The state-dependent $M/G/1$ queue with orbit

Opher Baron
Rotman School of Management, University of Toronto, Toronto, Canada
email: opher.baron@rotman.utoronto.ca

Antonis Economou
Department of Mathematics, National and Kapodistrian University of Athens, Athens, Greece
email: aeconom@math.uoa.gr

Athanasia Manou
Department of Industrial Engineering, Koç University, Istanbul, Turkey
email: amanou@ku.edu.tr

December 5, 2017

Abstract: We consider a state-dependent single server queue with orbit. This is a versatile model for the study of service systems, where the server needs a non-negligible time to retrieve waiting customers every time he completes a service. This situation arises typically when the customers are not physically present at a system, but they have a remote access to it, as in a call center station, a communication node etc. We introduce a probabilistic approach for the performance evaluation of this queueing system, that we refer to as the Queueing and Markov Chain Decomposition approach. Moreover, we discuss the applicability of this approach for the performance evaluation of other non-Markovian service systems with state dependencies.

Keywords: state-dependent queueing system; orbit; retrial queue; non-negligible retrieval time; steady-state distribution; conditional sojourn time distributions; variable arrival rate; variable service speed; Queueing and Markov Chain Decomposition

1 Introduction

In many service systems, customers submit their petitions for service remotely. Then, they can begin their service immediately if they find an idle server, otherwise their petitions are queued and customers are assumed to wait at a virtual queue that is referred to as the *orbit*. Upon finishing a service, the server turns his attention to the orbit and looks for a customer to begin a new service. However, this seeking time is usually non-negligible, since the retrieval of a customer from the orbit may require some preliminary steps (e.g., to find an associated file, to call back the customer etc.). Then, if a new customer arrives during the seeking time, the retrieval process is abandoned and the server begins to serve the newly arrived customer. This situation occurs in a variety of service contexts: For example, consider customers that do not find an idle server at a call center, who may leave their contact details and wait to be called back later. Similarly, consider a person that answers emails for a commercial website. Such a person begins immediately to answer a newly arrived email if he is idle, otherwise he needs some time to retrieve a waiting email from his list.

Moreover, a node in certain communication networks begins immediately the transmission of a new message if it is idle, otherwise it needs some additional time to retrieve stored messages to be processed.

Our objective is to study a versatile model for the representation of such service systems, where the retrieval of waiting customers requires a non-negligible time. Indeed, the main motivation of the proposed model is the consideration of a framework that captures this feature. However, we notice that real service systems with this feature have other additional characteristics. Under this point of view, our model constitutes a first attempt for the study of such systems and more involved models should be developed. More concretely, we introduce a general state-dependent queue with a general retrial/retrieval policy. The consideration of the state-dependent version of this model seems particularly important mainly for four reasons. First, in many systems the arrival rate depends on the number of jobs in the orbit. An important example arises in the study of the strategic behavior of customers in such systems. Indeed, when the arriving customers assume a certain joining/balking strategy, then the arrival rates become state-dependent. Therefore, to evaluate a tagged customer's best response against a strategy of the others, one needs to use the performance measures of a state-dependent version of the model. Second, the consideration of state-dependencies allows the study of dynamic control policies in a system with orbit. Third, in many applications the length of a retrieval period depends on the number of customers in orbit. For example, in the classical retrial policy where each job from the orbit retries independently, the retrieval time becomes smaller as the number of customers in orbit increases. And fourth, the service distribution and speed may also depend on the number of customers in orbit (see e.g., Abouee-Mehrizi and Baron (2016) for a similar situation in the framework of the classical $M/G/1$ queue). These economic and control issues motivate us to consider the state-dependent version of the model. We focus on its performance evaluation and derive efficient recursive schemes for various performance descriptors.

As a specific application of our model consider load sharing in distributed systems, see e.g., Eager, Lazowska and Zahorjan (1986) and references within. The idea behind load sharing is to smooth the load on the processors in different stations of a network. There are many different algorithms for the application of load sharing. Each algorithm takes two decisions: when to transfer a task from a station and how to allocate transferred tasks among stations. It has been demonstrated in Lazowska, Zahorjan, Cheriton and Zwaencpocl (1984) that the processor time for a task transfer (packaging, transmitting, and unpackaging) is more instrumental than the network cost. Thus, the latter cost is often ignored in the analysis of algorithms for load sharing.

A simple and effective transfer policy is a threshold policy, where new tasks are transferred if and only if the number of tasks in queue is bigger than a threshold, T, when the station is at its transferring phase, and are accepted otherwise, when the station is in its processing phase. This transfer policy can be combined with various allocation policies, such as random allocation or allocation to the shortest queue.

Two important observations on the analysis of load sharing are as follows. First, the performance measures of the system can be obtained from the analysis of any individual station. As Eager, Lazowska and Zahorjan (1986) state, this analysis is accurate if all stations are identical, i.e., the network is homogeneous, asymptotically exact when the number of stations grows, and has been shown to be quite accurate in simulation results with 20 or more stations. Second, load balancing algorithms result in a state-dependent arrival rate to stations. For any policy, this rate depends on at least whether the station is in its processing or transferring phases. This rate is clearly state-dependent for allocation policies that use information on the state space, such as allocation to the shortest queue.

In the reasonable case where each station prioritizes its own work, new arrivals during a seeking time interrupt the seeking time and immediately get into service. Then, with the description above, our state-dependent M/G/1 queue with orbit is an appropriate model for analyzing the performance of load sharing algorithms, by setting the threshold T, modeling the processor task transfer time as seeking time, capturing the state-dependent arrival rate, and allowing for generally distributed processing times that differ between new and allocated arrivals.

Additional applications for our approach can be found in Phung-Duc, Rogiest and Wittevrongel (2017) for state-dependent retrial queues and Yajima and Phung-Duc (2017) for state-dependent service rate. See also Phung-Duc (2014, 2017a, 2017b) for further discussion of motivating applications and relevant extensions.

The main contributions of the paper can be summarized as follows:

- On the modeling side, we consider a general model of a service system where the retrieval of waiting customers requires, possibly, a non-negligible time. This system allows generally distributed service and seeking (retrieval) times and extensive state-dependencies regarding the arrival rates and the service and seeking time distributions.
- On the technical side, we develop an intuitive probabilistic methodology for the performance evaluation of non-Markovian queueing systems with state-dependencies, that we refer to as the Queueing and Markov Chain Decomposition (QMCD) approach. This approach is effectively applied for the model of interest and easily yields its performance descriptors. More importantly, we discuss how this methodology can be adapted for the study of other non-Markovian queueing systems with state-dependencies.

The paper is structured as follows: In Section 2 we present the literature review. In Section 3, we present the model of the state-dependent single server queue with orbit and the performance quantities of interest. In Section 4, we present an overview of the QMCD approach in general terms and then we briefly specify its application for the study of the model of interest. In Section 5, we apply this methodology to obtain efficient computational schemes for various descriptors of the model. Finally, in Section 6, we conclude our paper with some conclusions and future research directions. Proofs not in the text appear in the Appendix that also contains some technical analytic proofs.

2 Literature review

The analysis of queues with orbit, known also as retrial queues, constitutes an important subfield of Queueing Theory, since the consideration of the retrial feature enables the more accurate representation and quantification of real service systems with finite waiting spaces. Indeed, in such systems, it is assumed that arriving customers who do not find a free waiting position join a so-called retrial orbit and return later for service. The literature on retrial queues is very rich, see e.g., the books of Falin and Templeton (1997) and Artalejo and Gomez-Corral (2008) and references therein.

Most studies on retrial queueing systems assume the so-called classical retrial policy for the customers in orbit, according to which each retrying customer conducts attempts for service independently of the other customers, usually with exponentially distributed inter-retrial times. However, there are certain service situations in which the intervals between the successive retrials from the orbit are independent of the number of customers in it. Such cases occur when only the customer that is at the 'head' of the orbit is allowed to conduct retrials or, more realistically, when it is the server that looks for customers from the orbit after every service completion (see e.g., the situations of a call center with a call-back option, the email list and the communication node in the discussion above). This retrial policy is known as the constant retrial policy.

The constant retrial policy was introduced by Fayolle (1986) in a Markovian framework where both service and retrial (seeking) times are exponential. Later on, Choi, Park and Pearce (1993) considered the corresponding model where the seeking times are generally distributed, while Martin and Artalejo (1995) considered the model where the service times are generally distributed. Finally, Gomez-Corral (1999) considered the single server retrial queue with the constant retrial policy and general service and seeking times. This paper presented an exhaustive analysis of this general model which includes the stability condition, the probability generating function of the steady-state distribution and the Laplace-Stieltjes transforms of the waiting time, busy period and idle period distributions. Many papers have appeared since the seminal work of Gomez-Corral (1999) that extend his analysis in systems with additional characteristics. However, there is no generalization for the analysis of the state-dependent version of this model. One of our objectives is to fill this gap and to provide effective computational procedures for this case.

The performance evaluation of queueing models with state-dependencies has received recently considerable attention. Kerner (2008) analyzed the $M/G/1$ queue with state-dependent arrival rates, using an analytic approach, based on the supplementary variable method (see Cox (1955)). The same problem was analyzed by Economou and Manou (2015) using a probabilistic approach which we extend and clarify in the present paper. Abouee-Mehrizi and Baron (2016) studied a generalization where the service time distributions depend on the number of customers in the queue at a service beginning, in addition to state-dependent arrival rates. Oz, Adan and Haviv (2015, 2017) and Oz (2016) considered models with state-dependencies under a different viewpoint, using a rate balance principle. Finally, Manou, Economou and Karaesmen (2014) considered the analysis of a clearing system with state-dependent arrival rates as a necessary step towards the study of customer strategic behavior in a transportation station. In general, the study of customer strategic behavior in observable queueing systems requires the consideration of models with state-dependencies.

3 The model

We consider the following queueing system, which we refer to as the $M_n/G_n/1/1$ queue with orbit: There is one server that serves customers singly, no waiting space, and customers that arrive when the server is busy join a retrial orbit. When the server becomes free, after a service completion, he begins to retrieve a customer from the retrial orbit according to the FCFS discipline (we assume FCFS as it is standard in the literature, but this is not essential for our analysis). If the seeking (retrieval) time is completed before the next arrival, the server begins to serve the oldest customer from the orbit. Otherwise, i.e., when a new customer arrives during the seeking time, the retrieval process is interrupted and the new customer enters service. The retrieval process begins anew the next time the server is free. We consider a general system with state dependencies on the number of customers in service and in orbit, i.e., we allow inter-arrival, seeking, and service times to depend on the number of customers in service and in orbit. Moreover, service times may also depend on whether the served customer has come from the orbit or is a new arrival. This is important, e.g., when customers from the orbit require a shorter identification process by the server.

Regarding the state-dependencies, we assume that the time till the next arrival is exponentially distributed with rate λ_{ij}, whenever there are i customers in service ($i = 0$ or 1) and j customers in orbit ($j = 0, 1, 2, \ldots$). We denote a corresponding random variable by $T_{\lambda_{ij}}$, in accordance with the notation T_s for denoting an exponentially distributed random variable with rate s, that we will use in the paper. In other words, customers arrive to the system according to a state-dependent Poisson process. A typical service time of a customer from the orbit that observes j customers in the orbit upon the start of service (i.e., there were $j+1$ customers in the system before his service beginning) is denoted by B_{0j}, whereas B_{1j} represents a service time of a new customer that observes an idle or seeking server upon arrival and j customers in orbit. Finally, we let A_j denote a generic seeking time that begins when there are j customers in orbit. Moreover, we assume that the various service, seeking and inter-arrival times are independent. For any random variable Z that appears in the paper, we denote its expectation by $E[Z]$, its cumulative distribution function by $Z(x)$, its Laplace-Stieltjes transform (LST) by $\tilde{Z}(s) := \int_0^\infty e^{-sx} dZ(x)$, and its probability density function (in case that Z has a continuous distribution) by $z(x)$.

To obtain a Markovian description of the system, we define for each time t, the random variables $C(t)$, $Q(t)$, and $S(t)$ that record, respectively, the state of the server (0 or 1), the number of customers in the retrial orbit, and the remaining time till the next service/seeking completion. More specifically, $C(t) = 1$

when the server is busy (serving) at time t, whereas $C(t) = 0$ when the server is idle or seeking at time t, and $S(t)$ records the remaining service time at time t when $C(t) = 1$ or the remaining seeking time at time t when $C(t) = 0$. For times t such that $C(t) = Q(t) = 0$, the random variable $S(t)$ is not defined.

A moment of reflection shows that the stochastic process $\{(C(t), Q(t), S(t)) : t \geq 0\}$ is a continuous time Markov process with state space $\{(0,0)\} \cup \{(0, j, x) : j = 1, 2, \ldots; x \geq 0\} \cup \{(1, j, x) : j = 0, 1, \ldots; x \geq 0\}$. We are interested in the limiting behavior of this process as $t \to \infty$, i.e., in the steady-state probabilities $p_{ij} = \lim_{t\to\infty} \Pr[C(t) = i, Q(t) = j]$, for $(i, j) \in \{0, 1\} \times \{0, 1, 2, \ldots\}$. We will derive these probabilities using a method we refer to as the Queueing and Markov Chain Decomposition (QMCD). The QMCD approach is intuitive and uses simple probabilistic arguments. We assume throughout the paper that the system is stable. The stability condition is given in Theorem 5.3. In the special case where all quantities (arrival rates, seeking and service time distributions) become state-independent, the stability condition is given in Theorem 5.4 and coincides with the one in Gomez-Corral (1999) (Theorem 1).

4 Overview of the QMCD approach

The QMCD approach that will be described in detail below in the framework of the present model is a versatile technique for the performance evaluation of a variety of other non-Markovian queueing systems with state-dependencies.

The fundamental idea of the method is that a given queueing system can be seen as a network of its states, where a single customer (the so-called pointer) circulates in it indicating which is the actual state of the system at each time. Then, a number of useful equations can be derived, by relating the remaining generally distributed times (e.g., service times, inter-arrival times etc.) as the system changes states, by equating arrival and departure rates at each subsystem and by applying Little's law.

To be concrete, consider a queueing system represented by a continuous-time Markov chain with transition rates q_{ij} and stationary probabilities p_j. Then, we can associate a network of subsystems to it, in which a pointer circulates from state to state of the original system. By equating the arrival and the departure rates at each subsystem j we deduce that

$$\sum_{i\neq j} p_i q_{ij} = \sum_{i\neq j} p_j q_{ji}$$

which is the balance equation for state j. However, the same equation can be derived by applying Little's law at subsystem j. Indeed, for each subsystem j, we have

$$E[Q_j] = \Lambda_j E[W_j],$$

where $E[Q_j]$, Λ_j and $E[W_j]$ are, respectively, the mean number of customers, the arrival rate, and the mean sojourn time at subsystem j. Since there is only one customer (the pointer) that circulates in the network, we have that Q_j is 1 if the pointer is at subsystem j (with probability p_j) and 0 if it is not, i.e., $E[Q_j] = p_j$. Moreover, the arrival rate at subsystem j is $\Lambda_j = \sum_{i\neq j} p_i q_{ij}$, whereas the sojourn time at subsystem j is exponential with rate $\sum_{i\neq j} q_{ji}$, from which we have $E[W_j] = \left(\sum_{i\neq j} q_{ji}\right)^{-1}$. Hence, Little's law yields

$$p_j = \sum_{i\neq j} p_i q_{ij} \left(\sum_{i\neq j} q_{ji}\right)^{-1},$$

which is also the balance equation of state j.

In the case of a non-Markovian queueing system, the same idea applies, i.e., we can see the system as the network of its states with the pointer circulating in it. Then, we obtain two sets of equations that involve the steady-state probabilities, by equating arrival and departure rates at each subsystem and by applying Little's law at each subsystem. However, these two sets of equations are not identical in the non-Markovian case. Moreover, they involve two sets of new quantities: transition rates due to generally distributed time completions and the remaining generally distributed times when a subsystem is entered. We then need an additional set of equations to have sufficient number of equations to derive all three types of quantities. This last set of equations is derived by considering how the remaining generally distributed times evolve as the system changes states.

Thus, in a non-Markovian queueing system, even with state-dependencies, where only one generally distributed time runs each instant, we can apply the QMCD approach to derive sufficient number of equations for the steady-state probabilities, the transition rates due to generally distributed time completions and the LSTs of the remaining generally distributed times when a state is entered.

The application of QMCD in this paper and the discussion above focus on the case where each subsystem is of a single state. Nevertheless, QMCD is also applicable when the decomposition is to subsystems which correspond to sets of states. For example, see its application in Wang, Baron and Scheller-Wolf (2015) for a multi-server queue, and in Abouee-Mehrizi, Baron and Berman (2014) for a queueing network. Overall the QMCD approach includes the following four phases:

QMCD - Phase A: Decompose the system to subsystems. This phase reduces the complexity of the analysis from a combined system to smaller, simpler subsystems.

QMCD - Phase B: Tie the subsystems together – transitions among subsystems and other effects. It is instrumental to consider the effects of subsystems on each other *before* solving them. Such effects may modify the subsystems and therefore should be carefully captured in their analysis. An important premise of QMCD is that *only the average* effects among the subsystems should be captured. That is, QMCD analyzes subsystems with the same steady state distribution as the corresponding part of the system, acknowledging that the subsystems differ than the system in their sample paths.

QMCD - Phase C: Solve each subsystem. Solving the modified subsystems can often be done using known queueing approaches. An important part of this solution is to characterize the average time spent in each subsystem.

QMCD - Phase D: Use the normalization equation. The normalization of the solution is the last step going from the steady-states of the modified subsystems to the steady-state of the entire system. This step is tied to the stability condition of the entire system.

In the case where each subsystem is of a single state the phase B of the QMCD is divided to two steps. Then, the QMCD requires the completion of 5 steps (1 step for each of the phases A, C and D, and 2 steps for phase B).

To be concrete, we now describe these steps in the framework of our model. First, note that there are three fundamental quantities that are involved in the QMCD approach:

- the steady-state probabilities p_{ij},
- the steady-state rates μ_{ij} of service/seeking time completions per time unit that initiated from state $(C(t), Q(t)) = (i, j)$ (note that these are unconditional rates, i.e., rates of certain transitions per time unit, whereas λ_{ij} are conditional rates, i.e., rates of transitions per unit time that the pointer is in subsystem (i, j)),
- the random variables S_{ij} that represent the limiting conditional distribution of $S(t)$ given that $(C(t), Q(t)) = (i, j)$, for $(i, j) \neq (0, 0)$. That is, S_{1j} and S_{0j} represent, respectively, the conditional remaining service and seeking times given that there are j customers in orbit. We let $\tilde{S_{ij}}(s)$ denote their LSTs.

Step 1 of the QMCD approach consists in decomposing the original system into infinitely many subsystems, one for each single state of $(C(t), Q(t))$, i.e., instead of the original system, we consider a queueing network (where the nodes-subsystems correspond to the various states) and there is a single customer - the so-called pointer - that circulates in it. When the pointer is at a given subsystem, the original system is found in the corresponding state. Having this consideration in mind, the above quantities of interest acquire new meanings for each state (i, j) of $(C(t), Q(t))$:

- The random variable S_{ij} is the remaining customer service/seeking time, when the pointer is at subsystem (i, j).
- The probability p_{ij} is the long-run fraction of time that the pointer is at subsystem (i, j).
- The rate μ_{ij} is the rate of customer service/seeking time completions that force the pointer to leave subsystem (i, j), per time unit.
- The rate $\lambda_{ij} p_{ij}$ is the rate of customer arrivals, when the pointer is at subsystem (i, j).

In step 2 of the QMCD approach, we relate the LSTs of the random variables S_{ij}. In step 3, we equate the in and out rates at each subsystem (i, j) and obtain equations that involve the probabilities p_{ij} and the rates μ_{ij}. In step 4, we apply Little's law at each subsystem. Thus, we obtain the necessary number of equations to compute quantities of interest up to a multiplicative constant. This constant is determined by the normalization equation that requires the sum of the probabilities of the system being in the various states to be equal to 1. This constitutes the final step of the QMCD approach, which also gives the stability condition of the system.

We note that QMCD can be useful in studying queueing networks as well. For example, the consideration of each of the nodes/subnetworks in a tandem queueing network as a subsystem and a similar treatment using a pointer can facilitate its analysis. Indeed, one can think of the solution approach for the multi-echelon inventory system in Abouee-Mehrizi, Baron and Berman (2014), and the solution of unidirectional quasi-birth-death processees in Doroudi, Fralix and Harchol-Balter (2015) as applications of QMCD to queueing networks. Of course, other ideas of queueing decomposition for the approximate analysis of non product form networks have been developed extensively in the literature. Reiser and Kobayashi (1974) presented a key method that has subsequently been used by many authors. Its application to manufacturing was pursued by Bitran and Tirupati (1988) and Caldentey (2001).

Similarly, for queueing systems with priorities, the overall service time of customers with other priorities can be seen as the single generally distributed time at each instant and thus analysis of such systems using QMCD seems also beneficial. Indeed, one can think of the solution approach for the $M/M/c$ queues with priorities in Wang, Baron and Scheller Wolf (2015) as an application of QMCD to queues with priorities.

5 Applying the QMCD approach

5.1 QMCD - Step 1: Decompose the system in subsystems

In this step, as explained above, we decompose the original system into subsystems. This is done by associating with each state of the original system a node (subsystem) of a closed queueing network with one customer. This customer is referred to as the pointer and circulates from subsystem to subsystem. His position at any given time shows the state of the original system.

5.2 QMCD - Step 2: Relate the remaining generally distributed times as the system evolves

In this step we derive recursive equations for the LSTs and the mean values of the random variables S_{ij}. First, we note that the rate with which service starts when there are j customers in orbit, i.e., when the pointer passes into subsystem $(1, j)$, is

$$\nu_{1j} := \mu_{0,j+1} + \lambda_{0j} p_{0j}. \tag{5.1}$$

The first term is the seeking completion rate that leads the original system from state $(0, j+1)$ to state $(1, j)$. After such a transition, a service of a customer from the orbit starts. This customer observes j customers in the orbit upon his service beginning. Thus, his service time is distributed as B_{0j}. The second term is the customers' arrival rate that leads the original system from $(0, j)$ to $(1, j)$. After this transition a service of a new customer starts. This customer observed an idle (if $j = 0$) or a seeking (if $j > 0$) server and j customers in the orbit upon arrival. Thus, his service time is distributed as B_{1j}. Then, $\mu_{0,j+1}/\nu_{1j}$ is the probability that a service time that started with j customers in the orbit of the original system is distributed as B_{0j} and $\lambda_{0j} p_{0j}/\nu_{1j}$ is the probability that this service time is distributed as B_{1j}. Therefore, a service time that started with j customers in the orbit has the representation

$$B_j \stackrel{d}{=} \begin{cases} B_{0j} & \text{with probability } \frac{\mu_{0,j+1}}{\nu_{1j}} \\ B_{1j} & \text{with probability } \frac{\lambda_{0j} p_{0j}}{\nu_{1j}}, \end{cases} \tag{5.2}$$

i.e., B_j is a mixture of B_{0j} and B_{1j} and its LST, $\tilde{B}_j(s)$, is given by

$$\tilde{B}_j(s) = \frac{1}{\nu_{1j}} \left(\mu_{0,j+1} \tilde{B}_{0j}(s) + \lambda_{0j} p_{0j} \tilde{B}_{1j}(s) \right). \tag{5.3}$$

Now, recall that S_{ij} stands for a random variable with the limiting conditional distribution of $S(t)$ given that the pointer is at subsystem (i, j). Since customers arrive according to a Poisson process with rate λ_{ij} as long as the pointer is at the subsystem (i, j), we can use the conditional Poisson Arrivals See Time Averages (PASTA) property (see van Doorn and Regterschot (1988), Theorem 1). We conclude that S_{ij} is identically distributed with the remaining service/seeking time observed by arriving customers that find the original system at state (i, j).

Let a_j be the conditional probability that an arriving customer is the first during the ongoing service time, given that the state after his arrival is $(1, j)$. Using a_j, we are able to state the following distributional relationships for the random variables S_{ij}.

So far, we have introduced all the parameters involved in our analysis. Table 1 summarizes the system parameters and their definitions.

Proposition 5.1 *(Relating the conditional seeking and service times) Given a_j we have*

$$S_{0j} \stackrel{d}{=} (A_j - T_{\lambda_{0j}} | A_j \geq T_{\lambda_{0j}}), \ j \geq 1, \tag{5.4}$$

$$S_{10} \stackrel{d}{=} (B_0 - T_{\lambda_{10}} | B_0 \geq T_{\lambda_{10}}), \tag{5.5}$$

$$S_{1j} \stackrel{d}{=} \begin{cases} (B_j - T_{\lambda_{1j}} | B_j \geq T_{\lambda_{1j}}) & \textit{with probability } a_j \\ (S_{1,j-1} - T_{\lambda_{1j}} | S_{1,j-1} \geq T_{\lambda_{1j}}) & \textit{with probability } 1 - a_j, \end{cases} \quad j \geq 1. \tag{5.6}$$

Proof. Consider a tagged customer that arrives at the original system and his arrival corresponds to a transition from state $(0, j)$ to $(1, j)$, for some $j \geq 1$. Then, this customer interrupted the ongoing seeking time and no other arrivals had occurred during this seeking time. Therefore, the arrival of the tagged

λ_{ij}	arrival rate, when there are i customers in service and j customers in orbit
p_{ij}	steady-state probability that there are i customers in service and j customers in orbit
T_s	exponentially distributed random variable with rate s
B_{0j}	service time of a customer from the orbit that observes j customers in the orbit upon the start of service
B_{1j}	service time of a new customer that observes an idle or seeking server upon arrival and j customers in orbit
A_j	seeking time that begins when there are j customers in orbit
S_{0j}	remaining seeking time, when there are j customers in the orbit
S_{1j}	remaining service time, when there are j customers in the orbit
μ_{0j}	rate of seeking time completions, when there are j customers in the orbit
μ_{1j}	rate of service time completions, when there are j customers in the orbit
a_j	the conditional probability that an arriving customer is the first during the ongoing service time, given that the state after his arrival is $(1, j)$.

Table 1: Main quantities of the model

customer is the first event of a Poisson process with rate λ_{0j} during a seeking time A_j and the remaining seeking time at the arrival instant is distributed as $A_j - T_{\lambda_{0j}}$. Of course, the existence of such a customer requires that $A_j \geq T_{\lambda_{0j}}$ and we obtain (5.4).

The proof of (5.5) that refers to a tagged customer that arrives at the system and his arrival corresponds to a transition from $(1, 0)$ to $(1, 1)$ is similar after replacing A_j by B_0.

Finally, consider a tagged customer that arrives at the original system and his arrival corresponds to a transition from $(1, j)$ to $(1, j+1)$, for some $j \geq 1$. There are two cases for such a transition. Either it is the first arrival during the ongoing service time, leading to the first branch of (5.6) (similarly to the justification of (5.5)), or it is not the first during the ongoing service time. In that case, the arrival time of the tagged customer is the first event of a Poisson process with rate λ_{1j} during the remaining service time $S_{1,j-1}$ of the previous customer who arrived at state $(1, j-1)$. This argument yields the second branch of (5.6). ∎

Now, we can translate (5.4)-(5.6) in terms of LSTs, using the following Lemma 5.1 that can be easily proved by direct calculations (for a proof see Lemma 3.3 in Economou and Manou (2015)).

Lemma 5.1 *(LST of a residual time) Let X be a non-negative generally distributed random variable with LST $\tilde{F}(s)$ and T_λ an exponential random variable with rate λ, independent of X. Let $\tilde{H}(s)$ denote the conditional LST of $X - T_\lambda$ given that $X \geq T_\lambda$, i.e., $\tilde{H}(s) = E[e^{-s(X-T_\lambda)} | X \geq T_\lambda]$. Then, for $s \neq \lambda$,*

$$\tilde{H}(s) = \frac{\lambda}{s - \lambda} \cdot \frac{\tilde{F}(\lambda) - \tilde{F}(s)}{1 - \tilde{F}(\lambda)}. \tag{5.7}$$

Therefore, Proposition 5.1, in light of (5.7), yields Corollary 5.1.

Corollary 5.1 *The LSTs $\tilde{S}_{ij}(s)$ satisfy the 1st QMCD system of equations:*

$$\tilde{S}_{0j}(s) = \frac{\lambda_{0j}}{s-\lambda_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)}{1-\tilde{A}_j(\lambda_{0j})}, \ j \geq 1, \tag{5.8}$$

$$\tilde{S}_{10}(s) = \frac{\lambda_{10}}{s-\lambda_{10}} \cdot \frac{\tilde{B}_0(\lambda_{10}) - \tilde{B}_0(s)}{1-\tilde{B}_0(\lambda_{10})}, \tag{5.9}$$

$$\tilde{S}_{1j}(s) = \frac{\lambda_{1j}}{s-\lambda_{1j}} \left(a_j \frac{\tilde{B}_j(\lambda_{1j}) - \tilde{B}_j(s)}{1-\tilde{B}_j(\lambda_{1j})} + (1-a_j) \frac{\tilde{S}_{1,j-1}(\lambda_{1j}) - \tilde{S}_{1,j-1}(s)}{1-\tilde{S}_{1,j-1}(\lambda_{1j})} \right), \ j \geq 1. \tag{5.10}$$

We can calculate the expectations of S_{ij} from their LSTs. We have $E[S_{ij}] = -\tilde{S}'_{ij}(0)$. Therefore,

$$\begin{aligned} E\left[S_{0j}\right] &= \left(\frac{\lambda_{0j}}{(s-\lambda_{0j})^2} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)}{1-\tilde{A}_j(\lambda_{0j})} + \frac{\lambda_{0j}}{s-\lambda_{0j}} \cdot \frac{\frac{d\tilde{A}_j(s)}{ds}}{1-\tilde{A}_j(\lambda_{0j})} \right) \Bigg|_{s=0} \\ &= \left(\frac{1}{\lambda_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - 1}{1-\tilde{A}_j(\lambda_{0j})} + \frac{E\left[A_j\right]}{1-\tilde{A}_j(\lambda_{0j})} \right) \\ &= \frac{E\left[A_j\right]}{1-\tilde{A}_j(\lambda_{0j})} - \frac{1}{\lambda_{0j}}, \ j \geq 1 \end{aligned} \tag{5.11}$$

and similarly

$$E\left[S_{10}\right] = \frac{E\left[B_0\right]}{1-\tilde{B}_0(\lambda_{10})} - \frac{1}{\lambda_{10}}, \tag{5.12}$$

$$E\left[S_{1j}\right] = a_j \frac{E\left[B_j\right]}{1-\tilde{B}_j(\lambda_{1j})} + (1-a_j) \frac{E\left[S_{1,j-1}\right]}{1-\tilde{S}_{1,j-1}(\lambda_{1j})} - \frac{1}{\lambda_{1j}}, \ j \geq 1. \tag{5.13}$$

Now, in Proposition 5.2, we proceed to the computation of a_j.

Proposition 5.2 *(Probability of an arrival to be the first during a service time) The conditional probability a_j that an arriving customer is the first during an ongoing service time, given that he finds the state $(C(t), Q(t)) = (1, j)$ upon arrival, is*

$$a_j = \frac{\nu_{1j}(1-\tilde{B}_j(\lambda_{1j}))}{\lambda_{1j} p_{1j}}, \tag{5.14}$$

where ν_{1j} is defined in (5.1).

Proof. Because of the regenerative nature of the original system the probability a_j can be interpreted as the long-run rate of customers that see state $(C(t), Q(t)) = (1, j)$ upon arrival and who are the first in the ongoing service time (excluding the customer that initiated the service), denoted by b_j, over the long-run rate of customers that see state $(C(t), Q(t)) = (1, j)$ upon arrival, denoted by c_j:

$$a_j = \frac{b_j}{c_j}. \tag{5.15}$$

Using the interpretation of the quantities p_{ij}, $\lambda_{ij} p_{ij}$, μ_{0j} and μ_{1j} in Section 4, we have that

$$c_j = \lambda_{1j} p_{1j}. \tag{5.16}$$

To compute b_j, note that each tagged arriving customer that sees state $(C(t), Q(t)) = (1, j)$ upon arrival and who is the first in the ongoing service time is in one-to-one correspondence with a service time that before its start either (i) the state was $(C(t), Q(t)) = (0, j+1)$ and a seeking process was terminated or (ii) the state was $(C(t), Q(t)) = (0, j)$ and a new arrival occurred. In other words, each such arriving customer is in one-to-one correspondence with a service beginning with j customers in orbit. In addition, this service time should exceed the arrival time of the tagged customer. Hence, the arrival rate of customers that see state $(C(t), Q(t)) = (1, j)$ upon arrival and who are the first in the ongoing service time is equal to the product of the rate of service beginnings at state $(C(t), Q(t)) = (1, j)$ times the conditional probability that at least one customer arrived during a service time given that the service time started from state $(C(t), Q(t)) = (1, j)$.

Recalling that ν_{1j} denotes the steady-state rate of service beginnings at state $(C(t), Q(t)) = (1, j)$ and that the conditional probability that at least one customer arrived during a service time given that the service time started from state $(C(t), Q(t)) = (1, j)$ is simply $\Pr[T_{\lambda_{1j}} < B_j] = 1 - \tilde{B}_j(\lambda_{1j})$, we obtain that

$$b_j = \nu_{1j}(1 - \tilde{B}_j(\lambda_{1j})). \tag{5.17}$$

Combining (5.15), (5.16) and (5.17) yields (5.14). ∎

Substituting a_j from Proposition 5.2 into (5.8)-(5.10) and (5.11)-(5.13) yields recursive schemes for the LSTs and the mean values of the random variables S_{ij}.

5.3 QMCD - Step 3: Equate in and out rates for each subsystem

In this step we relate the probabilities p_{ij} with the rates μ_{ij}, by equating the 'in' and the 'out' rates at each subsystem (i, j), i.e., the rates in which the pointer enters and leaves a given subsystem.

Proposition 5.3 *(Equating 'in' and 'out' rates at the subsystems) We have the formulas:*

$$\begin{aligned} \mu_{10} &= \lambda_{00} p_{00}, & (5.18)\\ \mu_{1j} &= \lambda_{0j} p_{0j} + \mu_{0j},\ j \geq 1, & (5.19)\\ \lambda_{00} p_{00} + \mu_{01} &= \mu_{10} + \lambda_{10} p_{10}, & (5.20)\\ \lambda_{0j} p_{0j} + \lambda_{1,j-1} p_{1,j-1} + \mu_{0,j+1} &= \mu_{1j} + \lambda_{1j} p_{1j},\ j \geq 1 & (5.21) \end{aligned}$$

that correspond to subsystems $(0,0)$, $(0,j)$, *with* $j \geq 1$, $(1,0)$, *and* $(1,j)$, *with* $j \geq 1$, *respectively.*

Proof. The 'in' rate at subsystem $(0, 0)$ is μ_{10}, since an entrance of the pointer at $(0, 0)$ happens only when a service completion occurs that changes the state of the original system from $(1, 0)$ to $(0, 0)$. The 'out' rate at subsystem $(0, 0)$ is $\lambda_{00} p_{00}$ since a departure of the pointer from $(0, 0)$ happens only when an arrival occurs that changes the state of the original system from $(0, 0)$ to $(1, 0)$. By equating these two rates we obtain (5.18).

Similarly, for a subsystem $(0, j)$ with $j \geq 1$, we have that the 'in' rate is μ_{1j}, whereas the 'out' rate is $\lambda_{0j} p_{0j} + \mu_{0j}$, since a departure of the pointer from subsystem $(0, j)$ occurs either because of new customer arrival or the completion of a seeking time, leading to (5.19). Similarly, for subsystem $(1, 0)$, we have that the 'in' rate is $\lambda_{00} p_{00} + \mu_{01}$, whereas the 'out' rate is $\mu_{10} + \lambda_{10} p_{10}$. Equating these rates yields (5.20).

Finally, the 'in' rate for a subsystem $(1, j)$, with $j \geq 1$, is $\lambda_{0j} p_{0j} + \lambda_{1,j-1} p_{1,j-1} + \mu_{0,j+1}$, whereas the 'out' rate is $\mu_{1j} + \lambda_{1j} p_{1j}$. Again, equating the two rates yields (5.21). ∎

Note that (5.18)-(5.21) can be simplified considerably to get:

Corollary 5.2 *The steady-state probabilities p_{ij} and rates μ_{ij} satisfy the 2nd QMCD system of equations:*

$$\begin{aligned} \lambda_{00} p_{00} &= \mu_{10}, & (5.22)\\ \lambda_{0j} p_{0j} &= \mu_{1j} - \mu_{0j},\ j \geq 1, & (5.23)\\ \lambda_{1j} p_{1j} &= \mu_{0,j+1},\ j \geq 0. & (5.24) \end{aligned}$$

This system can be also derived by equating upcrossing and downcrossing rates for each possible number j of customers in the original system and its orbit. For example, let $\Sigma_1 = \{(i,k) : i = 0,1 \text{ and } k \leq j\}$ and $\Sigma_2 = \{(i,k) : i = 0,1 \text{ and } k > j\}$. The pointer can move from Σ_1 to Σ_2 only if an arrival that finds the system at $(1,j)$ occurs, i.e., at rate $\lambda_{1j} p_{1j}$. Moreover, the pointer can move from Σ_2 to Σ_1 only if a seeking time completion occurs with $j+1$ customers in the orbit, i.e., at rate $\mu_{0,j+1}$. Equating the rates, we obtain (5.24). However, we preferred to expose here the more general idea of equating the 'in' and 'out' rates of the corresponding subsystems, since it can be applied in other models as well. The 2nd QMCD system can be used to express the steady-state probabilities p_{ij} in terms of the rates μ_{ij}. Moreover, an implication of (5.24) is that ν_{1j} from (5.1) becomes

$$\nu_{1j} = \mu_{0,j+1} + \lambda_{0j} p_{0j} = \lambda_{1j} p_{1j} + \lambda_{0j} p_{0j}. \tag{5.25}$$

5.4 QMCD - Step 4: Apply Little's law for each subsystem

In this step we apply Little's law at each subsystem. For deriving the mean sojourn time of the pointer at each subsystem, we will need an auxiliary result that is presented in Lemma 5.2.

Lemma 5.2 *(Expected value of the minimum of an exponential and a general random variable) Let X be a non-negative generally distributed random variable with LST $\tilde{F}(s)$ and T_λ an exponential random variable with rate λ, independent of X. Then*

$$E[\min(X, T_\lambda)] = \frac{1 - \tilde{F}(\lambda)}{\lambda}. \tag{5.26}$$

Proposition 5.4 *(Applying Little's law at each subsystem) The steady-state probabilities satisfy the following formulas:*

$$p_{00} = \frac{\mu_{10}}{\lambda_{00}}, \tag{5.27}$$

$$p_{0j} = \frac{\mu_{1j}(1 - \tilde{A}_j(\lambda_{0j}))}{\lambda_{0j}}, \; j \geq 1, \tag{5.28}$$

$$p_{10} = \frac{\lambda_{00} p_{00}(1 - \tilde{B}_{10}(\lambda_{10})) + \mu_{01}(1 - \tilde{B}_{00}(\lambda_{10}))}{\lambda_{10}}, \tag{5.29}$$

$$p_{1j} = \frac{\lambda_{0j} p_{0j}(1 - \tilde{B}_{1j}(\lambda_{1j})) + \lambda_{1,j-1} p_{1,j-1}(1 - \tilde{S}_{1,j-1}(\lambda_{1j})) + \mu_{0,j+1}(1 - \tilde{B}_{0j}(\lambda_{1j}))}{\lambda_{1j}}, \; j \geq 1. \tag{5.30}$$

Proof. Applying Little's law at each subsystem (i,j) yields

$$E[Q_{ij}] = \Lambda_{ij} E[W_{ij}], \; i = 0,1, \; j \geq 0, \tag{5.31}$$

where $E[Q_{ij}]$ is the mean number of customers in subsystem (i,j), Λ_{ij} is the arrival rate at subsystem (i,j) and $E[W_{ij}]$ the mean sojourn time at subsystem (i,j). Since, there is only one customer (the pointer) that circulates in the network of the subsystems, we have that Q_{ij} is 1 with probability p_{ij} and 0 with the complementary probability $1 - p_{ij}$. Therefore,

$$E[Q_{ij}] = p_{ij}, \; i = 0,1, \; j \geq 0. \tag{5.32}$$

For a given subsystem (i,j), the arrival rate Λ_{ij} is the 'in' rate at this subsystem. The left-hand sides of (5.18)-(5.21) give

$$\Lambda_{0j} = \mu_{1j}, \; j \geq 0, \tag{5.33}$$

$$\Lambda_{10} = \lambda_{00} p_{00} + \mu_{01}, \tag{5.34}$$

$$\Lambda_{1j} = \lambda_{0j} p_{0j} + \lambda_{1,j-1} p_{1,j-1} + \mu_{0,j+1}, \; j \geq 1. \tag{5.35}$$

We now derive expressions for the mean sojourn times $E[W_{ij}]$. When the pointer enters subsystem $(0,0)$, then the system is empty, so the pointer will leave this subsystem at the time of the first customer arrival that happens after an exponential time with rate λ_{00}. Therefore

$$E[W_{00}] = \frac{1}{\lambda_{00}}. \tag{5.36}$$

When the pointer enters a subsystem $(0,j)$ with $j \geq 1$, then a seeking time starts anew, so the pointer will leave the subsystem when the seeking time ends or a new customer arrives, whatever occurs first. The seeking time is distributed as A_j, with distribution $A_j(x)$, whereas the time till the next new customer's arrival (if nothing else happens) is exponentially distributed with rate λ_{0j}. Moreover, these two times are independent. Using (5.26), we conclude that

$$E[W_{0j}] = E[\min(A_j, T_{\lambda_{0j}})] = \frac{1 - \tilde{A}_j(\lambda_{0j})}{\lambda_{0j}},\ j \geq 1. \tag{5.37}$$

Finally, the pointer may enter a subsystem $(1,j)$ with $j \geq 1$, either from the subsystem $(0,j)$ due to a new arrival, or from the subsystem $(1,j-1)$ due to a new arrival, or from the subsystem $(0,j+1)$ due to a seeking time completion. The corresponding probabilities are $\frac{\lambda_{0j}p_{0j}}{\Lambda_{1j}}$, $\frac{\lambda_{1,j-1}p_{1,j-1}}{\Lambda_{1j}}$ and $\frac{\mu_{0,j+1}}{\Lambda_{1j}}$.

Conditioning on the arrival to subsystem $(1,j)$ from each of the subsystems the analysis is similar to the one leading to (5.37). That is, there is a competition of an exponential time to next arrival with rate λ_{1j} with an independent new service time B_{1j}, residual service time $S_{1,j-1}$, and new service time B_{0j}, when arriving from subsystems $(0,j)$, $(1,j-1)$ and $(0,j+1)$, respectively. Then, using the above probabilities we get:

$$E[W_{1j}] = \frac{\lambda_{0j}p_{0j}}{\Lambda_{1j}} \cdot \frac{1 - \tilde{B}_{1j}(\lambda_{1j})}{\lambda_{1j}} + \frac{\lambda_{1,j-1}p_{1,j-1}}{\Lambda_{1j}} \cdot \frac{1 - \tilde{S}_{1,j-1}(\lambda_{1j})}{\lambda_{1j}} + \frac{\mu_{0,j+1}}{\Lambda_{1j}} \cdot \frac{1 - \tilde{B}_{0j}(\lambda_{1j})}{\lambda_{1j}},\ j \geq 1. \tag{5.38}$$

For $j = 0$, i.e., for $E[W_{10}]$, there are no entrances from subsystem $(1,-1)$, i.e., no middle term in (5.38), so:

$$E[W_{10}] = \frac{\lambda_{00}p_{00}}{\Lambda_{10}} \cdot \frac{1 - \tilde{B}_{10}(\lambda_{10})}{\lambda_{10}} + \frac{\mu_{01}}{\Lambda_{10}} \cdot \frac{1 - \tilde{B}_{00}(\lambda_{10})}{\lambda_{10}}. \tag{5.39}$$

Now, Little's law, i.e., (5.31), applied to subsystems $(0,0)$, $(0,j)$, $(1,0)$ and $(1,j)$, taking into account the various expressions for $E[Q_{ij}]$, Λ_{ij} and $E[W_{ij}]$ yields (5.27)-(5.30) ■

Multiplying each of (5.28)-(5.30) by the denominator of its right-hand side (RHS), separating the terms that involve LSTs from the other terms, simplifying using (5.19)-(5.21) and taking into account (5.3), yields Corollary 5.3.

Corollary 5.3 *The steady-state probabilities p_{ij}, rates μ_{ij} and LSTs $\tilde{S}_{ij}(s)$ satisfy the 3rd QMCD system of equations:*

$$\begin{aligned}
\mu_{0j} &= \mu_{1j}\tilde{A}_j(\lambda_{0j}),\ j \geq 1, &(5.40)\\
\mu_{10} &= \lambda_{00}p_{00}\tilde{B}_{10}(\lambda_{10}) + \mu_{01}\tilde{B}_{00}(\lambda_{10}) = \nu_{10}\tilde{B}_0(\lambda_{10}), &(5.41)\\
\mu_{1j} &= \lambda_{0j}p_{0j}\tilde{B}_{1j}(\lambda_{1j}) + \lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}) + \mu_{0,j+1}\tilde{B}_{0j}(\lambda_{1j}) &\\
&= \nu_{1j}\tilde{B}_j(\lambda_{1j}) + \lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}),\ j \geq 1. &(5.42)
\end{aligned}$$

5.5 QMCD - Recursive schemes

The three QMCD systems (Corollaries 5.1, 5.2 and 5.3), in combination with the formula for the probability a_j (Proposition 5.2) provide sufficient number of equations to compute all the quantities of interest $\tilde{S}_{ij}(s)$, μ_{ij} and p_{ij}, given that p_{00} is known. In practice p_{00} is treated as a multiplicative constant that is computed at the end of all computations, using the normalization equation. The recursive scheme that computes the LSTs $\tilde{S}_{ij}(s)$ and the rates μ_{ij} is given in the following theorem.

Theorem 5.1 *(Recursive scheme for the rates μ_{ij} and the LSTs $\tilde{S}_{ij}(s)$) Given the probability p_{00}, the quantities μ_{ij} and $\tilde{S}_{1j}(s)$ can be computed by the recursive scheme*

$$\mu_{0,j+1} = \frac{1-\tilde{B}_{1j}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})}\mu_{1j} + \frac{\tilde{B}_{1j}(\lambda_{1j}) - \tilde{S}_{1,j-1}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})}\mu_{0j}, \ j \geq 1, \tag{5.43}$$

$$\mu_{1,j+1} = \frac{1}{\tilde{A}_{j+1}(\lambda_{0,j+1})}\mu_{0,j+1}, \ j \geq 1, \tag{5.44}$$

$$\tilde{S}_{1j}(s) = \frac{\lambda_{1j}}{s-\lambda_{1j}}\left(a_j\frac{\tilde{B}_j(\lambda_{1j})-\tilde{B}_j(s)}{1-\tilde{B}_j(\lambda_{1j})} + (1-a_j)\frac{\tilde{S}_{1,j-1}(\lambda_{1j})-\tilde{S}_{1,j-1}(s)}{1-\tilde{S}_{1,j-1}(\lambda_{1j})}\right), \ j \geq 1, \tag{5.45}$$

where

$$\begin{aligned} a_j &= \left(1-\tilde{B}_j(\lambda_{1j})\right)\left(1+\frac{\mu_{1j}-\mu_{0j}}{\mu_{0,j+1}}\right) \\ &= \frac{\mu_{0,j+1}\left(1-\tilde{B}_{0j}\left(\lambda_{1j}\right)\right) + \left(\mu_{1j}-\mu_{0j}\right)\left(1-\tilde{B}_{1j}\left(\lambda_{1j}\right)\right)}{\mu_{0,j+1}}, \ j \geq 1, \end{aligned} \tag{5.46}$$

and initial conditions

$$\mu_{10} = \lambda_{00}p_{00}, \tag{5.47}$$

$$\mu_{01} = \frac{\lambda_{00}(1-\tilde{B}_{10}(\lambda_{10}))}{\tilde{B}_{00}(\lambda_{10})}p_{00}, \tag{5.48}$$

$$\mu_{11} = \frac{\lambda_{00}(1-\tilde{B}_{10}(\lambda_{10}))}{\tilde{A}_1(\lambda_{01})\tilde{B}_{00}(\lambda_{10})}p_{00}, \tag{5.49}$$

$$\tilde{S}_{10}(s) = \frac{\lambda_{10}}{s-\lambda_{10}} \cdot \frac{\tilde{B}_0(\lambda_{10})-\tilde{B}_0(s)}{1-\tilde{B}_0(\lambda_{10})}. \tag{5.50}$$

Moreover, we have

$$\tilde{S}_{0j}\left(s\right) = \frac{\lambda_{0j}}{s-\lambda_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j})-\tilde{A}_j(s)}{1-\tilde{A}_j(\lambda_{0j})}, \ j \geq 1. \tag{5.51}$$

Note that the initial conditions (5.47)-(5.50) as well as (5.51) can all be calculated based upon the problem's primitives. Using their values, we can calculate (5.43) followed by (5.44), for $j=1$. The second line of (5.46) can now also be calculated and is used in calculating (5.45), for $j=1$. And then, given the quantities $\mu_{0,j+1}$, $\mu_{1,j+1}$, $\tilde{S}_{1j}$ and a_j for a given j, we can follow another recursive step, using (5.43)-(5.46), and find them for $j+1$. We can now proceed to a recursive scheme for the steady state probabilities p_{ij}.

Theorem 5.2 *(Recursive scheme for the steady-state probabilities) Given* p_{00}*, the steady-state probabilities* p_{ij} *can be computed by the recursive scheme*

$$p_{0j} = \frac{\lambda_{1,j-1}\left(1-\tilde{A}_j\left(\lambda_{0j}\right)\right)}{\lambda_{0j}\tilde{A}_j\left(\lambda_{0j}\right)} p_{1,j-1}, \; j \geq 1, \tag{5.52}$$

$$\begin{aligned} p_{1j} &= \frac{\lambda_{0j}\left(1-\tilde{B}_{1j}\left(\lambda_{1j}\right)\right)}{\lambda_{1j}\tilde{B}_{0j}\left(\lambda_{1j}\right)} p_{0j} + \frac{\lambda_{1,j-1}\left(1-\tilde{S}_{1,j-1}\left(\lambda_{1j}\right)\right)}{\lambda_{1j}\tilde{B}_{0j}\left(\lambda_{1j}\right)} p_{1,j-1} \\ &= \frac{\lambda_{0j}\left(1-\tilde{B}_{j}\left(\lambda_{1j}\right)\right)}{\lambda_{1j}\tilde{B}_{j}\left(\lambda_{1j}\right)} p_{0j} + \frac{\lambda_{1,j-1}\left(1-\tilde{S}_{1,j-1}\left(\lambda_{1j}\right)\right)}{\lambda_{1j}\tilde{B}_{j}\left(\lambda_{1j}\right)} p_{1,j-1}, \; j \geq 1, \end{aligned} \tag{5.53}$$

with $\tilde{S}_{1j}(s)$ *given in* (5.45) *and* (5.50) *and the initial condition*

$$p_{10} = \frac{\lambda_{00}\left(1-\tilde{B}_{10}\left(\lambda_{10}\right)\right)}{\lambda_{10}\tilde{B}_{00}\left(\lambda_{10}\right)} p_{00} = \frac{\lambda_{00}\left(1-\tilde{B}_{0}\left(\lambda_{10}\right)\right)}{\lambda_{10}\tilde{B}_{0}\left(\lambda_{10}\right)} p_{00}. \tag{5.54}$$

Next corollary provides the dependence of μ_{ij}, $\tilde{S}_{ij}(s)$, $\tilde{B}_j(s)$, ν_{1j} and p_{ij} on p_{00}. This result will be used in the following subsection to derive the stability condition and compute the quantity p_{00}.

Corollary 5.4 *The rates* μ_{ij} *and* ν_{1j}*, and the probabilities* p_{ij} *are linear in* p_{00}*. The LSTs* $\tilde{S}_{ij}(s)$ *and* $\tilde{B}_j(s)$ *are independent of* p_{00}*.*

5.6 QMCD - Step 5: Use the normalization equation

In this step we note that whenever the system is stable, the normalization equation holds, i.e.,

$$\sum_{i=0}^{1}\sum_{j=0}^{\infty} p_{ij} = 1. \tag{5.55}$$

From Corollary 5.4, we have that the probabilities p_{ij} are linear in p_{00}. Thus, the normalization equation enables to determine the probability p_{00} and derive the stability condition of the system. The results are stated in the following theorem.

Theorem 5.3 *(Stability condition and the steady-state probability* p_{00}*) Let* k_j*,* l_j*, for* $j \geq 1$*, and* q *defined by*

$$k_j = \frac{\lambda_{1,j-1}\left[\left(1-\tilde{B}_{1j}(\lambda_{1j})\right)\left(1-\tilde{A}_j(\lambda_{0j})\right) + \tilde{A}_j(\lambda_{0j})\left(1-\tilde{S}_{1,j-1}(\lambda_{1j})\right)\right]}{\lambda_{1j}\tilde{B}_{0j}(\lambda_{1j})\tilde{A}_j(\lambda_{0j})}, \; j \geq 1, \tag{5.56}$$

$$l_j = \frac{\lambda_{1,j-1}\left(1-\tilde{A}_j(\lambda_{0j})\right)}{\lambda_{0j}\tilde{A}_j(\lambda_{0j})}, \; j \geq 1 \tag{5.57}$$

and

$$q = \frac{\lambda_{00}\left(1-\tilde{B}_{10}(\lambda_{10})\right)}{\lambda_{10}\tilde{B}_{00}(\lambda_{10})}. \tag{5.58}$$

The stability condition for the system is

$$\sum_{j=1}^{\infty}(k_j + l_j)\prod_{i=1}^{j-1} k_i < \infty \tag{5.59}$$

and, under the stability condition, the stationary probability p_{00} is given by

$$p_{00} = \frac{1}{1 + q + q\sum_{j=1}^{\infty}(k_j + l_j)\prod_{i=1}^{j-1} k_i}. \tag{5.60}$$

5.7 The state-independent case

We now consider the state-independent case, where $\lambda_{ij} = \lambda$, $B_{ij}(x) = B(x)$ and $A_j(x) = A(x)$, for all i, j. Then, using the previous results, we can derive the stability condition of this system and the probability generating function of the stationary distribution of the number of customers in system. Thus, we have alternative proofs of the main results of Gomez-Corral (1999), where the state-independent case was treated (see Theorems 1 and 2 of that paper, in particular equation (16)).

Theorem 5.4 *(Stability condition and probability generating function of the steady-state number of customers in system) The stability condition for the system in the state-independent case is*

$$\lambda E[B] < \tilde{A}(\lambda). \tag{5.61}$$

Let $K(z)$ be the generating function of the number of customers in system in steady-state. Then, under the stability condition,

$$K(z) = \frac{(\tilde{A}(\lambda) - \lambda E[B])(1-z)\tilde{B}(\lambda(1-z))}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z))}. \tag{5.62}$$

6 Conclusion and Future Work

In this paper we presented the basic steps of the QMCD approach for the study of queueing systems with state-dependencies and generally distributed times and applied this methodology for the performance evaluation of the state-dependent $M/G/1$ queue with orbit. The QMCD approach constitutes a probabilistic alternative to the classical analytic approaches that use supplementary variables and transform methods for the analysis of queueing systems with generally distributed inter-arrival and service times. Among its advantages, we can list the following:

1. The QMCD approach provides a direct relationship of the conditional remaining service times of customers that find n and $n-1$ customers upon arrival at the system, in terms of random variables, rather than in terms of transforms. Therefore, it outperforms the analytic methods (e.g., the supplementary variable method) in explanatory power, as it provides a direct link between the system dynamics and the recursive relationships of conditional remaining service times of customers.

2. Relating the conditional remaining service times in terms of random variables enables the use of sample path arguments for the study of stochastic comparison issues in the framework of the underlying system.

3. The arguments of the probabilistic approach are still valid even when the underlying distributions are not absolutely continuous. Indeed, the probabilistic approach uses relations of random variables that are then interpreted in terms of Laplace-Stieltjes transforms. In particular the existence of probability densities is not required for the generally distributed times, in contrast to the supplementary variables method.

4. The probabilistic approach seems more economic, in the sense that it avoids long calculations that are indispensable when applying the analytic methods.

5. The QMCD approach is applicable also in settings where the supplementary variables method is not, e.g., see its application in Wang, Baron and Scheller-Wolf (2015) for a multi-server queue, and in Abouee-Mehrizi, Baron and Berman (2014) for a queueing network.

Apart from the methodological contribution, the study of the state-dependent $M/G/1$ queue with orbit allowed us to determine the customer equilibrium behavior, regarding the joining/balking dilemma, in the observable version of this queueing system. The results are reported in Baron, Economou and Manou (2017).

The next steps for future work regarding the methodological dimension of the paper is to characterize which systems are solvable using the QMCD approach and how it can be extended by considering subsystems that do not correspond to single states, e.g. in multi-server queues with priorities.

Acknowledgments: We thank the Associate Editor, and two anonymous reviewers for their valuable comments that greatly improved the paper. We thank Professor Mor Harchol-Balter of Carnegie Mellon University for pointing us to the application of our model for the analysis of load sharing algorithms.

References

[1] Abouee-Mehrizi, H. and Baron, O. (2016) State-dependent $M/G/1$ queueing systems. *Queueing Systems* **82**, 121-148.

[2] Abouee-Mehrizi, H., Baron, O. and Berman, O. (2014) Exact analysis of capacitated two-echelon inventory systems with priorities. *Manufacturing & Service Operations Management* **16**, 561-577.

[3] Artalejo, J.R. and Gomez-Corral, A. (2008) *Retrial Queueing Systems, A Computational Approach.* Springer.

[4] Baron, O., Economou, A. and Manou, A. (2017) Equilibrium customer strategies for joining the M/G/1 queue with orbit. *Working paper.*

[5] Bitran, G.R. and Tirupati, D. (1988) Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference. *Management Science* **34**, 75-100.

[6] Caldentey, R. (2001) Approximations for multi-class departure processes. *Queueing Systems* **38**, 205-212.

[7] Choi, B.D., Park, K.K. and Pearce, C.E.M. (1993) An $M/M/1$ retrial queue with control policy and general retrial times. *Queueing Systems* **14**, 275-292.

[8] Cox, D.R. (1955) A use of complex probabilities in the theory of stochastic processes. *Proceedings of the Cambridge Philosophical Society* **51**, 313-319.

[9] Doroudi, S., Fralix, B. and Harchol-Balter, M. (2016) Clearing analysis of phases: Exact limiting probabilities for skip-free, unidirectional, quasi-birth-death processes. *Stochastic Systems* **6**, 420-458.

[10] Eager, D.L., Lazowska, E.D. and Zahorjan, J. (1986) Adaptive load sharing in homogeneous distributed systems. *IEEE Transactions in Software Engineering* **12**, 662-675.

[11] Economou, A. and Manou, A. (2015) A probabilistic approach for the analysis of the $M_n/G/1$ queue. To appear in *Annals of Operations Research.*

[12] Falin, G.I. and Templeton, J.G.C. (1997) *Retrial Queues.* Chapman and Hall, London.

[13] Fayolle, G. (1986) A simple telephone exchange with delayed feedbacks. In Boxma, O.J., Cohen, J.W. and Tijms, H.C. (editors) *Teletraffic Analysis and Computer Performance Evaluation*, Elsevier, Amsterdam, pp. 245-253.

[14] Gomez-Corral, A. (1999) Stochastic analysis of a single server retrial queue with general retrial times. *Naval Research Logistic* **46**, 561-581.

[15] Kerner, Y. (2008) The conditional distribution of the residual service time in the $M_n/G/1$ queue. *Stochastic Models* **24**, 364-375.

[16] Lazowska, E.D., Zahorjan, J., Cheriton, D.R. and Zwaencpocl, W. (1984) File access performance of diskless workstations. *Department of Computer Science, University of Washington, Seattle, Technical Report* 84-06-01.

[17] Manou, A., Economou, A. and Karaesmen, F. (2014) Strategic customers in a transportation station: When is it optimal to wait? *Operations Research* **62**, 910-925.

[18] Martin, M. and Artalejo, J.R. (1995) Analysis of an $M/G/1$ queue with two types of impatient units. *Advances in Applied Probability* **27**, 840-861.

[19] Oz, B., Adan, I. and Haviv, M. (2017) A rate balance principle and its application to queueing models. *Queueing Systems* **87**, 95-111.

[20] Oz, B., Adan, I. and Haviv, M. (2015) The conditional distribution of the remaining service or vacation time in the $M_n/G_n/1$ queue with vacations. *Preprint.*

[21] Oz, B. (2016) *Strategic Behavior in Queues.* PhD Thesis, The Hebrew University of Jerusalem.

[22] Phung-Duc, T., Rogiest, W. and Wittevrongel, S. (2017) Single server retrial queues with speed scaling: Analysis and performance evaluation. To appear in *Journal of Industrial and Management Optimization.*

[23] Phung-Duc, T. (2014) Server farms with batch arrival and staggered setup. *Proceedings of the 15th Symposium on Information and Communication Technology (SoICT), Hanoi, December 4-5, 2014*, 240-247.

[24] Phung-Duc, T. (2017a) Exact solutions for $M/M/c/$Setup queues. *Telecommunication Systems* **64**, 309-324.

[25] Phung-Duc, T. (2017b) Single server retrial queues with setup time. *Journal of Industrial and Management Optimization* **13**, 1329-1345.

[26] Reiser, M. and Kobayashi, H. (1974) Accuracy of diffusion approximation for some queueing systems. *IBM Journal of Research and Development* **18**, 110-124.

[27] van Doorn, E.A. and Regterschot, D.J.K. (1988) Conditional PASTA. Operations Research Letters **7**, 229-232.

[28] Wang, J., Baron, O. and Scheller-Wolf, A. (2015) $M/M/c$ queue with two priority classes. *Operations Research* **63**, 733-749.

[29] Yajima, M. and Phung-Duc, T. (2017) Batch arrival single-server queue with variable service speed and setup time. *Queueing Systems* **86**, 241-260.

7 Appendix: Technical analytic proofs

In the Appendix, we provide several technical analytic proofs of various results of the paper. The proofs are quite tedious but straightforward, depending mainly on algebraic manipulations and they are given for completeness.

Corollary 5.2 - Proof. In light of (5.18), (5.20) yields

$$\mu_{01} = \lambda_{10} p_{10}. \tag{7.1}$$

Now, by substituting μ_{1j} by the RHS of (5.19) into (5.21) and canceling $\lambda_{0j}p_{0j}$ yields $\lambda_{1,j-1}p_{1,j-1}+\mu_{0,j+1} = \mu_{0j} + \lambda_{1j}p_{1j}$, for $j \geq 1$, that can be written as

$$\mu_{0,j+1} - \mu_{0j} = \lambda_{1j}p_{1j} - \lambda_{1,j-1}p_{1,j-1},\ j \geq 1. \tag{7.2}$$

Summing (7.2) for consecutive values of j and canceling equal terms yields

$$\mu_{0,j+1} - \mu_{01} = \lambda_{1j}p_{1j} - \lambda_{10}p_{10},\ j \geq 1, \tag{7.3}$$

which reduces to $\mu_{0,j+1} = \lambda_{1j}p_{1j}$, for $j \geq 1$, using (7.1). ∎

Lemma 5.2 - Proof. We have

$$\begin{aligned} E[\min(X,T_\lambda)] &= \int_0^\infty \Pr[\min(X,T_\lambda) > x]dx = \int_0^\infty \Pr[X > x]e^{-\lambda x}dx \\ &= \int_0^\infty \int_x^\infty dF(y)e^{-\lambda x}dx = \int_0^\infty \int_0^y e^{-\lambda x}dxdF(y) \\ &= \int_0^\infty \frac{1-e^{-\lambda y}}{\lambda}dF(y) = \frac{1-\tilde{F}(\lambda)}{\lambda}. \end{aligned}$$

∎

Theorem 5.1 - Proof. We first establish the initial conditions. The initial condition (5.47) is simply (5.22). By (5.47) and (5.41) we have that

$$\lambda_{00}p_{00} = \lambda_{00}p_{00}\tilde{B}_{10}(\lambda_{10}) + \mu_{01}\tilde{B}_{00}(\lambda_{10}),$$

which yields (5.48). By (5.40) for $j = 1$ we have

$$\mu_{01} = \mu_{11}\tilde{A}_1(\lambda_{01}),$$

which, in combination with (5.48), yields (5.49). Also, (5.50) is identical to (5.9).

Now, solving (5.42) for $\mu_{0,j+1}$ yields (5.43). Indeed, we have that

$$\begin{aligned} \mu_{0,j+1} &= \frac{\mu_{1j} - \lambda_{0j}p_{0j}\tilde{B}_{1j}(\lambda_{1j}) - \lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})} \\ &= \frac{\mu_{1j} - (\mu_{1j} - \mu_{0j})\tilde{B}_{1j}(\lambda_{1j}) - \mu_{0j}\tilde{S}_{1,j-1}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})} \\ &= \mu_{1j}\frac{1-\tilde{B}_{1j}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})} + \mu_{0j}\frac{\tilde{B}_{1j}(\lambda_{1j}) - \tilde{S}_{1,j-1}(\lambda_{1j})}{\tilde{B}_{0j}(\lambda_{1j})},\ j \geq 1, \end{aligned}$$

where in the second equality we used the equations (5.23) and (5.24). Now, (5.44) immediately follows from (5.40) for $j \geq 1$ and (5.45) is identical to (5.10). Substituting (5.1) into (5.14) and using (5.23) yields

$$\begin{aligned} a_j &= \frac{\mu_{0,j+1}\left(1-\tilde{B}_j(\lambda_{1j})\right)+\lambda_{0j}p_{0j}\left(1-\tilde{B}_j(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}} \\ &= \frac{\mu_{0,j+1}\left(1-\tilde{B}_j(\lambda_{1j})\right)+(\mu_{1j}-\mu_{0j})\left(1-\tilde{B}_j(\lambda_{1j})\right)}{\mu_{0,j+1}} \\ &= \left(1-\tilde{B}_j(\lambda_{1j})\right)\left(1+\frac{\mu_{1j}-\mu_{0j}}{\mu_{0,j+1}}\right), \end{aligned}$$

which gives the first line in (5.46). The second line follows using (5.1), (5.3) and (5.23). Indeed, we have

$$\begin{aligned} a_j &= \left(1-\frac{\mu_{0,j+1}\tilde{B}_{0j}\left(\lambda_{1j}\right)+\left(\mu_{1j}-\mu_{0j}\right)\tilde{B}_{1j}\left(\lambda_{1j}\right)}{\mu_{0,j+1}+\mu_{1j}-\mu_{0j}}\right)\frac{\mu_{0,j+1}+\mu_{1j}-\mu_{0j}}{\mu_{0,j+1}} \\ &= \left(\frac{\mu_{0,j+1}+\mu_{1j}-\mu_{0j}-\mu_{0,j+1}\tilde{B}_{0j}\left(\lambda_{1j}\right)-\left(\mu_{1j}-\mu_{0j}\right)\tilde{B}_{1j}\left(\lambda_{1j}\right)}{\mu_{0,j+1}+\mu_{1j}-\mu_{0j}}\right)\frac{\mu_{0,j+1}+\mu_{1j}-\mu_{0j}}{\mu_{0,j+1}} \\ &= \frac{\mu_{0,j+1}\left(1-\tilde{B}_{0j}\left(\lambda_{1j}\right)\right)+\left(\mu_{1j}-\mu_{0j}\right)\left(1-\tilde{B}_{1j}\left(\lambda_{1j}\right)\right)}{\mu_{0,j+1}}. \end{aligned}$$

Finally, (5.51) is identical to (5.8). ∎

Theorem 5.2 - Proof. Using (5.23) and (5.24), we deduce that

$$\mu_{1j}=\lambda_{0j}p_{0j}+\mu_{0j}=\lambda_{0j}p_{0j}+\lambda_{1,j-1}p_{1,j-1},\ j\geq 1. \tag{7.4}$$

Equations (5.24) and (5.40) yield

$$\lambda_{1,j-1}p_{1,j-1}=\mu_{0j}=\mu_{1j}\tilde{A}_j(\lambda_{0j}),\ j\geq 1,$$

so, using (7.4), we conclude that

$$\lambda_{1,j-1}p_{1,j-1}=\left(\lambda_{0j}p_{0j}+\lambda_{1,j-1}p_{1,j-1}\right)\tilde{A}_j(\lambda_{0j}),\ j\geq 1. \tag{7.5}$$

Solving for p_{0j} yields (5.52). Next, equations (5.41) and (5.22) imply that

$$\lambda_{00}p_{00}=\mu_{01}\tilde{B}_{00}\left(\lambda_{10}\right)+\lambda_{00}p_{00}\tilde{B}_{10}\left(\lambda_{10}\right)=\nu_{10}\tilde{B}_0\left(\lambda_{10}\right)=\left(\lambda_{10}p_{10}+\lambda_{00}p_{00}\right)\tilde{B}_0\left(\lambda_{10}\right),$$

where the second and third equality follow by (5.3) and (5.25), respectively. Noting that $\mu_{01}=\lambda_{10}p_{10}$ from (5.24) and solving for p_{10} yields both expressions in (5.54).

Finally, substituting (7.4) and (5.24) into the left hand side and right hand side of (5.42) respectively, we get

$$\begin{aligned} \lambda_{0j}p_{0j}+\lambda_{1,j-1}p_{1,j-1} &= \lambda_{1j}p_{1j}\tilde{B}_{0j}(\lambda_{1j})+\lambda_{0j}p_{0j}\tilde{B}_{1j}(\lambda_{1j})+\lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}) \\ &= \nu_{1j}\tilde{B}_j\left(\lambda_{1j}\right)+\lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}\left(\lambda_{1j}\right),\ j\geq 1. \end{aligned} \tag{7.6}$$

Rearranging terms in the first equality yields

$$\lambda_{1j}p_{1j}\tilde{B}_{0j}(\lambda_{1j})=\lambda_{0j}p_{0j}\left(1-\tilde{B}_{1j}\left(\lambda_{1j}\right)\right)+\lambda_{1,j-1}p_{1,j-1}\left(1-\tilde{S}_{1,j-1}\left(\lambda_{1j}\right)\right),\ j\geq 1.$$

Solving for p_{1j} yields the first line in (5.53). Substituting (5.25) in the second line of (7.6) implies that

$$\lambda_{0j}p_{0j} + \lambda_{1,j-1}p_{1,j-1} = (\lambda_{1j}p_{1j} + \lambda_{0j}p_{0j})\,\tilde{B}_j(\lambda_{1j}) + \lambda_{j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}),\ j \geq 1.$$

Solving for p_{1j} yields the second line in (5.53). ■

Corollary 5.4 - Proof. We are going to prove the result by induction. First, note that (5.47)-(5.49) imply that μ_{10}, μ_{01} and μ_{11} are linear in p_{00}. Also, (5.51) shows that LSTs $\tilde{S}_{0j}(s)$ are independent of p_{00}, for $j \geq 1$. Since μ_{01} is linear in p_{00}, (5.1) for $j = 1$ implies that ν_{10} is linear in p_{00}. Moreover, since ν_{10} and μ_{01} are linear in p_{00}, (5.3) implies that $\tilde{B}_0(s)$ is independent of p_{00} and consequently $\tilde{S}_{10}(s)$ is independent of p_{00}. From formulas (5.52) for $j = 1$ and (5.54), we have that p_{01} and p_{10} are linear in p_{00}.
We make the induction hypothesis: Assume that the rates $\mu_{01},\ \mu_{02}, \ldots,\ \mu_{0j}$, the rates $\mu_{10},\ \mu_{11}, \ldots,\ \mu_{1j}$, the rates $\nu_{10},\ \nu_{11}, \ldots,\ \nu_{1,j-1}$, the probabilities $p_{10},\ p_{11}, \ldots,\ p_{1,j-1}$, and the probabilities $p_{01},\ p_{02}, \ldots,\ p_{0j}$ are linear in p_{00}. Moreover, assume that the LSTs $\tilde{B}_0(s),\ \tilde{B}_1(s), \ldots,\ \tilde{B}_{j-1}(s)$ and $\tilde{S}_{10}(s),\ \tilde{S}_{11}(s), \ldots,\ \tilde{S}_{1,j-1}(s)$ are independent of p_{00}.
We will prove the following: (i) $\mu_{0,j+1}$ is linear in p_{00}, (ii) $\mu_{1,j+1}$ is linear in p_{00}, (iii) ν_{1j} is linear in p_{00}, (iv) $\tilde{B}_j(s)$ is independent of p_{00}, (v) $\tilde{S}_{1j}(s)$ is independent of p_{00}, (vi) p_{1j} is linear in p_{00} and (vii) $p_{0,j+1}$ is linear in p_{00}.
(i) Since μ_{0j} and μ_{1j} are linear in p_{00} and $\tilde{S}_{1,j-1}(s)$ is independent of p_{00}, (5.43) implies that $\mu_{0,j+1}$ is linear in p_{00}.
(ii) Using that $\mu_{0,j+1}$ is linear in p_{00}, (5.44) gives that $\mu_{1,j+1}$ is linear in p_{00}.
(iii) Since $\mu_{0,j+1}$ and p_{0j} are linear in p_{00}, (5.1) implies that ν_{1j} is linear in p_{00}.
(iv) Since $\mu_{0,j+1}$, p_{0j} and ν_{1j} are linear in p_{00}, (5.3) implies that $\tilde{B}_j(s)$ is independent of p_{00}.
(v) Using that μ_{1j}, μ_{0j}, and $\mu_{0,j+1}$ are linear in p_{00}, (5.46) gives that a_j is independent of p_{00}. Then, using that a_j, $\tilde{S}_{1,j-1}(s)$ and $\tilde{B}_j(s)$ are independent of p_{00}, (5.45) implies that $\tilde{S}_{1j}(s)$ is independent of p_{00}.
(vi) Since p_{0j} and $p_{1,j-1}$ are linear in p_{00} and $\tilde{S}_{1,j-1}(s)$ is independent of p_{00}, (5.53) shows that p_{1j} is linear in p_{00}.
(vii) (5.52), using that p_{1j} is linear in p_{00}, yields that $p_{0,j+1}$ is linear in p_{00}. ■

Theorem 5.3 - Proof. Using (5.53) in combination with (5.52) gives

$$p_{1j} = \frac{\lambda_{1,j-1}\left[\left(1-\tilde{B}_{1j}(\lambda_{1j})\right)\left(1-\tilde{A}_j(\lambda_{0j})\right) + \tilde{A}_j(\lambda_{0j})\left(1-\tilde{S}_{1,j-1}(\lambda_{1j})\right)\right]}{\lambda_{1j}\tilde{B}_{0j}(\lambda_{1j})\tilde{A}_j(\lambda_{0j})}p_{1,j-1}. \tag{7.7}$$

Thus,

$$p_{1j} = k_j p_{1,j-1},\ j \geq 1. \tag{7.8}$$

Also, (5.52) and (5.54) can be written as

$$p_{0j} = l_j p_{1,j-1},\ j \geq 1 \tag{7.9}$$

and

$$p_{10} = q p_{00}, \tag{7.10}$$

respectively. Then,

$$p_{0j} + p_{1j} \overset{(7.8),(7.9)}{=} (k_j + l_j)p_{1,j-1} \overset{(7.8)}{=} (k_j + l_j)\prod_{i=1}^{j-1} k_i p_{10} \overset{(7.10)}{=} q(k_j + l_j)\prod_{i=1}^{j-1} k_i p_{00},\ j \geq 1. \tag{7.11}$$

Normalization equation yields

$$
\begin{aligned}
& \sum_{j=0}^{\infty}(p_{0j}+p_{1j}) = 1 \\
\Rightarrow\quad & p_{00}+p_{10}+\sum_{j=1}^{\infty}(p_{0j}+p_{1j}) = 1 \\
\overset{(7.10),(7.11)}{\Rightarrow}\quad & p_{00}+qp_{00}+\sum_{j=1}^{\infty}q(k_j+l_j)\prod_{i=1}^{j-1}k_i p_{00} = 1 \\
\Rightarrow\quad & \left(1+q+q\sum_{j=1}^{\infty}(k_j+l_j)\prod_{i=1}^{j-1}k_i\right)p_{00} = 1.
\end{aligned}
$$

Thus, the stability condition is given by (5.59) and p_{00} is given by (5.60). ■

Theorem 5.4 - Proof. Let $\tilde{P}_{ij}(s) = p_{ij}\tilde{S}_{ij}(s)$ be the Laplace transform of the steady state density

$$
p_{ij}(r) = \lim_{t\to\infty}\lim_{dr\to 0^+}\frac{\Pr[C(t)=i, Q(t)=j, S(t)\in(r,r+dr]]}{dr},
$$

i.e.,

$$
\tilde{P}_{ij}(s) = \int_0^{\infty} e^{-sr}p_{ij}(r)dr.
$$

Multiplying (5.8)-(5.10) with the corresponding p_{ij}s yields the following equations for $\tilde{P}_{ij}(s)$:

$$
\tilde{P}_{0j}(s) = \frac{1}{s-\lambda}\left(\frac{\lambda p_{0j}\tilde{A}(\lambda)}{1-\tilde{A}(\lambda)} - \frac{\lambda p_{0j}}{1-\tilde{A}(\lambda)}\tilde{A}(s)\right),\ j\geq 1, \tag{7.12}
$$

$$
\tilde{P}_{10}(s) = \frac{1}{s-\lambda}\left(\frac{\lambda p_{10}\tilde{B}(\lambda)}{1-\tilde{B}(\lambda)} - \frac{\lambda p_{10}}{1-\tilde{B}(\lambda)}\tilde{B}(s)\right), \tag{7.13}
$$

$$
\begin{aligned}
\tilde{P}_{1j}(s) = \frac{1}{s-\lambda}\Bigg(&\frac{\lambda p_{1j}a_j\tilde{B}(\lambda)}{1-\tilde{B}(\lambda)} + \frac{\lambda p_{1j}(1-a_j)\tilde{S}_{1,j-1}(\lambda)}{1-\tilde{S}_{1,j-1}(\lambda)} \\
& - \frac{\lambda p_{1j}a_j}{1-\tilde{B}(\lambda)}\tilde{B}(s) - \frac{\lambda p_{1j}(1-a_j)}{p_{1,j-1}(1-\tilde{S}_{1,j-1}(\lambda))}\tilde{P}_{1,j-1}(s)\Bigg),\ j\geq 1.
\end{aligned} \tag{7.14}
$$

We simplify the coefficients appearing in these equations, using the formulas (5.22)-(5.24) and (5.40)-(5.42) of the 2nd and 3rd QMCD system. We have that

$$
\nu_{1j} = \mu_{0,j+1}+\lambda p_{0j} = \lambda p_{1j}+\lambda p_{0j}, \quad j\geq 1, \tag{7.15}
$$

$$
a_j = \frac{\nu_{1j}(1-\tilde{B}(\lambda))}{\lambda p_{1j}}, \quad j\geq 1, \tag{7.16}
$$

$$
1-a_j = \frac{\lambda p_{1,j-1}(1-\tilde{S}_{1,j-1}(s))}{\lambda p_{1j}}, \quad j\geq 1, \tag{7.17}
$$

so we can easily see (using (5.14), (5.22)-(5.24) and (5.40)-(5.42)) that

$$\frac{\lambda p_{0j}\tilde{A}(\lambda)}{1-\tilde{A}(\lambda)} = \mu_{0j}, \quad j \geq 1, \tag{7.18}$$

$$\frac{\lambda p_{0j}}{1-\tilde{A}(\lambda)} = \mu_{1j}, \quad j \geq 1, \tag{7.19}$$

$$\frac{\lambda p_{10}\tilde{B}(\lambda)}{1-\tilde{B}(\lambda)} = \mu_{10}, \tag{7.20}$$

$$\frac{\lambda p_{10}}{1-\tilde{B}(\lambda)} = \nu_{10}, \tag{7.21}$$

$$\frac{\lambda p_{1j}a_j}{1-\tilde{B}(\lambda)} = \nu_{1j}, \quad j \geq 1, \tag{7.22}$$

$$\frac{\lambda p_{1j}(1-a_j)}{p_{1,j-1}(1-\tilde{S}_{1,j-1}(\lambda))} = \lambda, \quad j \geq 1 \tag{7.23}$$

and

$$\frac{\lambda p_{1j}a_j\tilde{B}(\lambda)}{1-\tilde{B}(\lambda)} + \frac{\lambda p_{1j}(1-a_j)\tilde{S}_{1,j-1}(\lambda)}{1-\tilde{S}_{1,j-1}(\lambda)} = \mu_{1j}, \quad j \geq 1. \tag{7.24}$$

Substituting (7.18)-(7.24) in (7.12)-(7.14) yields

$$\tilde{P}_{0j}(s) = \frac{1}{s-\lambda}\left(\mu_{0j} - \mu_{1j}\tilde{A}(s)\right), \ j \geq 1, \tag{7.25}$$

$$\tilde{P}_{10}(s) = \frac{1}{s-\lambda}\left(\mu_{10} - \nu_{10}\tilde{B}(s)\right), \tag{7.26}$$

$$\tilde{P}_{1j}(s) = \frac{1}{s-\lambda}\left(\mu_{1j} - \nu_{1j}\tilde{B}(s) - \lambda\tilde{P}_{1,j-1}(s)\right) \ j \geq 1. \tag{7.27}$$

We now apply a generating function approach to determine the equilibrium distribution of the number of customers in the system. To this end, we introduce the generating functions

$$\tilde{P}_0(s,z) = \sum_{j=1}^{\infty}\tilde{P}_{0j}(s)z^j, \quad \tilde{P}_1(s,z) = \sum_{j=0}^{\infty}\tilde{P}_{1j}(s)z^j, \tag{7.28}$$

$$\tilde{M}_0(z) = \sum_{j=1}^{\infty}\mu_{0j}, \quad \tilde{M}_1(z) = \sum_{j=0}^{\infty}\mu_{1j}. \tag{7.29}$$

Multiplying (7.25) by $(s-\lambda)z^j$, and summing for $j \geq 1$ yields

$$(\lambda - s)\tilde{P}_0(s,z) = (\tilde{M}_1(z) - \mu_{10})\tilde{A}(s) - \tilde{M}_0(z). \tag{7.30}$$

Similarly, multiplying (7.27) by $(s-\lambda)z^j$, summing for $j \geq 1$, and summing also (7.26) yields

$$(\lambda - s)\tilde{P}_1(s,z) = \sum_{j=0}^{\infty}(\mu_{0,j+1} + \lambda p_{0j})z^j\tilde{B}(s) + \lambda z\tilde{P}_1(s,z) - \tilde{M}_1(z). \tag{7.31}$$

Note, however, that

$$\sum_{j=0}^{\infty}\mu_{0,j+1}z^j = \frac{\tilde{M}_0(z)}{z},$$

$$\sum_{j=0}^{\infty}\lambda p_{0j}z^j = \lambda p_{00} + \lambda\sum_{j=1}^{\infty}\tilde{P}_{0j}(0)z^j = \lambda p_{00} + \lambda\tilde{P}_0(0,z),$$

so (7.31) assumes the form

$$(\lambda - s)\tilde{P}_1(s,z) = \frac{1}{z}\tilde{M}_0(z)\tilde{B}(s) + \lambda p_{00}\tilde{B}(s) + \lambda\tilde{P}_0(0,z)\tilde{B}(s) + \lambda z\tilde{P}_1(s,z) - \tilde{M}_1(z). \tag{7.32}$$

Setting $s = 0$ in (7.30), $s = 0$ in (7.32), $s = \lambda$ in (7.30), and $s = \lambda(1-z)$ in (7.32) yields the system

$$\lambda\tilde{P}_0(0,z) = \tilde{M}_1(z) - \mu_{10} - \tilde{M}_0(z), \tag{7.33}$$

$$\lambda(1-z)\tilde{P}_1(0,z) = \frac{1}{z}\tilde{M}_0(z) + \lambda p_{00} + \lambda\tilde{P}_0(0,z) - \tilde{M}_1(z), \tag{7.34}$$

$$\tilde{M}_1(z)\tilde{A}(\lambda) = \mu_{10}\tilde{A}(\lambda) + \tilde{M}_0(z), \tag{7.35}$$

$$\tilde{M}_1(z) = \frac{1}{z}\tilde{M}_0(z)\tilde{B}(\lambda(1-z)) + \lambda p_{00}\tilde{B}(\lambda(1-z)) + \lambda\tilde{P}_0(0,z)\tilde{B}(\lambda(1-z)), \tag{7.36}$$

in the unknowns $\tilde{M}_0(z)$, $\tilde{M}_1(z)$, $\tilde{P}_0(0,z)$, and $\tilde{P}_1(0,z)$. We can easily solve the system by using sequentially (7.35), (7.33), and (7.34), to express the unknowns $\tilde{M}_1(z)$, $\tilde{P}_0(0,z)$, and $\tilde{P}_1(0,z)$ in terms of $\tilde{M}_0(z)$. We get:

$$\tilde{M}_1(z) = \lambda p_{00} + \frac{1}{\tilde{A}(\lambda)}\tilde{M}_0(z), \tag{7.37}$$

$$\tilde{P}_0(0,z) = \frac{1-\tilde{A}(\lambda)}{\lambda\tilde{A}(\lambda)}\tilde{M}_0(z), \tag{7.38}$$

$$\tilde{P}_1(0,z) = \frac{1}{\lambda z}\tilde{M}_0(z). \tag{7.39}$$

$$\tag{7.40}$$

We can then plug (7.37)-(7.39) and obtain the equation

$$\lambda p_{00} + \frac{1}{\tilde{A}(\lambda)}\tilde{M}_0(z) = \left(\frac{1}{z}\tilde{M}_0(z) + \lambda p_{00} + \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\tilde{M}_0(z)\right)\tilde{B}(\lambda(1-z)) \tag{7.41}$$

for $\tilde{M}_0(z)$. Solving (7.41) for $\tilde{M}_0(z)$ and substituting in (7.37)-(7.39) yields the following explicit expressions for $\tilde{M}_0(z)$, $\tilde{M}_1(z)$, $\tilde{P}_0(0,z)$, and $\tilde{P}_1(0,z)$:

$$\tilde{M}_0(z) = \frac{\lambda\tilde{A}(\lambda)z(1-\tilde{B}(\lambda(1-z)))}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z)))}p_{00}, \tag{7.42}$$

$$\tilde{M}_1(z) = \frac{\lambda\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z)}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z)))}p_{00}, \tag{7.43}$$

$$\tilde{P}_0(0,z) = \frac{(1-\tilde{A}(\lambda))z(1-\tilde{B}(\lambda(1-z)))}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z)))}p_{00}, \tag{7.44}$$

$$\tilde{P}_1(0,z) = \frac{\tilde{A}(\lambda)(1-\tilde{B}(\lambda(1-z)))}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z)))}p_{00}. \tag{7.45}$$

The generating function of the number of customers in system in steady-state is

$$K(z) = \sum_{j=0}^{\infty} p_{0j}z^j + \sum_{j=1}^{\infty} p_{1,j-1}z^j = p_{00} + \tilde{P}_0(0,z) + z\tilde{P}_1(0,z). \tag{7.46}$$

Using (7.44)-(7.45), we obtain

$$K(z) = \frac{\tilde{A}(\lambda)(1-z)\tilde{B}(\lambda(1-z))}{\tilde{A}(\lambda)\tilde{B}(\lambda(1-z))(1-z) - z(1-\tilde{B}(\lambda(1-z)))}p_{00}. \tag{7.47}$$

The probability of an empty system p_{00} is determined from the normalization equation $K(1) = 1$. Using L'Hospital's rule we obtain

$$p_{00} = 1 - \frac{\lambda(-1)\tilde{B}'(0)}{\tilde{A}(\lambda)} = 1 - \frac{\lambda E[B]}{\tilde{A}(\lambda)}. \tag{7.48}$$

The system is stable if and only if $p_{00} > 0$, which gives the stability condition (5.61). Substituting (7.48) in (7.47) yields (5.62). ∎

In the rest of the Appendix, we use the well-known supplementary variable method, introduced by Cox (1955), to provide analytic proofs for the three QMCD Systems (Corollaries 5.1-5.3) which constitute the basis for the recursive schemes for the performance analysis of the model.

To this end, we introduce the following transient probabilities and densities

$$\begin{aligned} p_{ij}^{(t)} &= \Pr[C(t) = i, Q(t) = j],\ i = 0, 1,\ j \geq 0, & (7.49)\\ p_{0j}^{(t)}(r) &= \lim_{dr \to 0^+} \frac{\Pr[C(t) = 0, Q(t) = j, S(t) \in (r, r + dr]]}{dr},\ j \geq 1, & (7.50)\\ p_{1j}^{(t)}(r) &= \lim_{dr \to 0^+} \frac{\Pr[C(t) = 1, Q(t) = j, S(t) \in (r, r + dr]]}{dr},\ j \geq 0, & (7.51) \end{aligned}$$

and their steady-state counterparts

$$\begin{aligned} p_{ij} &= \lim_{t \to \infty} p_{ij}^{(t)},\ i = 0, 1,\ j \geq 0, & (7.52)\\ p_{0j}(r) &= \lim_{t \to \infty} p_{0j}^{(t)}(r),\ j \geq 1, & (7.53)\\ p_{1j}(r) &= \lim_{t \to \infty} p_{1j}^{(t)}(r),\ j \geq 0. & (7.54) \end{aligned}$$

A moment of reflection reveals that the analytic quantity $p_{ij}(0)$ is identical to the corresponding rate μ_{ij} of service/seeking time completions that we considered in the probabilistic derivations, for all i, j. Therefore, similarly to (5.1) and (5.3), we have in the present analytic framework the equations

$$\begin{aligned} \nu_{1j} &= p_{0,j+1}(0) + \lambda_{0j} p_{0j}, & (7.55)\\ \tilde{B}_j(s) &= \frac{1}{\nu_{1j}} \left(\mu_{0,j+1} \tilde{B}_{0j}(s) + \lambda_{0j} p_{0j} \tilde{B}_{1j}(s) \right). & (7.56) \end{aligned}$$

Moreover, we denote by $\tilde{P}_{ij}(s)$ the Laplace transform (LT) of the steady-state density $p_{ij}(r)$, i.e.,

$$\tilde{P}_{ij}(s) = \int_0^\infty e^{-sr} p_{ij}(r) dr = p_{ij} \tilde{S}_{ij}(s),\ (i, j) \in \{0, 1\} \times \{0, 1, 2, \ldots\} \setminus \{(0, 0)\}. \tag{7.57}$$

We consider the evolution of the continuous time Markov process $\{(C(t), Q(t), S(t)) : t \geq 0\}$ in the interval $[t, t + dt]$. Then, we have Lemma 7.1.

Lemma 7.1 *The steady-state probabilities p_{0j}, $j \geq 1$, the steady-state densities $p_{ij}(0)$ and the LTs $\tilde{P}_{ij}(s)$, $(i, j) \in \{0, 1\} \times \{0, 1, 2, \ldots\} \setminus \{(0, 0)\}$ satisfy the following system of equations.*

$$\begin{aligned} \lambda_{00} p_{00} - p_{10}(0) &= 0 & (7.58)\\ (s - \lambda_{0j}) \tilde{P}_{0j}(s) &= p_{0j}(0) - p_{1j}(0) \tilde{A}_j(s),\ j \geq 1, & (7.59)\\ (s - \lambda_{10}) \tilde{P}_{10}(s) &= p_{10}(0) - p_{01}(0) \tilde{B}_{00}(s) - \lambda_{00} p_{00} \tilde{B}_{10}(s), & (7.60)\\ (s - \lambda_{1j}) \tilde{P}_{1j}(s) &= p_{1j}(0) - p_{0,j+1}(0) \tilde{B}_{0j}(s) - \lambda_{0j} p_{0j} \tilde{B}_{1j}(s) - \lambda_{1,j-1} \tilde{P}_{1,j-1}(s). & (7.61) \end{aligned}$$

Proof. Considering the evolution of $\{(C(t), Q(t), R(t)) : t \geq 0\}$ in the interval $[0, t+dt]$ and conditioning on its value at time t, we have the equations

$$
\begin{aligned}
p_{00}^{(t+dt)} &= p_{00}^{(t)}(1-\lambda_{00}dt) + p_{10}^{(t)}(0)dt + o(dt), && (7.62)\\
p_{0j}^{(t+dt)}(r-dt) &= p_{0j}^{(t)}(r)(1-\lambda_{0j}dt) + p_{1j}^{(t)}(0)a_j(r)dt + o(dt),\ j \geq 1, && (7.63)\\
p_{10}^{(t+dt)}(r-dt) &= p_{10}^{(t)}(r)(1-\lambda_{10}dt) + p_{01}^{(t)}(0)dtb_{00}(r) + p_{00}^{(t)}\lambda_{00}dtb_{10}(r) + o(dt), && (7.64)\\
p_{1j}^{(t+dt)}(r-dt) &= p_{1j}^{(t)}(r)(1-\lambda_{1j}dt) + p_{0,j+1}^{(t)}(0)dtb_{0j}(r) + p_{0j}^{(t)}\lambda_{0j}dtb_{1j}(r)\\
&\quad + p_{1,j-1}^{(t)}(r)\lambda_{1,j-1}dt + o(dt),\ j \geq 1, && (7.65)
\end{aligned}
$$

for $dt \to 0^+$. Taking the limits of (7.62)-(7.65), as $t \to \infty$ and using (7.52)-(7.54) yields

$$
\begin{aligned}
p_{00} &= p_{00}(1-\lambda_{00}dt) + p_{10}(0)dt + o(dt), && (7.66)\\
p_{0j}(r-dt) &= p_{0j}(r)(1-\lambda_{0j}dt) + p_{1j}(0)a_j(r)dt + o(dt),\ j \geq 1, && (7.67)\\
p_{10}(r-dt) &= p_{10}(r)(1-\lambda_{10}dt) + p_{01}(0)dtb_{00}(r) + p_{00}\lambda_{00}dtb_{10}(r) + o(dt), && (7.68)\\
p_{1j}(r-dt) &= p_{1j}(r)(1-\lambda_{1j}dt) + p_{0,j+1}(0)dtb_{0j}(r) + p_{0j}\lambda_{0j}dtb_{1j}(r)\\
&\quad + p_{1,j-1}(r)\lambda_{1,j-1}dt + o(dt),\ j \geq 1, && (7.69)
\end{aligned}
$$

for $dt \to 0^+$. Rearranging the terms appropriately, dividing by dt and taking the limits as $dt \to 0^+$ in(7.66)-(7.69) yields

$$
\begin{aligned}
p_{10}(0) &= \lambda_{00}p_{00}, && (7.70)\\
p'_{0j}(r) &= \lambda_{0j}p_{0j}(r) - p_{1j}(0)a_j(r),\ j \geq 1, && (7.71)\\
p'_{10}(r) &= \lambda_{10}p_{10}(r) - p_{01}(0)b_{00}(r) - \lambda_{00}p_{00}b_{10}(r), && (7.72)\\
p'_{1j}(r) &= \lambda_{1j}p_{1j}(r) - p_{0,j+1}(0)b_{0j}(r) - \lambda_{0j}p_{0j}b_{1j}(r) - \lambda_{1,j-1}p_{1,j-1}(r),\ j \geq 1. && (7.73)
\end{aligned}
$$

Equation (7.70) yields immediately (7.58). Also, multiplying (7.71)-(7.73) by e^{-sr} and integrating with respect to $s \in [0, \infty)$ yields equations (7.59)-(7.61). ■

Now, we can provide an analytic proof of the 2nd and the 3rd QMCD systems.

Lemma 7.2 *The steady-state probabilities p_{ij} and the steady-state densities $p_{ij}(0)$ satisfy the 2nd QMCD system:*

$$
\begin{aligned}
\lambda_{00}p_{00} &= p_{10}(0), && (7.74)\\
\lambda_{0j}p_{0j} &= p_{1j}(0) - p_{0j}(0),\ j \geq 1, && (7.75)\\
\lambda_{1j}p_{1j} &= p_{0,j+1}(0),\ j \geq 0. && (7.76)
\end{aligned}
$$

Together with the LSTs $\tilde{S}_{ij}(s)$, they also satisfy the 3rd QMCD system

$$
\begin{aligned}
p_{0j}(0) &= p_{1j}(0)\tilde{A}_j(\lambda_{0j}),\ j \geq 1, && (7.77)\\
p_{10}(0) &= \lambda_{00}p_{00}\tilde{B}_{10}(\lambda_{10}) + p_{01}(0)\tilde{B}_{00}(\lambda_{10}) = \nu_{10}\tilde{B}_0(\lambda_{10}), && (7.78)\\
p_{1j}(0) &= \lambda_{0j}p_{0j}\tilde{B}_{1j}(\lambda_{1j}) + \lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}) + p_{0,j+1}(0)\tilde{B}_{0j}(\lambda_{1j})\\
&= \nu_{1j}\tilde{B}_j(\lambda_{1j}) + \lambda_{1,j-1}p_{1,j-1}\tilde{S}_{1,j-1}(\lambda_{1j}),\ j \geq 1. && (7.79)
\end{aligned}
$$

Proof. Equation (7.58) yields immediately (7.74). Taking $s = 0$ in (7.59)-(7.61) yields

$$
\begin{aligned}
\lambda_{0j}p_{0j} &= p_{1j}(0) - p_{0j}(0),\ j \geq 1 && (7.80)\\
\lambda_{10}p_{10} &= p_{01}(0) + \lambda_{00}p_{00} - p_{10}(0), && (7.81)\\
\lambda_{1j}p_{1j} &= p_{0,j+1}(0) + \lambda_{0j}p_{0j} + \lambda_{1,j-1}p_{1,j-1} - p_{1j}(0),\ j \geq 1. && (7.82)
\end{aligned}
$$

Equation (7.80) is identical to equation (7.75). Also, using (7.74), equation (7.81) yields (7.76) for $j = 0$.

Using (7.75), equation (7.82) assumes the form

$$\lambda_{1j} p_{1j} = p_{0,j+1}(0) - p_{0j}(0) + \lambda_{1,j-1} p_{1,j-1},\ j \geq 1.$$

Iterating this equation yields

$$\begin{aligned} \lambda_{1j} p_{1j} &= \sum_{i=1}^{j} \left(p_{0,i+1}(0) - p_{0i}(0)\right) + \lambda_{10} p_{10} \\ &= p_{0,j+1}(0) - p_{01}(0) + \lambda_{10} p_{10},\ j \geq 1. \end{aligned}$$

Then, using (7.76) for $j = 0$, we deduce (7.76) for $j \geq 1$.

Equation (7.59) for $s = \lambda_{0j}$ yields (7.77). Also, setting $s = \lambda_{10}$ in equation (7.60) and using (7.56) yields (7.78). Similarly, equation (7.61) for $s = \lambda_{1j}$ in combination with (7.56) and (7.57) yields (7.79). ■

Finally, we provide an analytic proof of the 1st QMCD system, i.e., of Corollary 5.1 with a_j given by (5.14).

Corollary 5.1 - Proof. (Analytic proof of (5.8)-(5.10) with a_j given by (5.14)).

Equation (7.59) can be written as

$$\tilde{P}_{0j}(s) = \frac{1}{s - \lambda_{0j}} \left(p_{0j}(0) - p_{1j}(0)\tilde{A}_j(s)\right),\ j \geq 1.$$

Using (7.77), the above equation yields

$$\tilde{P}_{0j}(s) = \frac{1}{s - \lambda_{0j}} p_{1j}(0) \left(\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)\right),\ j \geq 1.$$

Dividing by p_{0j}, we deduce that

$$\begin{aligned} \tilde{S}_{0j}(s) &= \frac{\lambda_{0j}}{s - \lambda_{0j}} \cdot \frac{p_{1j}(0)}{\lambda_{0j} p_{0j}} \left(\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)\right) \\ &= \frac{\lambda_{0j}}{s - \lambda_{0j}} \cdot \frac{p_{1j}(0)\left(1 - \tilde{A}_j(\lambda_{0j})\right)}{\lambda_{0j} p_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)}{1 - \tilde{A}_j(\lambda_{0j})} \\ &\overset{(7.77)}{=} \frac{\lambda_{0j}}{s - \lambda_{0j}} \cdot \frac{p_{1j}(0) - p_{0j}(0)}{\lambda_{0j} p_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)}{1 - \tilde{A}_j(\lambda_{0j})} \\ &\overset{(7.75)}{=} \frac{\lambda_{0j}}{s - \lambda_{0j}} \cdot \frac{\tilde{A}_j(\lambda_{0j}) - \tilde{A}_j(s)}{1 - \tilde{A}_j(\lambda_{0j})},\ j \geq 1, \end{aligned}$$

that proves formula (5.8).

Equation (7.60) can be written as

$$\tilde{P}_{10}(s) = \frac{1}{s - \lambda_{10}} \left(p_{10}(0) - p_{01}(0)\tilde{B}_{00}(s) - p_{00}\lambda_{00}\tilde{B}_{10}(s)\right).$$

Using (7.78) we obtain

$$\tilde{P}_{10}(s) = \frac{1}{s - \lambda_{10}} \left[p_{01}(0) \left(\tilde{B}_{00}(\lambda_{10}) - \tilde{B}_{00}(s)\right) + p_{00}\lambda_{00} \left(\tilde{B}_{10}(\lambda_{10}) - \tilde{B}_{10}(s)\right)\right]$$

and dividing by p_{10} yields

$$\begin{aligned}
\tilde{S}_{10}(s) &= \frac{\lambda_{10}}{s-\lambda_{10}}\left[\frac{p_{01}(0)}{\lambda_{10}p_{10}}\left(\tilde{B}_{00}(\lambda_{10})-\tilde{B}_{00}(s)\right)+\frac{\lambda_{00}p_{00}}{\lambda_{10}p_{10}}\left(\tilde{B}_{10}(\lambda_{10})-\tilde{B}_{10}(s)\right)\right] \\
&\overset{(7.56)}{=} \frac{\lambda_{10}}{s-\lambda_{10}}\cdot\frac{\nu_{10}}{\lambda_{10}p_{10}}\left(\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)\right) \\
&= \frac{\lambda_{10}}{s-\lambda_{10}}\cdot\frac{\nu_{10}\left(1-\tilde{B}_{0}(\lambda_{10})\right)}{\lambda_{10}p_{10}}\cdot\frac{\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)}{1-\tilde{B}_{0}(\lambda_{10})} \\
&\overset{(7.78)}{=} \frac{\lambda_{10}}{s-\lambda_{10}}\cdot\frac{\nu_{10}-p_{10}(0)}{\lambda_{10}p_{10}}\cdot\frac{\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)}{1-\tilde{B}_{0}(\lambda_{10})} \\
&\overset{(7.74)}{=} \frac{\lambda_{10}}{s-\lambda_{10}}\cdot\frac{\nu_{10}-\lambda_{00}p_{00}}{\lambda_{10}p_{10}}\cdot\frac{\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)}{1-\tilde{B}_{0}(\lambda_{10})} \\
&\overset{(7.55)}{=} \frac{\lambda_{10}}{s-\lambda_{10}}\cdot\frac{p_{01}(0)}{\lambda_{10}p_{10}}\cdot\frac{\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)}{1-\tilde{B}_{0}(\lambda_{10})} \\
&\overset{(7.76)}{=} \frac{\lambda_{10}}{s-\lambda_{10}}\frac{\tilde{B}_{0}(\lambda_{10})-\tilde{B}_{0}(s)}{1-\tilde{B}_{0}(\lambda_{10})},
\end{aligned}$$

that proves formula (5.9).

Similarly, equation (7.61) is written as

$$\tilde{P}_{1j}(s)=\frac{1}{s-\lambda_{1j}}\left(p_{1j}(0)-p_{0,j+1}(0)\tilde{B}_{0j}(s)-p_{0j}\lambda_{0j}\tilde{B}_{1j}(s)-\lambda_{1,j-1}\tilde{P}_{1,j-1}(s)\right),\ j\geq 1.$$

Using (7.79) and (7.57) we obtain

$$\begin{aligned}
\tilde{P}_{1j}(s) &= \frac{1}{s-\lambda_{1j}}\Big[p_{0,j+1}(0)\left(\tilde{B}_{0j}(\lambda_{1j})-\tilde{B}_{0j}(s)\right)+\lambda_{0j}p_{0j}\left(\tilde{B}_{1j}(\lambda_{1j})-\tilde{B}_{1j}(s)\right) \\
&\quad+\lambda_{1,j-1}\left(\tilde{P}_{1,j-1}(\lambda_{1j})-\tilde{P}_{1,j-1}(s)\right)\Big],\ j\geq 1.
\end{aligned}$$

Dividing by p_{1j} yields

$$\begin{aligned}
\tilde{S}_{1j}(s) &= \frac{\lambda_{1j}}{s-\lambda_{1j}}\Big[\frac{p_{0,j+1}(0)}{\lambda_{1j}p_{1j}}\left(\tilde{B}_{0j}(\lambda_{1j})-\tilde{B}_{0j}(s)\right)+\frac{\lambda_{0j}p_{0j}}{\lambda_{1j}p_{1j}}\left(\tilde{B}_{1j}(\lambda_{1j})-\tilde{B}_{1j}(s)\right) \\
&\quad+\frac{\lambda_{1,j-1}p_{1,j-1}}{\lambda_{1j}p_{1j}}\left(\tilde{S}_{1,j-1}(\lambda_{1j})-\tilde{S}_{1,j-1}(s)\right)\Big] \\
&\overset{(7.56)}{=} \frac{\lambda_{1j}}{s-\lambda_{1j}}\left[\frac{\nu_{1j}}{\lambda_{1j}p_{1j}}\left(\tilde{B}_{j}(\lambda_{1j})-\tilde{B}_{j}(s)\right)+\frac{\lambda_{1,j-1}p_{1,j-1}}{\lambda_{1j}p_{1j}}\left(\tilde{S}_{1,j-1}(\lambda_{1j})-\tilde{S}_{1,j-1}(s)\right)\right] \\
&= \frac{\lambda_{1j}}{s-\lambda_{1j}}\Bigg[\frac{\nu_{1j}\left(1-\tilde{B}_{j}(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}}\cdot\frac{\tilde{B}_{j}(\lambda_{1j})-\tilde{B}_{j}(s)}{1-\tilde{B}_{j}(\lambda_{1j})} \\
&\quad+\frac{\lambda_{1,j-1}p_{1,j-1}\left(1-\tilde{S}_{1,j-1}(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}}\cdot\frac{\tilde{S}_{1,j-1}(\lambda_{1j})-\tilde{S}_{1,j-1}(s)}{1-\tilde{S}_{1,j-1}(\lambda_{1j})}\Bigg] \\
&= \frac{\lambda_{1j}}{s-\lambda_{1j}}\left[\alpha_j\frac{\tilde{B}_{j}(\lambda_{1j})-\tilde{B}_{j}(s)}{1-\tilde{B}_{j}(\lambda_{1j})}+(1-\alpha_j)\frac{\tilde{S}_{1,j-1}(\lambda_{1j})-\tilde{S}_{1,j-1}(s)}{1-\tilde{S}_{1,j-1}(\lambda_{1j})}\right],\ j\geq 1,
\end{aligned}$$

where

$$\alpha_j = \frac{\nu_{1j}\left(1-\tilde{B}_j(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}},\ j \geq 1 \tag{7.83}$$

and

$$\begin{aligned}
&\frac{\lambda_{1,j-1}p_{1,j-1}\left(1-\tilde{S}_{1,j-1}(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}} \\
\overset{(7.79)}{=}\ &\frac{\lambda_{1,j-1}p_{1,j-1} - p_{1j}(0) + p_j^1\tilde{B}_j(\lambda_{1j})}{\lambda_{1j}p_{1j}} \\
\overset{(7.75)}{=}\ &\frac{\lambda_{1,j-1}p_{1,j-1} - \lambda_{0j}p_{0j} - p_{0j}(0) + p_j^1\tilde{B}_j(\lambda_{1j})}{\lambda_{1j}p_{1j}} \\
\overset{(7.76)}{=}\ &\frac{-\lambda_{0j}p_{0j} + p_j^1\tilde{B}_j(\lambda_{1j})}{\lambda_{1j}p_{1j}} \\
\overset{(7.55)}{=}\ &\frac{p_{0,j+1}(0) - p_j^1 + p_j^1\tilde{B}_j(\lambda_{1j})}{\lambda_{1j}p_{1j}} \\
\overset{(7.76)}{=}\ &\frac{\lambda_{1j}p_{1j} - p_j^1\left(1-\tilde{B}_j(\lambda_{1j})\right)}{\lambda_{1j}p_{1j}} \\
\overset{(7.83)}{=}\ &1-\alpha_j,\ j \geq 1.
\end{aligned}$$

Thus, formula (5.10) has been proved with a_j given by (5.14). ■