rho)\alpha_{i+1,2k-1}}{\rho(1 - \alpha_{i+1,2k-1})} c_{i+1,2k-1}, \tag{3.35}$$

$$d_{i+1,2k} = \frac{\alpha_{i+1,2k}}{\rho + \alpha_{i+1,2k}} c_{i+1,2k}, \tag{3.36}$$

with

$$d_{0,1} = \frac{\alpha_{0,1}}{\rho + \alpha_{0,1}} c_{0,1}.$$

Taking into consideration that $\alpha_{i+1,2k} = \frac{\rho \alpha_{i,k}}{1+\rho-\alpha_{i,k}}$, Equation (3.33) assumes the form

$$p(m,-n) \propto (1 - \rho)\rho^m \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}^{m+n}}{(1 + \rho - \alpha_{i,k})^{m+1}}, \ \ m \geq 0, n \geq 2. \tag{3.37}$$

This leaves two probabilities: $p(0,0)$ and $p(0,-1)$. By using Equations (3.11) and (3.12), we obtain

$$p(0,-1) = (p(1,0) + p(0,-2))/\rho, \ \ p(0,0) = p(0,-1)/\rho.$$

Equivalently, we have that

$$p(0,-1) \propto \frac{1 - \rho^2}{\rho} \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1 + \rho - \alpha_{i,k}}$$

and

$$p(0,0) \propto \frac{1 - \rho^2}{\rho^2} \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1 + \rho - \alpha_{i,k}}. \tag{3.38}$$

There remains the task of proving that these formal expressions are in fact absolutely convergent and to obtain the normalization constant so as to ensure that the probabilities add up to 1.

3.4 Step 3: Absolute convergence of the solution

The analysis that follows is similar to the analysis performed in [6]. First, we analyze the limiting behavior of the sequences $\{\alpha_{i,k}\}_{i\in\mathbb{N}_0, k=1,\ldots,2^i}$ and the corresponding sequences of coefficients. The following proposition provides geometric bounds for the α's.

Proposition 3.1. *The $\{\alpha_{i,k}\}_{i\in\mathbb{N}_0, k=1,\ldots,2^i}$ appearing in (3.27)-(3.34) satisfy the properties*

i) $\rho^2 = \alpha_{0,1} \geq \max\{\alpha_{1,1}, \alpha_{1,2}\} \geq \max\{\alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}, \alpha_{2,4}\} \geq \ldots \geq 0$.

ii) $\alpha_{i,1} = \rho^{i+2}$ *and* $\alpha_{i,2^m} = \frac{\rho^{i+2}}{(1+\rho)^m(1-\rho^{i-m+2})+\rho^{i+2}}$, $i \geq 0$, $m = 0, 1, \ldots, i$.

iii) $\alpha_{i,2^i} \leq \alpha_{i,k} \leq \alpha_{i,1}$, $i \geq 0$, $k = 1, 2, 3, \ldots, 2^i$.

Proof. i) Remember that $\alpha_{0,1} = \rho^2$. Furthermore, $\alpha_{1,1} = \rho\alpha_{0,1}$ and $\alpha_{1,2} = \frac{\rho\alpha_{0,1}}{1+\rho-\alpha_{0,1}}$. Then, $0 \leq \alpha_{0,1} < \max\{\alpha_{1,1}, \alpha_{1,2}\} \leq \rho^2$, so that assertion (i) follows upon repeating this argument.

ii) First of all we observe that $\alpha_{i,1} = \rho\alpha_{i-1,1} = \rho^2\alpha_{i-2,1} = \ldots = \rho^i\alpha_{0,1} = \rho^{i+2}$. Furthermore, let $g(x) = \frac{\rho x}{1+\rho-x}$, then $\alpha_{i,2^m} = g(\alpha_{i-1,2^{m-1}}) = g^{(2)}(\alpha_{i-2,2^{m-2}}) = \ldots = g^{(m)}(\alpha_{i-m,1}) = g^{(m)}(\rho^{i-m+2}) = \frac{\rho^{i+2}}{(1+\rho)^m(1-\rho^{i-m+2})+\rho^{i+2}}$, where $g^{(i)}(x) = g^{(i-1)}(g(x)) = g(g^{(i-1)}(x))$ denotes the i-th composition of the function $g(\cdot)$.

iii) The proof follows easily after observing the structure of the α's as given in Equations (3.28) and (3.29) and assertion (ii). ∎

It follows from Proposition 3.1 that the sequence of $\{\alpha_{i,k}\}$ tends to zero as i tends to infinity. In the next lemma we state the geometrical asymptotic behavior of the α's and the coefficients.

Lemma 3.4. *As $i \to \infty$,*

$$\frac{\alpha_{i+1,2k-1}}{\alpha_{i,k}} \to \rho \quad and \quad \frac{\alpha_{i+1,2k}}{\alpha_{i,k}} \to \frac{\rho}{1+\rho}, \tag{3.39}$$

$$\frac{c_{i+1,2k-1}}{c_{i,k}} \to 1 \quad and \quad \frac{c_{i+1,2k}}{c_{i,k}} \to 0, \tag{3.40}$$

$$\frac{d_{i+1,2k-1}}{d_{i,k}} \to \rho^2 \quad and \quad \frac{d_{i+1,2k}}{d_{i,k}} \to 0. \tag{3.41}$$

Proof. The proof of the lemma is immediate in light of Equations (3.28), (3.29), (3.30), (3.31), (3.35) and (3.36). ∎

We are now in position to prove the absolute convergence of the series appearing in Equations (3.27)-(3.34) .

Theorem 3.2. i) *The series $\sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k}\alpha_{i,k}^{m+n}$, which defines $p(m,n)$ for $m \geq 1$, $n \geq 0$, converges absolutely for all $m \geq 1$, $n \geq 0$.*

ii) *The series $\sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k}\alpha_{i,k}^{m+n}(1+\rho-\alpha_{i,k})^{-m-1}$, which defines $p(m,-n)$ for $m \geq 0, n \geq 2$, converges absolutely for all $m \geq 0$, $n \geq 2$.*

iii) *The series $\sum_{i=0}^{\infty} \sum_{k=1}^{2^i} d_{i,k}\alpha_{i,k}^{m}$, which defines $p(m,-1)$ for $m \geq 1$, converges absolutely for all $m \geq 0$.*

iv) *The series $\sum_{m+n>0}(1-\rho^m)\sum_{i=0}^{\infty}\sum_{k=1}^{2^i} c_{i,k}\alpha_{i,k}^{m+n}$, $\sum_{m>0}\sum_{i=0}^{\infty}\sum_{k=1}^{2^i} d_{i,k}\alpha_{i,k}^{m}$ and $\sum_{m+n>0}\rho^m\sum_{i=0}^{\infty}\sum_{k=1}^{2^i} c_{i,k}\alpha_{i,k}^{m+n}(1+\rho-$ which define the $\sum_{m+n>0}(p(m,n)+p(m,-n))$, converge absolutely.*

Proof. For $m \geq 1$ and $n \geq 0$ the series that define $p(m,n)$ converge absolutely if

$$\sum_{i=0}^{\infty}\sum_{k=1}^{2^i} |c_{i,k}|\alpha_{i,k}^{m+n} < \infty.$$

15

Consider now a fixed $m \geq 1$ and $n \geq 0$. Then, define the ratio of two consecutive terms

$$R_u^{(i)}(m,n) = \frac{|c_{i+1,2k-1}|\,\alpha_{i+1,2k-1}^{m+n}}{|c_{i,k}|\,\alpha_{i,k}^{m+n}} \quad \text{and} \quad R_l^{(i)}(m,n) = \frac{|c_{i+1,2k}|\,\alpha_{i+1,2k}^{m+n}}{|c_{i,k}|\,\alpha_{i,k}^{m+n}},$$

with $k = 1, 2, 3, \ldots, 2^i$, $i = 0, 1, \ldots$. As $i \to \infty$, the corresponding α's tend to zero and we obtain that

$$R_u^{(i)}(m,n) \to R_u(m,n) = \rho^{m+n} \quad \text{and} \quad R_l^{(i)}(m,n) \to R_l(m,n) = 0.$$

Hence, the absolute convergence of the tree-structured series is determined by the spectral radius of the matrix of rates

$$R(m,n) = \begin{bmatrix} R_u(m,n) & R_u(m,n) \\ R_l(m,n) & R_l(m,n) \end{bmatrix},$$

in particular $\sigma(R(m,n)) = R_u(m,n) = \rho^{m+n}$, where $\sigma(\cdot)$ denotes the spectral radius of a matrix (see e.g. [6]). It is now apparent that for $m \geq 1$ and $n \geq 0$ we have that $\sigma(R(m,n)) < 1$, hence the corresponding series converges for all $m \geq 1$ and $n \geq 0$. The rest of the series can be shown to converge in a similar way. ∎

3.5 Step 4: Normalization constant

In order to complete our analysis it remains to obtain the normalization constant. This is the last step needed to complete the proof of our main result presented in Theorem 3.1.

Proof of Theorem 3.1. Exploiting the fact that the total number of customers in the system is distributed as the queue length distribution of the $M/M/1$, with arrival rate λ and service rate μ, we get

$$p(0,0) = 1 - \rho.$$

On the other hand we show in Equation (3.38) that

$$p(0,0) = C^{-1} \frac{1-\rho^2}{\rho^2} \sum_{i=0}^{\infty} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1+\rho-\alpha_{i,k}}.$$

Hence, we can directly obtain an expression for the normalization constant C, which concludes the proof. ∎

4 Boundary value problem

In this section we describe the generating function approach for analyzing the shorter queue polling model. For reasons of convenience we refer back to the original set of random variables describing our model, $\{X(t), Y(t)\}$ defined in Section 2.

We define the following PGF's:

$$\Pi_{\text{up}}(x,y) = \mathbb{E}[x^X y^Y \mathbb{1}_{\{X \leq Y\}}] = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \pi(i,j) x^i y^j + \pi(0,0),$$

$$\Pi_{\text{lo}}(x,y) = \mathbb{E}[x^X y^Y \mathbb{1}_{\{X > Y\}}] = \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \pi(i,j) x^i y^j,$$

$$\Pi(x,y) = \Pi_{\text{up}}(x,y) + \Pi_{\text{lo}}(x,y) = \mathbb{E}[x^X y^Y] = \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \pi(i,j) x^i y^j + \pi(0,0).$$

16

The balance equations of the model are given in (2.2), (2.3) and (2.4). Multiplying these equations with $x^i y^j$, x^i and 1, respectively, and summing over all (i,j) yields, after some manipulations, the following functional equation:

$$k_{\mathrm{up}}(x,y)\Pi_{\mathrm{up}}(x,y) + k_{\mathrm{lo}}(x,y)\Pi_{\mathrm{lo}}(x,y) \tag{4.1}$$
$$= \mu(x-1)\pi(0,0) + \mu x \left(\mathbb{E}[x^Y \mathbb{1}_{\{X=1\}}] - \mathbb{E}[y^Y \mathbb{1}_{\{X=1\}}] \right),$$

with

$$k_{\mathrm{up}}(x,y) = (\lambda+\mu)x - \mu - \lambda x^2 = (x-1)(\mu - \lambda x),$$
$$k_{\mathrm{lo}}(x,y) = (\lambda+\mu)x - \mu - \lambda xy.$$

The objective of this section is to demonstrate how to solve the functional equation (4.1). We briefly outline the notation needed and the steps of our approach, and then we point out some differences and similarities between this method and the compensation approach.

Remark 4.1. *Setting $x = y$ we immediately obtain that*

$$\Pi_{up}(x,x) + \Pi_{lo}(x,x) = \mathbb{E}[x^{X+Y}] = \frac{\pi(0,0)}{1-\rho x} = \frac{1-\rho}{1-\rho x}.$$

In analyzing the functional equation (4.1) we employ an idea similar to Wiener-Hopf factorization. First we introduce an appropriate translation and then we fix one of the two variables and define a smooth closed contour in the other variable, to be called the free variable. We next show that the functional equation presented in (4.1) can be separated into two parts, one analytic on the inside of the contour and one analytic on the outside of the contour, in terms of the free variable. Moreover, the partial PGFs appearing in both parts are bounded in terms of the free variable. Hence, using Liouville's Theorem we immediately obtain that the partial bivariate generating functions, Π_{up} and Π_{lo}, behave as "constant" in the free variable. In that sense we decompose the original functional equation into two functional equations, one for the upper part of quadrant, $\{(i,j) \in S_{X,Y} : i \leq j\}$, and one for the lower part of the quadrant, $\{(i,j) \in S_{X,Y} : i > j\}$, that we solve. The two new equations are easier to treat and their solution follows the steps described by Cohen in [13].

More explicitly, the two functional equations that we derive following the above approach and their solutions are stated in the following main Theorem.

Theorem 4.1. *For all x,y, with $|x|, |y| \leq 1$,*

$$(x-1)(\mu - \lambda x)\Pi_{up}(x,y) = (\mu - \lambda)(x-1) \tag{4.2}$$
$$+ \lambda(1-x)x\mathbb{E}[(xy)^Y \mathbb{1}_{\{X=Y\}}]$$
$$+ \mu x \left(\mathbb{E}[(xy)^Y \mathbb{1}_{\{X=1\}}] - \mathbb{E}[y^Y \mathbb{1}_{\{X=1\}}] \right),$$
$$((\lambda+\mu)x - \lambda xy - \mu)\Pi_{lo}(x,y) = -\lambda(1-x)x\mathbb{E}[(xy)^Y \mathbb{1}_{\{X=Y\}}] \tag{4.3}$$
$$- \mu x \left(\mathbb{E}[(xy)^Y \mathbb{1}_{\{X=1\}}] - \mathbb{E}[x^Y \mathbb{1}_{\{X=1\}}] \right),$$

where

$$\mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}] := \Omega(\tau) = \sum_{\ell=0}^{\infty} \sum_{i=1}^{2^\ell} \frac{\omega(\tau_{\ell,i})}{\tau - \tau_{\ell,i}}, \tag{4.4}$$

$$\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}] := B(\tau) = \sum_{\ell=0}^{\infty} \sum_{i=1}^{2^\ell} \frac{b(\tau_{\ell,i})}{\tau - \tau_{\ell,i}}, \tag{4.5}$$

17

with

$$\tau_{\ell+1,2i-1} = \tau_{\ell,i}/\rho, \qquad \tau_{\ell+1,2i} = ((1+\rho)\tau_{\ell,i} - 1)/\rho,$$

and $\tau_{0,1} = 1/\rho^2$. Moreover, the residues are calculated recursively by

$$\omega(\tau_{\ell+1,2i-1}) = \frac{\rho(\tau_{\ell+1,2i-1} - 1)}{\rho^2\,\tau_{\ell+1,2i-1} - 1}\omega(\tau_{\ell,i}),$$

$$\omega(\tau_{\ell+1,2i}) = \frac{1+\rho}{\rho}\,\frac{1-\rho^2}{1-\rho^2(\tau_{\ell+1,2i}+1)}\omega(\tau_{\ell,i}),$$

and

$$b(\tau_{\ell+1,2i-1}) = \frac{1+\rho}{\rho^2(\tau_{\ell+1,2i-1} - 1)}\omega(\tau_{\ell+1,2i-1}),$$

$$b(\tau_{\ell+1,2i}) = \frac{1}{1-\rho}\omega(\tau_{\ell+1,2i}).$$

All residues containing the factor $\omega(\tau_{0,1})$ are uniquely determined by

$$\Omega(0) = \rho(1-\rho) = -\sum_{\ell=0}^{\infty}\sum_{i=1}^{2^{\ell}}\frac{\omega(\tau_{\ell,i})}{\tau_{\ell,i}}, \qquad b(\tau_{0,1}) = \frac{1}{1-\rho}\omega(\tau_{0,1}).$$

In particular

$$(\mu - \lambda)\Pi_{up}(1,y) = \mu - \lambda - \lambda B(y) + \mu\Omega'(y), \tag{4.6}$$

$$\lambda(1-y)\Pi_{lo}(1,y) = \mu\left(\Omega(1) - \Omega(y)\right), \tag{4.7}$$

$$\Pi_{lo}(1,1) = \frac{\lambda}{\mu}B(1) = \frac{\mu}{\lambda}\Omega'(1). \tag{4.8}$$

4.1 Construction of two functional equations

Taking the translation $x = uv$ and $y = u/v$, with $|v| = 1$, we demonstrate that the bivariate PGF can be "decomposed" into two functions, one analytic inside the smooth contour $|v| = 1$ and one analytic outside $|v| = 1$, each one constituting the analytic continuation of the other through the contour $|v| = 1$. Via the translation $x = uv$ and $y = u/v$, Equation (4.1) assumes the form

$$\psi(u,v) := ((\lambda + \mu)uv - \mu - \lambda u^2 v^2)\mathbb{E}[u^{X+Y}v^{X-Y}\mathbb{1}_{\{X \le Y\}}] \tag{4.9}$$
$$+ \mu uv\mathbb{E}[u^Y v^{-Y}\mathbb{1}_{\{X=1\}}]$$
$$= -((\lambda + \mu)uv - \mu - \lambda u^2)\mathbb{E}[u^{X+Y}v^{X-Y}\mathbb{1}_{\{X > Y\}}] \tag{4.10}$$
$$+ \mu uv\mathbb{E}[u^Y v^Y\mathbb{1}_{\{X=1\}}] + \mu(uv - 1)\pi(0,0).$$

We observe that the terms in (4.9) are regular for $|v| > 1$ and continuous for $|v| \ge 1$. Moreover, for $|u| \le 1$ and $|v| \ge 1$, we have

$$0 \le \left|\mathbb{E}[u^{X+Y}v^{X-Y}\mathbb{1}_{\{X \le Y\}}]\right| \le 1.$$

On the other hand, the terms in (4.10) are regular for $|v| < 1$ and continuous for $|v| \le 1$. Moreover, for $|u| \le 1$ and $|v| \le 1$, we have

$$0 \le \left|\mathbb{E}[u^{X+Y}v^{X-Y}\mathbb{1}_{\{X > Y\}}]\right| \le 1.$$

Consequently $\psi(u,v)$ is regular in the whole v-plane and because the terms in (4.9) behave as $|v|^2$ for $|v| \to \infty$ it follows from Liouville's Theorem that $\psi(u,v)$ is a polynomial of the second degree in v. More specifically, for $|u| \le 1$

$$\psi(u,v) = A_2(u)v^2 + A_1(u)v + A_0(u),$$

with $A_2(u)$, $A_1(u)$ and $A_0(u)$ independent of v. Considering in Equations (4.9) and (4.10) the series expansions in powers of v and of v^{-1} and equating the coefficients of equal powers immediately yields

$$A_2(u) = -\lambda u^2 \mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y\}}],$$
$$A_1(u) = (\lambda u^2 + \mu)u\mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y+1\}}] + u(\lambda + \mu)\pi(0,0),$$
$$A_0(u) = -\mu\pi(0,0).$$

Then, substituting these last expressions for $A_2(u)$, $A_1(u)$ and $A_0(u)$ in (4.9) and (4.10) yields

$$h_{\mathrm{up}}(u,v)\mathbb{E}[u^{X+Y}v^{X-Y} \mathbb{1}_{\{X \le Y\}}] = -\mu uv\mathbb{E}[u^Y v^{-Y} \mathbb{1}_{\{X=1\}}] - \lambda u^2 v^2\mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y\}}] \tag{4.11}$$
$$+ (\lambda u^2 + \mu)uv\mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y+1\}}]$$
$$+ ((\lambda + \mu)uv - \mu)\pi(0,0)$$

and

$$h_{\mathrm{lo}}(u,v)\mathbb{E}[u^{X+Y}v^{X-Y} \mathbb{1}_{\{X>Y\}}] = \mu uv\mathbb{E}[u^Y v^Y \mathbb{1}_{\{X=1\}}] + \lambda u^2 v^2\mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y\}}] \tag{4.12}$$
$$- (\lambda u^2 + \mu)uv\mathbb{E}[u^{2Y} \mathbb{1}_{\{X=Y+1\}}] + \lambda uv\pi(0,0),$$

with

$$h_{\mathrm{up}}(u,v) = ((\lambda + \mu)uv - \mu - \lambda u^2 v^2),$$
$$h_{\mathrm{lo}}(u,v) = ((\lambda + \mu)uv - \mu - \lambda u^2).$$

From the above two functional equations, (4.11) and (4.12), it is seen that the determination of the bivariate PGF $\Pi(x,y) = \mathbb{E}[x^X y^Y]$ requires the construction of the functions $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y\}}]$, $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y+1\}}]$ and $\mathbb{E}[z^Y \mathbb{1}_{\{X=1\}}]$, which should satisfy the following conditions:

i) they are not identically zero,

ii) they are regular for $|z| < 1$ and the sum of coefficients in their series expansions in powers z^n converges absolutely, and

iii) they satisfy (4.1), (4.11) and (4.12).

Once the functions $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y\}}]$, $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y+1\}}]$ and $\mathbb{E}[z^Y \mathbb{1}_{\{X=1\}}]$ have been constructed and satisfy the conditions (i) - (iii), the bivariate PGF, $\Pi(x,y)$, is determined via (4.11) and (4.12). For this we take advantage of the zeros of the two kernels $h_{\mathrm{up}}(u,v)$ and $h_{\mathrm{lo}}(u,v)$.

Observing that Equation (4.11) involves the product uv and the variable v, and the kernel $h_{\mathrm{up}}(u,v)$ only involves the product uv, we introduce a new translation: $\tau = u^2$ and $\sigma = u/v$. This translation is selected to facilitate our calculations and to reveal the correspondence of this approach with the compensation approach. Then, Equation (4.11) under the new translation assumes the form

$$h_{\mathrm{up}}(\tau,\sigma)\mathbb{E}[(\tau/\sigma)^X \sigma^Y \mathbb{1}_{\{X \ge Y\}}] = -\mu\sigma\mathbb{E}[\sigma^Y \mathbb{1}_{\{X=1\}}] - \lambda\tau\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}] \tag{4.13}$$
$$+ (\lambda\tau + \mu)\sigma\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y+1\}}]$$
$$+ \lambda\tau\pi(0,0),$$

19

with

$$h_{\mathrm{up}}(\tau,\sigma) := (\lambda+\mu)\tau\sigma - \mu\sigma^2 - \lambda\tau^2 = (\sigma-\tau)(\lambda\tau - \mu\sigma).$$

Moreover, Equation (4.12) for $\tau = u^2$ and $\zeta = uv$ assumes the form

$$h_{\mathrm{lo}}(\tau,\zeta)\mathbb{E}[\zeta^X(\tau/\zeta)^Y \mathbb{1}_{\{X<Y\}}] = -\mu\mathbb{E}[\zeta^Y \mathbb{1}_{\{X=1\}}] - \lambda\zeta\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}] \tag{4.14}$$
$$+ (\lambda\tau+\mu)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y+1\}}] + \lambda\pi(0,0),$$

with

$$h_{\mathrm{lo}}(\tau,\zeta) := (\lambda+\mu)\zeta - \lambda\tau - \mu.$$

Exploiting the zeros of the two kernels $h_{\mathrm{up}}(\tau,\sigma)$ and $h_{\mathrm{lo}}(\tau,\zeta)$ provides us with all the additional information we need to determine the unknown functions $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y\}}]$, $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y+1\}}]$ and $\mathbb{E}[z^Y \mathbb{1}_{\{X=1\}}]$.

The first step of our analysis is to substitute $\tau = \sigma$ in Equation (4.13). This creates a key functional equation that connects the two unknown functions with the third one, reducing the number of unknown functions to two. More explicitly, setting $\tau = \sigma$ in Equation (4.13) yields the following key equation:

$$\mu\mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}] + \lambda\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}] = (\lambda\tau+\mu)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y+1\}}] + \lambda\pi(0,0). \tag{4.15}$$

Equations (4.13) and (4.14) now simplify to the following set of equations:

$$h_{\mathrm{up}}(\tau,\sigma)\mathbb{E}[(\tau/\sigma)^X\sigma^Y \mathbb{1}_{\{X\geq Y\}}] = -\mu\sigma\big(\mathbb{E}[\sigma^Y \mathbb{1}_{\{X=1\}}] + \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}]\big) \tag{4.16}$$
$$- \lambda(\tau-\sigma)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}]$$
$$+ \lambda(\tau-\sigma)\pi(0,0),$$
$$h_{\mathrm{lo}}(\tau,\zeta)\mathbb{E}[\zeta^X(\tau/\zeta)^Y \mathbb{1}_{\{X<Y\}}] = -\mu\big(\mathbb{E}[\zeta^Y \mathbb{1}_{\{X=1\}}] + \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}]\big)$$
$$+ \lambda(1-\zeta)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}]. \tag{4.17}$$

Note that, after a change of variables, Equations (4.16) and (4.17) exactly correspond to Equations (4.2) and (4.3). Moreover, looking closely at Equations (4.16) and (4.17) it is evident that the two unknown functions, $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y\}}]$ and $\mathbb{E}[z^Y \mathbb{1}_{\{X=1\}}]$, share the same poles on $|z| > 1$. We can meromorphically continue the two functions, $\mathbb{E}[z^Y \mathbb{1}_{\{X=1\}}]$ and $\mathbb{E}[z^Y \mathbb{1}_{\{X=Y\}}]$, out of $|z| \leq 1$ and into $|z| > 1$ going from one pole to the next one. This analysis is identical to the one performed in [13] and for this reason is omitted. However, for reasons of self-containedness we show in the following subsection how the two functions are meromorphically continued outside the unit disk. Furthermore, in the next subsection we explain how to obtain all the poles formulating a recursion based on the two kernels and how to calculate the residues of the unknown functions at these poles, which permits us to establish the power series expansions of the two unknown series.

4.2 Properties of the zero tuples of the kernels

Setting $\sigma = \lambda\tau/\mu = \rho\tau$ in Equation (4.16) and $\zeta = (\mu+\lambda\tau)/(\mu+\lambda) = (1+\rho t)/(1+\rho)$ in Equation (4.17), where $\rho = \lambda/\mu$, yields

$$\mathbb{E}[(\rho\tau)^Y \mathbb{1}_{\{X=1\}}] - \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}] = -(1-\rho)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}] \tag{4.18}$$
$$+ (1-\rho)\pi(0,0),$$
$$\mathbb{E}[(\frac{1+\rho\tau}{1+\rho})^Y \mathbb{1}_{\{X=1\}}] - \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}] = \frac{\rho^2}{1+\rho}(1-\tau)\mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}]. \tag{4.19}$$

20

We define

$$\Omega(\tau) = \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}],$$
$$B(\tau) = \mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}].$$

Observe now that Equations (4.18) and (4.19) are functional equations in one variable:

$$\Omega(\rho\tau) - \Omega(\tau) = -(1-\rho)B(\tau) + (1-\rho)\pi(0,0), \tag{4.20}$$

$$\Omega\Big(\frac{1+\rho\tau}{1+\rho}\Big) - \Omega(\tau) = \frac{\rho^2}{1+\rho}(1-\tau)B(\tau). \tag{4.21}$$

Remark 4.2. *One direct approach is to substitute the second equation for $B(\tau)$ into the first equation. This leads to the following functional equation in one variable:*

$$\Omega(\tau) = -\frac{(1-\rho)(1-\tau)}{1-\rho^2\tau}\pi(0,0) + \frac{1-\rho^2}{1-\rho^2\tau}\Omega\Big(\frac{1+\rho\tau}{1+\rho}\Big) + \frac{\rho^2(1-\tau)}{1-\rho^2\tau}\Omega(\rho\tau). \tag{4.22}$$

Observe that the terms in the functional equation for $\Omega(\tau)$, cf. Equation (4.22), are by definition regular in $|\tau| < 1$ and continuous in $|\tau| \le 1$. It is easily seen that $\Omega(\tau)$ can be meromorphically continued for all τ with $|\tau| < 1/\rho^2$. Then, $\tau = 1/\rho^2$ is the smallest pole of $\Omega(\tau)$. The next two poles are obtained by demanding $\rho\tau = 1/\rho^2$ and $\frac{1+\rho\tau}{1+\rho} = 1/\rho^2$. Continuing in this manner we obtain all poles of the function $\Omega(\tau)$ and we can recursively show that the function can be meromorphically continued going from one pole to the next one.

Moreover it is interesting to note that the functional equation (4.22) can be solved with two methods, either calculating the residues recursively or by a double iteration. The latter method produces a solution that grows like a binary tree but unfortunately it seems very difficult to write the solution as a power series expression so as to compare it with the compensation method and for this reason we do not proceed further with this solution. The interested reader is referred to Fayolle et al. [15] and Jacquet and Merle [22] for more details on solving this type of functional equations.

Let us now return to the system of functional equations given in (4.20) and (4.21). Let T be a pole of $\Omega(\tau)$ and $B(\tau)$, then we define the residues of these two functions at the pole T as follows

$$\omega(T) = \lim_{\tau \to T}(\tau - T)\Omega(\tau),$$
$$b(T) = \lim_{\tau \to T}(\tau - T)B(\tau).$$

Consider now the pole closest to the origin, i.e., the pole with the smallest absolute value, say $\mathcal{T}$. Then multiplying both sides of Equations (4.20) and (4.21) with $\tau - \mathcal{T}$ and taking the limit as $\tau \to \mathcal{T}$ yields

$$-\omega(\mathcal{T}) + (1-\rho)b(\mathcal{T}) = -\frac{1}{\rho}\omega(\rho\mathcal{T}), \tag{4.23}$$

$$-\omega(\mathcal{T}) - \frac{\rho^2}{1+\rho}(1-\mathcal{T})b(\mathcal{T}) = -\frac{1+\rho}{\rho}\omega\Big(\frac{1+\rho\mathcal{T}}{1+\rho}\Big). \tag{4.24}$$

Observe that

   i) $|\rho\mathcal{T}| < |\mathcal{T}|$, for all $\mathcal{T} \in \mathbb{C} \smallsetminus \{0\}$, and

   ii) $\big|\frac{1+\rho\mathcal{T}}{1+\rho}\big| < |\mathcal{T}|$, for all $|\mathcal{T}| > 1$.

Moreover, we assumed that $\mathcal{T}$ is the 'smallest' pole, hence $\omega(\rho\mathcal{T}) = \omega\big(\frac{1+\rho\mathcal{T}}{1+\rho}\big) = 0$. Then the system consisting of Equations (4.23) and (4.24) is dependent and the corresponding determinant of the coefficients reveals the 'smallest' pole which is the departure point of our recursion. More specifically
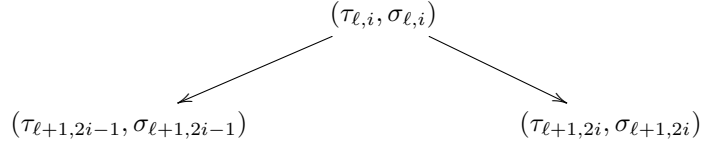
$$\det \begin{bmatrix} -1 & 1-\rho \\ -1 & -\frac{\rho^2}{1+\rho}(1-\mathcal{T}) \end{bmatrix} = 0$$

reveals that the starting pole is $\mathcal{T} = 1/\rho^2$.

4.2.1 Evolution of the poles and notation

Figure 6 depicts the two linear functions for $h_{\text{up}}(\tau, \sigma) = 0$ (dashed and solid lines) and the linear function $h_{\text{lo}}(\tau, \zeta) = 0$ (dotted line).

Let $(\tau_{\ell,i}, \sigma_{\ell,i})$, $\ell = 0, 1, \ldots$, $i = 1, 2, \ldots, 2^{\ell}$, be a zero-tuple of $h_{\text{up}}(\tau, \sigma) = 0$. Starting from $(\tau_{\ell,i}, \sigma_{\ell,i})$ we construct a series of zero tuples

$$(\tau_{\ell,i}, \sigma_{\ell,i})$$

$$(\tau_{\ell+1,2i-1}, \sigma_{\ell+1,2i-1}) \qquad\qquad (\tau_{\ell+1,2i}, \sigma_{\ell+1,2i})$$

recursively as follows

i) $\tau_{\ell,i} = \sigma_{\ell,i}$

ii) $\tau_{\ell+1,2i-1} = \tau_{\ell,i}/\rho$

ii) $\tau_{\ell+1,2i} = ((1+\rho)\tau_{\ell,i} - 1)/\rho$

The sequence of $(\tau_{\ell+1,2i-1}, \sigma_{\ell+1,2i-1})$ is referred to as the ladder generated by $\tau_{\ell,i}$ on $h_{\text{up}}(\tau, \sigma)$. The "up"-ladder is unbounded, while the "down"-ladder is finite and stops on the first index for which either $\tau < 1/\rho^2$ or $\sigma < 1/\rho^2$. The zero-tuple $(\tau_{\ell,i}, \sigma_{\ell,i})$ induces also a new tuple on $h_{\text{lo}}(\tau, \zeta)$, which we denoted by the sequence $(\tau_{\ell+1,2i}, \sigma_{\ell+1,2i})$. Hence, a zero-tuple produces exactly two terms: the even one (the left on the schematic representation above) which is created by the kernel $h_{\text{up}}(\tau, \sigma)$ and the odd one (the right one) which is created by the kernel $h_{\text{lo}}(\tau, \zeta)$. In this sense, for every pair $(\tau_{\ell,i}, \sigma_{\ell,i})$ we construct a similar tree structure as the one depicted in Figure 4.
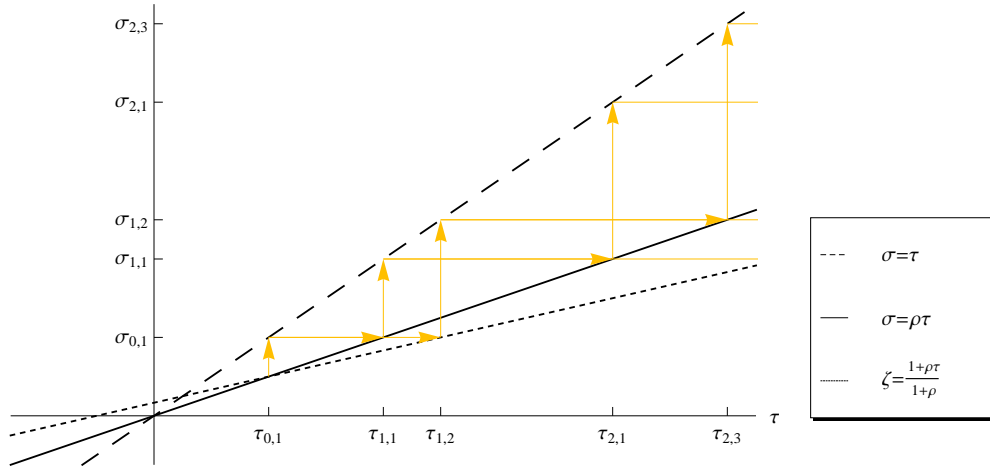

Figure 6: Zero tuples of $h_{\text{up}}(\tau, \sigma) = 0$ and $h_{\text{lo}}(\tau, \zeta) = 0$ for $\rho = 0.5$.

4.3   Solution of the functional equations

For the starting pole $\tau_{0,1} = 1/\rho^2$ the system of residues described in Equations (4.23) and (4.24) yields

$$-\omega(\tau_{0,1}) + (1 - \rho)b(\tau_{0,1}) = 0, \tag{4.25}$$

$$-\omega(\tau_{0,1}) - \frac{\rho^2}{1 + \rho}(1 - \tau_{0,1})b(\tau_{0,1}) = 0, \tag{4.26}$$

where the RHS of both (4.25) and (4.26) are zero because $\omega(\rho\tau_{0,1}) = \omega\left(\frac{1+\rho\tau_{0,1}}{1+\rho}\right) = 0$. Observe that the system (4.25) and (4.26) consists of two linearly dependent equations, hence we obtain

$$b(\tau_{0,1}) = \frac{1}{1 - \rho}\omega(\tau_{0,1}). \tag{4.27}$$

We continue now with the poles of the form $\tau_{\ell+1,2i-1}$, $\ell = 0, 1, \ldots$, $i = 1, 2, \ldots, 2^\ell$. Then the system of residues described in Equations (4.23) and (4.24) yields

$$-\omega(\tau_{\ell+1,2i-1}) + (1 - \rho)b(\tau_{\ell+1,2i-1}) = -\frac{1}{\rho}\omega(\tau_{\ell,i}), \tag{4.28}$$

$$-\omega(\tau_{\ell+1,2i-1}) - \frac{\rho^2}{1 + \rho}(1 - \tau_{\ell+1,2i-1})b(\tau_{\ell+1,2i-1}) = 0. \tag{4.29}$$

Note that $\tau_{\ell+1,2i-1} = \tau_{\ell,i}/\rho$. Plugging Equation (4.29) for $b(\cdot)$ into (4.28) reduces the system of equations to

$$\omega(\tau_{\ell+1,2i-1}) = \frac{1}{\rho}\frac{1}{1 - \frac{1-\rho^2}{\rho^2}\frac{1}{\tau_{\ell+1,2i-1}-1}}\omega(\tau_{\ell,i}),$$

$$b(\tau_{\ell+1,2i-1}) = \frac{1 + \rho}{\rho^2(\tau_{\ell+1,2i-1} - 1)}\omega(\tau_{\ell+1,2i-1}).$$

Performing a similar analysis for $\tau_{\ell+1,2i} = ((1 + \rho)\tau_{\ell,i} - 1)/\rho$ yields

$$-\omega(\tau_{\ell+1,2i}) + (1 - \rho)b(\tau_{\ell+1,2i-1}) = 0, \tag{4.30}$$

$$-\omega(\tau_{\ell+1,2i}) - \frac{\rho^2}{1 + \rho}(1 - \tau_{\ell+1,2i-1})b(\tau_{\ell+1,2i-1}) = -\frac{1 + \rho}{\rho}\omega(\tau_{\ell,i}), \tag{4.31}$$

which in its turn simplifies to

$$\omega(\tau_{\ell+1,2i}) = \frac{1 + \rho}{\rho}\frac{1 - \rho^2}{1 - \rho^2(\tau_{\ell+1,2i} + 1)}\omega(\tau_{\ell,i}),$$

$$b(\tau_{\ell+1,2i}) = \frac{1}{1 - \rho}\omega(\tau_{\ell+1,2i}).$$

Note, now, that we built an explicit expression for the functions $\Omega(\tau) = \mathbb{E}[\tau^Y \mathbb{1}_{\{X=1\}}]$ and $B(\tau) = \mathbb{E}[\tau^Y \mathbb{1}_{\{X=Y\}}]$ by calculating their residues moving from one pole to the next one, such that they

i) are not identically zero,

ii) satisfy the conditions (i)-(iii) described in Subsection 4.1.

We can now determine $\mathbb{E}[x^X y^Y]$, $|x|, |y| < 1$, via the generating functions $\mathbb{E}[x^X y^Y \mathbb{1}_{\{X \leq Y\}}]$ and $\mathbb{E}[x^X y^Y \mathbb{1}_{\{X>Y\}}]$. Then it follows that the Kolmogorov equations for the equilibrium distribution possess an absolutely convergent non null solution. By applying the well known Foster criterion, cf. [21], it follows that the joint queue length process (with irreducible state space) is positive recurrent and further that there is only one solution, which satisfies (4.1) and the normalizing condition.

5   Numerical results

For our numerical results we proceed as follows. First we calculate the following truncated series

$$x_K(m,n) = (1 - \rho^m) \sum_{i=0}^{K} \sum_{k=1}^{2^i} c_{i,k} \alpha_{i,k}^{m+n}, \ m \geq 1, \ n \geq 0, \tag{5.1}$$

$$x_K(m,-n) = (1 - \rho)\rho^m \sum_{i=0}^{K} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}^{m+n}}{(1 + \rho - \alpha_{i,k})^{m+1}}, \ m \geq 0, \ n \geq 2, \tag{5.2}$$

$$x_K(m,-1) = \sum_{i=0}^{K} \sum_{k=1}^{2^i} d_{i,k} \alpha_{i,k}^{m}, \ m \geq 1, \tag{5.3}$$

$$x_K(0,-1) = \frac{1 - \rho^2}{\rho} \sum_{i=0}^{K} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1 + \rho - \alpha_{i,k}}, \tag{5.4}$$

$$x_K(0,0) = \frac{1 - \rho^2}{\rho} \sum_{i=0}^{K} \sum_{k=1}^{2^i} c_{i,k} \frac{\alpha_{i,k}}{1 + \rho - \alpha_{i,k}}. \tag{5.5}$$

Observe that these quantities are immediately related to the probabilities $p(m,n)$ as given by Theorem 3.1. To get the corresponding probabilities we need to normalize and for this reason we define

$$norm := \frac{1}{\sum\limits_{m+|n| \leq b} x_K(m,n)}; \tag{5.6}$$

an alternative approach is to truncate the normalizing constant C^{-1} defined in Equation (3.21). Then, $norm \cdot x_K(m,n)$ is "equal" to the probability $p(m,n)$ for a carefully selected truncation level K and a sufficiently big number of states given by the level b. By employing this simplistic technique we observe that the closer we are to the origin (state $(0,0)$) and the bigger the value of ρ is, the larger the error we make in the calculation of the equilibrium distribution. On the other hand, the further away we are from the origin the more accurate our results are, even for values of ρ close to 1. Of course the accuracy of our results depends on the truncation level K; by increasing K we reduce the error, hence we improve the accuracy.

At this point we should also mention how we can calculate the accuracy of our numerical approach. Remember that $p(0,0) = 1 - \rho$, hence we define the relative accuracy as

$$\left| 1 - \frac{|norm \cdot x_K(0,0) - (1 - \rho)|}{1 - \rho} \right| 100\%.$$

According to our numerical findings for $\rho \geq 0.65$ the relative accuracy is below 50% even for large values of K, such as $K = 27$.

The problems of numerical accuracy are solved by exploiting the fact that

$$p(m,n) \approx norm \cdot x_K(m,n) \tag{5.7}$$

for m or n sufficiently big, as implied by Theorem 3.2. This is due to the fast convergence of the corresponding series further away from the origin, while the convergence of the series for states near the origin might be slower. So we select states far away from the origin and use Equation (5.7) to approximate their equilibrium distribution. Then, for the states close to the origin we solve numerically the linear system of balance equations. More explicitly:
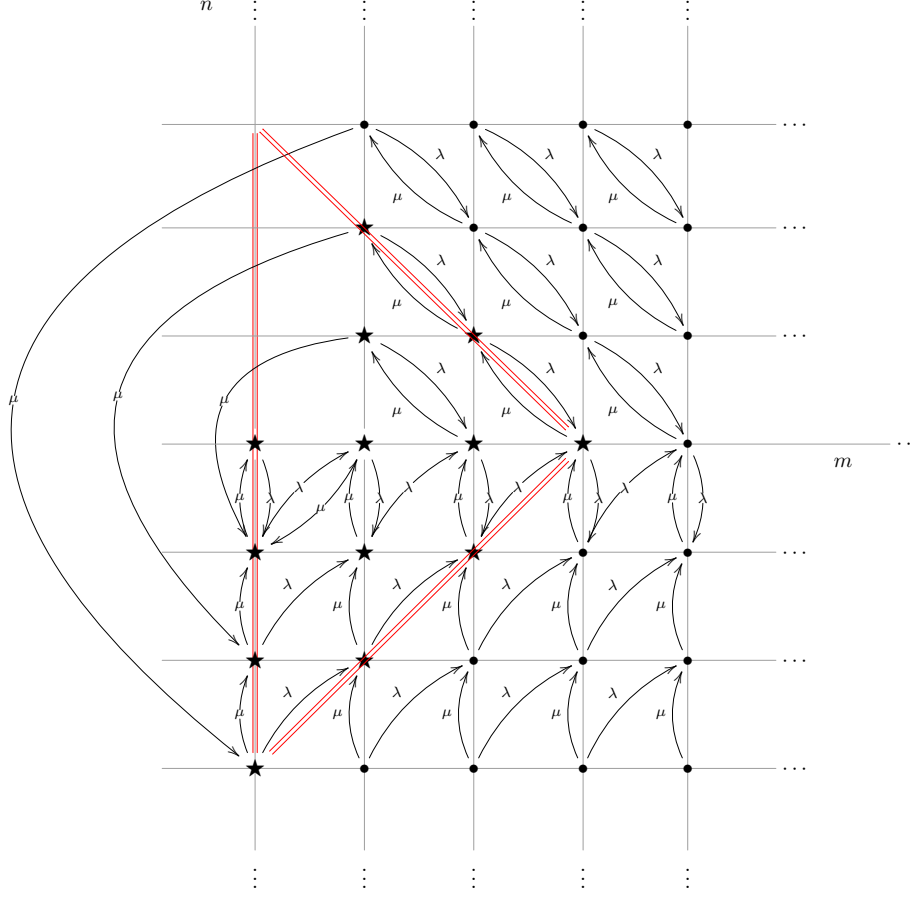
Figure 7: The triangular area containing the states (m, n), with $m + |n| \leq B$, for $B = 3$.

Step 1 We select a number of states around the origin. In particular, we select all states (m, n) such that $m + |n| \leq B$ for a positive integer B. These states are marked with a star in Figure 7, where we choose $B = 3$.

Step 2 Calculate the quantities $x_K(m, n)$ for all states (m, n) such that $B < m + |n| \leq b$, as given by Equations (5.1)-(5.5) for a specific truncation level K.

Step 3 Solve the linear system of the balance equations described in (3.1)-(3.13) for all states (m, n), with $m + |n| \leq B$ and obtain the corresponding values for $x_K(m, n)$.

Step 4a Normalize the quantities $x_K(m, n)$ for all m, n $(0 \leq m + |n| \leq b)$ as follows:

$$p_K(m, n) = norm \cdot x_K(m, n).$$

Define the relative accuracy as

$$\left| 1 - \frac{|p_K(0, 0) - (1 - \rho)|}{1 - \rho} \right| 100\%.$$

Step 4b Rescale the quantities $x_K(m, n)$ for all m, n $(0 \leq m + |n| \leq b)$ as follows

$$p_K(m, n) = x_K(m, n) \frac{1 - \rho}{x_K(0, 0)}.$$

ρ	p(0,0)	p(1,0)	p(0,-1)	p(1,1)	p(2,0)	p(1,-1)	p(0,-2)	relative inaccuracy
0.3	0.7000	0.0597	0.2100	0.0040	0.0052	0.0146	0.0033	$2.3929 \times 10^{-8}\,\%$
0.5	0.5000	0.1116	0.2500	0.0173	0.0260	0.0424	0.0134	$1.1419 \times 10^{-6}\,\%$
0.7	0.3000	0.1236	0.2100	0.0314	0.0533	0.0632	0.0234	$9.6145 \times 10^{-4}\,\%$
0.9	0.0999	0.0644	0.0899	0.0229	0.0435	0.0414	0.0165	$6.6667 \times 10^{-2}\,\%$

Table 1: Calculation of the equilibrium distribution.

Define the relative accuracy as

$$\left| 1 - \left| \sum_{m+|n| \leq b} p_K(m,n) - 1 \right| \right| 100\%. \tag{5.8}$$

Observe that the normalization approach described in Step (4a) and the scaling approach described in Step (4b) are both very accurate and the only actual difference between them is that the first one gives us the relative accuracy of a single term, $p(0,0)$, while the second approach gives us an overall accuracy by taking into account the total error made in all the calculations for all the probabilities. For a sufficiently large b the two methods give the same accuracy.

First of all using the normalization method we calculate the equilibrium distribution and we present the probabilities closest to the origin in Table 1. More specifically, we assume for our calculations that the series truncation level is $K = 2$, the triangular area is obtained for $B = \lceil 60 \cdot \rho \rceil$ and the total number of states is selected according to $b = 2B$. Looking at the results in Table 1 it is evident that for small values of ρ the truncated series give extremely accurate results even if we only take into consideration the $2^{K+1} - 1 = 7$ first terms. Of course as ρ increases we can either increase the series truncation level K or equivalently increase B. If we increase K we increase the number of terms in a geometric way, however by increasing B the complexity of solving the linear system does not change (as much) due to the sparsity of the corresponding linear system matrix. Hence, as ρ increases it is better to keep K at a fixed small value, such as $K = 2, \ldots, 8$, and increase B.

We present the probability that the server is located at the shortest queue, $\mathbb{P}[X \leq Y]$. Having accurately calculated the equilibrium distribution we can simply add all those equilibrium probabilities and obtain the result under question. We plot $\mathbb{P}[X \leq Y]$ as a function of ρ, for $\rho \in [0.01, 0.99]$. Figure 8 depicts the probability $\mathbb{P}[X \leq Y]$ using the normalization method. For $\rho \leq 0.9$ the relative accuracy of our results is more than 99.9999%, but as ρ approaches 1, the relative accuracy decreases and slightly drops below 99%.

In Figure 8 we present the probability $\mathbb{P}[X \leq Y]$ as a function of ρ ($\rho \in [0.01, 0.99]$). Here, we assume for our calculations that the series truncation level is $K = 5$, the triangular area is obtained for $B = \lceil 120 \cdot \rho \rceil$ and the total number of states is selected according to $b = 2B$. For the rest of the numerical results we consider that the parameters K, B and b remain unaltered.

We observe that as $\rho \to 1$ the probability the server is at the shorter queue converges to one. This numerical result contradicted our initial intuition. However, after some reflection and taking into account the results presented by Feller in [18] on the Ballot Theorem and the analysis presented below the results are sensible and fully verifiable.

Indeed, assume that at time 'zero' the server switches to queue one, meeting N customers in this queue (queue one)
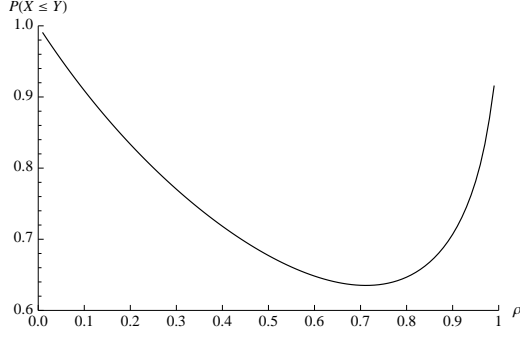
Figure 8: Probability that the server is at the shortest queue versus ρ.

(a) mean queue lengths vs $\rho \in [0.01, 0.9]$.

(b) mean queue lengths vs $\rho \in [0.9, 0.99]$.

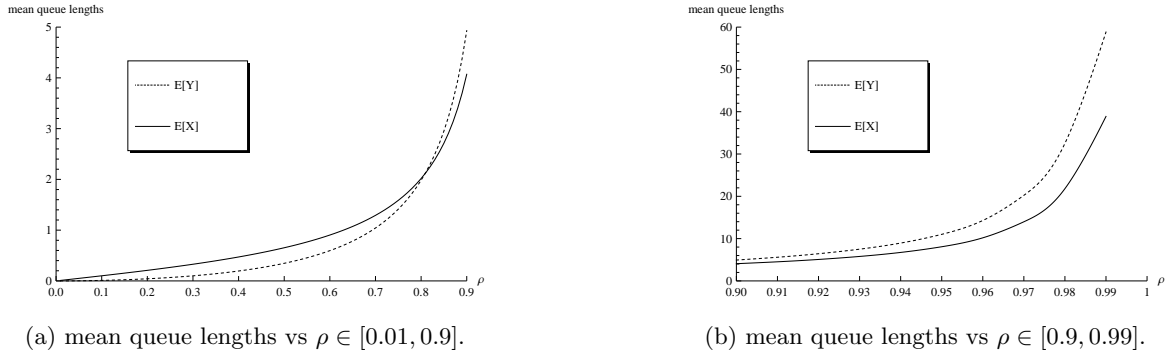Figure 9: Busy, $\mathbb{E}[X]$ (continuous line), and idle, $\mathbb{E}[Y]$ (dotted line), mean queue lengths versus ρ.

and leaving zero customers back in queue two. Hence, we start with $X(0) = N >> 1$, and $Y(0) = 0$. We are interested in the case that $\rho \to 1$. To this we define $\rho = 1 - \epsilon$ and we rescale the service rate to $\mu = 1$, then the arrival rate is $\lambda = 1 - \epsilon$. After roughly $N/2$ time units, both queue lengths become equal (mean queue length of both queues is roughly equal to $N/2$). Let us say that when the two queue lengths meet for the first time this constitutes the completion of the first phase and hence the second phase begins. The second phase represents the time it takes the queue under service to empty for the first time. We show that the length of the first phase $= o$(length of the second phase), and that in the second phase the queue length at the busy queue is almost all the time smaller than the queue length at the idle queue. For this reason, consider the difference $D = Y - X$ at jump epochs of the second phase. Observe that the differences, say D_n, at these discrete time points do a random walk in $\{-1, 0, 1, \dots\}$. The random walk takes unit steps as follows: If $D_n \geq 0$ then

$$D_{n+1} = \begin{cases} D_n + 1, & \text{with probability } \frac{\mu}{\lambda + \mu} = \frac{1}{2 - \epsilon}, \\ D_n - 1, & \text{with probability } \frac{\lambda}{\lambda + \mu} = \frac{1 - \epsilon}{2 - \epsilon}, \end{cases}$$

and if $D_n = -1$, then $D_{n+1} = D_n + 1 = 0$. So our random walk at time zero, $n = 0$, starts from point zero, since both queue lengths are assumed to be equal, $D_0 = 0$. We are interested in the time it takes the queue currently being served to empty, i.e. the duration of the second phase. Note that at time zero both queues had $N/2$ customers, so at the time the busy queue empties the idle queue has at least $N/2$ customers. Hence we are interested in the expected number of steps it takes the random walk to exceed for the first time $N/2$. Let $V_{i,K}$ denote the number of steps needed to reach K for the first time starting at $i \leq K$. Then a simple recursion reveals that

$$\mathbb{E}[V_{0, N/2}] = \frac{(N + 4)N}{4}.$$

27

This leads us to the conclusion that it took $N/2$ steps to complete the first phase and at least $(N+4)N/4$ to complete the second phase. Of course there are some (rare) steps that the random walk spends during the second phase at state -1 (which corresponds to the case that the server is located at the longer queue), but they are negligible in comparison to the steps the random walk takes on the non-negative integers. Moreover, the mean total number of customers in the system is equal to $1/(1-\rho) = \epsilon^{-1}$. Then, at the start of a busy period there are at least $N = O(\epsilon^{-1})$ customers in the busy queue. To summarize, the server after switching queues finds approximately ϵ^{-1} customers waiting to be served. It takes approximately $\epsilon^{-1}/2$ time for both queues to have the same number of customers, i.e., to complete
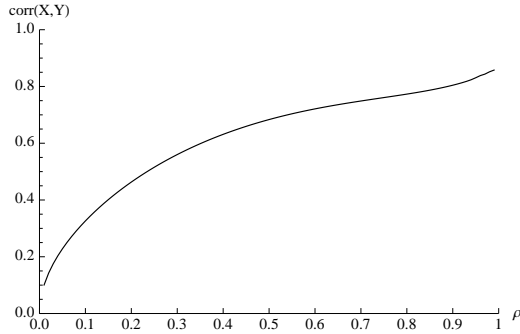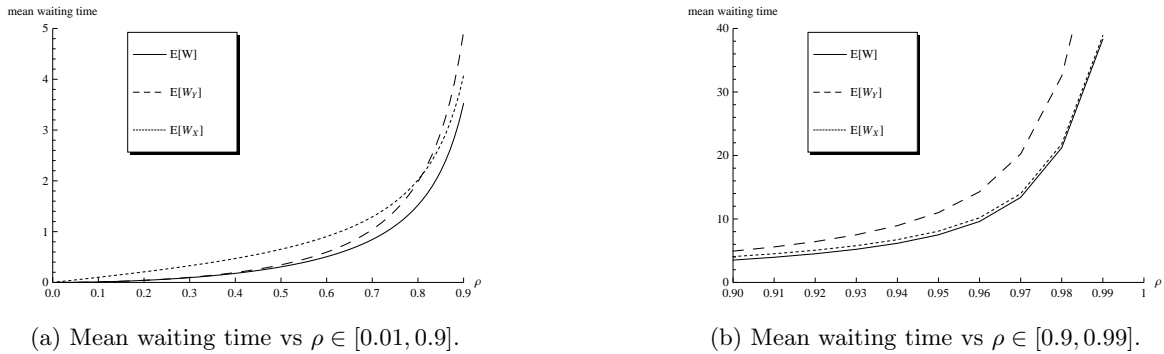

Figure 10: Correlation between the busy and the idle queue versus ρ.

(a) Mean waiting time vs $\rho \in [0.01, 0.9]$.

(b) Mean waiting time vs $\rho \in [0.9, 0.99]$.

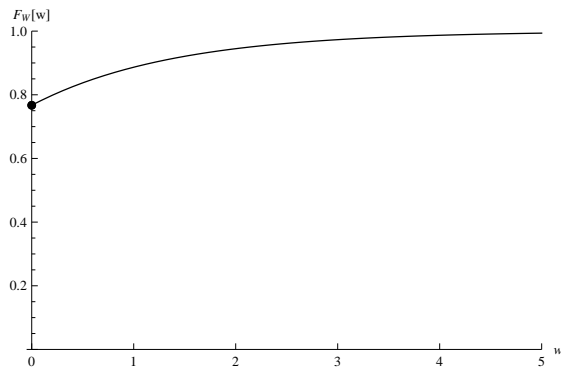Figure 11: Mean waiting time in the busy queue versus ρ, for $\mu = 1$.

Figure 12: Waiting time cdf for $\lambda = 0.5$ and $\mu = 1$.

the first phase. Then, it takes the server at least $O(\epsilon^{-2})$ to empty the busy queue, i.e., to complete the second phase.

Next, we present in Figure 9 the mean queue lengths for the busy and the idle queue, i.e. $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. The mean busy queue length as a function of ρ is sketched with a solid line, while the mean idle queue length is sketched with a dotted line. For these calculations both methods give identical results and we present here the scaling approach.

We continue with Figure 10 where we plot the correlation coefficient of the queue lengths for the busy and the idle queue, i.e. $\text{corr}(X,Y) = \frac{\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{Var[X]Var[Y]}}$. We observe that the correlation is always positive and increasing as ρ increases. It is interesting to remark that as $\rho \to 1$ the correlation reaches some critical value close to 0.7144. As ρ reaches one the correlation reaches its maximal strength, see also [12].

Moreover, we plot the mean waiting time of a customer taking into account the cost structure that motivated this paper. We assume that a customer has a zero cost while waiting in the idle queue and there is a fixed cost c that the customer pays per time unit spent waiting in the busy queue. Hence, the overall cost of a customer is calculated as $c \times \mathbb{E}[W]$, where the random variable W denotes the waiting time of a customer spent only in the busy queue. For simplicity we normalize the service rate, i.e. $\mu = 1$ and the waiting cost $c = 1$. We define a vector of costs (c_X, c_Y) that denote the cost per time unit spent in the busy and the idle queue, respectively. Adan et al. in [3] considered the case that $c_Y = 0$ and showed that the optimal dispatching policy is the JSQ discipline. In Figure 11 we plot the total mean waiting cost for a $(c_X = 1, c_Y = 0)$ cost. This function is plotted with a solid line. Moreover, we plot two additional functions:

i) $\mathbb{E}[W_X]$ corresponding to the total waiting costs for costs $(c_X = 1, c_Y = 1)$ given that the customer upon arrival found the busy queue less or equally congested as the idle queue, and

ii) $\mathbb{E}[W_Y]$ corresponding to the total waiting costs for costs $(c_X = 1, c_Y = 1)$ given that the customer upon arrival found the busy queue strictly more congested than the idle queue.

$\mathbb{E}[W_X]$ is plotted in Figure 11 with a dotted line while $\mathbb{E}[W_Y]$ is plotted with a dashed line. Note that the total waiting cost is as expected always less or equal to the minimum of the other two cost functions.

Finally, we numerically calculate the cumulative distribution function (cdf) of the waiting time of a customer spent only in the busy queue. In Figure 12 we present the cdf of the waiting time for $\lambda = 0.5$ and $\mu = 1$.

References

[1] ADAN, I.J.B.F. (1991). *A Compensation Approach for Queueing Problems.* PhD dissertation, Eindhoven University of Technology, Eindhoven.

[2] ADAN, I.J.B.F., BOXMA, O.J. and RESING J.A.C. (2001). Queueing models with multiple waiting lines. *Queueing Systems*, **37**(1-3) 65–98.

[3] ADAN, I.J.B.F., KULKARNI, V.G., LEE, N. and LEFEBER, A.A.J. (2013). Optimal routing of customers in polling systems. *Under preparation.*

[4] ADAN, I.J.B.F., and WESSELS, J. (1996). Shortest expected delay routing for Erlang servers, *Queueing Systems*, **23** 77–105.

[5] ADAN, I.J.B.F., WESSELS, J. and ZIJM, W.H.M. (1990). Analysis of the symmetric shorter queue problem. *Stochastic Models*, **6**(4) 691–713.

[6] ADAN, I.J.B.F., WESSELS, J. and ZIJM, W.H.M. (1991). Analysis of the asymmetric shorter queue problem. *Queueing Systems*, **9**(1) 1–58.

[7] ADAN, I.J.B.F., WESSELS, J. and ZIJM, W.H.M. (1993). A compensation approach for two-dimensional Markov processes. *Advances in Applied Probabibility*, **25**(4) 783–817.

[8] ADAN, I.J.B.F., KAPODISTRIA, S. and VAN LEEUWAARDEN, J.S.H. (2012). Erlang arrivals joining the shorter queue. *Queueing Systems*, DOI: 10.1007/s11134-012-9324-8.

[9] ALTMAN, E., JIMÉNEZ, T. and KOOLE, G. (2001). On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications*, **49**(9) 1659–1668.

[10] BLANC, J.P.C. (2009). Bad luck when joining the shortest queue. *European Journal of Operational Research*, **195**(1) 167–173.

[11] BOON, M.A.A., VAN DER MEI, R.D. and WINANDS, E.M.M. (2011). Applications of polling systems. *SORMS*, **16** 67–82.

[12] BOXMA, O.J. and VAN HOUTUM, G.J. (1993). The compensation approach applied to a 2×2 switch. *Probability in the Engineering and Informational Sciences*, **7**(4) 471–493.

[13] COHEN, J.W. (1998). Analysis of the asymmetrical shortest two-server queueing model. *Journal of Applied Mathematics and Stochastic Analysis*, **11**(2) 115–162.

[14] COHEN, J.W. and BOXMA, O.J. (1983). *Boundary Value Problems in Queueing System Analysis*, North-Holland, Amsterdam.

[15] FAYOLLE, G., FLAJOLET, P. and HOFRI, M. (1986) On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel. *Advances in Applied Probability*, **18** 441–472.

[16] FAYOLLE, G. and IASNOGORODSKI, R. (1979). Two coupled processors The reduction to a Riemann-Hilbert problem. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **47** 325–351.

[17] FAYOLLE, G., IASNOGORODSKI, R. and MALYSHEV, V. (1999). *Random Walks in the Quarter Plane*, Springer-Verlag, New York.

[18] FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York.

[19] FLATTO, L. and MCKEAN, H.P. (1977). Two queues in parallel. *Communications on Pure and Applied Mathematics*, **30**(2) 255–263.

[20] FOLEY, R.D. and MCDONALD, D.R. (2001). Join the shortest queue stability and exact asymptotics. *The Annals of Applied Probability*, **11**(3) 569–607.

[21] FOSTER, F.G. (1953). On the stochastic matrices associated with certain queueing processes. *Annals of Mathematical Statistics*, **24** 355–360.

[22] JACQUET, P. and MERLE, E. (1990). Analysis of a stack algorithm for CSMA-CD random length packet communication. *IEEE Transactions on Information Theory*, **36**(2) 420–426.

[23] GUPTA, V., HARCHOL BALTER, M., SIGMAN, K. and WHITT, W. (2007). Analysis of join-the-shortest-queue routing for web server farms. *Performance Evaluation*, **64**(9) 1062–1081.

[24] HALFIN, S. (1985). The shortest queue problem. *Journal of Applied Probability*, **22** 865–878.

[25] HORDIJK, A. and G. KOOLE (1990). On the optimality of the generalised shortest queue policy. *Probability in the Engineering and Informational Sciences*, **4** 477–487.

[26] HORDIJK, A. andSPIEKSMA, F. (1989). Constrained admission control to a queueing system. *Advances in Applied Probability*, **21**(2) 409–431.

[27] HOUCK, D.J. (1987). Comparison of policies for routing customers of parallel queueing systems, *Operations Research*, **35**(2) 306–310, 1987.

[28] KINGMAN, J.F.C. (1961). Two similar queues in parallel. *The Annals of Mathematical Statistics*, **32**(4) 1314–1323.

[29] LEVY, H. and SIDI, M. (1990). Polling models applications, modeling and optimization. *IEEE Transactions on Communications*, **38** 1750–1760.

[30] LI, H., MIYAZAWA, M. and ZHAO, Y. Q. (2007). Geometric decay in a QBD process with countable background states with applications to a join-the-shortest-queue model. *Stochastic Models*, **23**(3) 413–438.

[31] MENICH, R. (1987). Optimality of shortest queue routing for dependent service stations. *In 26th IEEE Conference on Decision and Control*, **26** 1069–1072.

[32] MENICH, R. and SERFOZO, R.F. (1991). Optimality of routing and servicing in dependent parallel processing systems. *Queueing Systems*, **9**(4) 403–418.

[33] RESING, J.A.C. (1993). Polling systems and multitype branching processes. *Queueing Systems*, **13**(4) 409–426, 1993.

[34] RESING, J.A.C. and RIETMAN, R. (2004). The M/M/1 queue with gated random order of service. *Statistica Neerlandica*, **58**(1) 97–110.

[35] STIDHAM, S. (1985). Optimal control of admission to a queueing system, *IEEE Transactions on Automatic Control*, **30**(8) 705–713, 1985.

[36] STIDHAM, S. (2009). *Optimal Design of Queueing Systems*. Chapman and Hall/CRC Press, Boca Raton, 2009.

[37] STIDHAM, S. and WEBER, R. (1993). A survey of Markov decision models for control of networks of queues. *Queueing Systems*, **13**(1) 291–314, 1993.

[38] TAKAGI, H. (1991). *Queueing Analysis*, volume 1. North-Holland, Amsterdam, 1991.

[39] TAKAGI, H. (1991). Application of polling models to computer networks. *Comput. Netw. ISDN Syst.*, **22** 193–211, 1991.

[40] TAKAGI, H. (2000). Analysis and application of polling models. In G. Haring, C. Lindemann, and M. Reiser, editors, *Performance Evaluation Origins and Directions*, volume 1769 of Lecture Notes in Computer Science, pages 424–442. Springer, Berlin.

[41] VISHNEVSKII, V.M. and SEMENOVA, O.V. (2006). Mathematical methods to study the polling systems. *Automation and Remote Control*, **67** 173–220.

[42] WHITT, W. (1986). Deciding which queue to join some counterexamples. *Operations research*, **34** 55–62.

[43] WINSTON, W. (1977). Optimality of the shortest line discipline. *Journal of Applied Probability*, **14** 181–189.

[44] YAO, H. and KNESSL, C. (2005). On the infinite server shortest queue problem Symmetric case. *Stochastic Models*, **21**(1) 101–132.

[45] YAO, H. and KNESSL, C. (2008). On the shortest queue version of the Erlang loss model. *Studies in Applied Mathematics*, **120**(2) 129–212.

[46] YECHIALI, U. (1972). Customers' optimal joining rules for the $GI/M/s$ queue. *Management Science*, **18**(7) 434–443.