Ann Oper Res
https://doi.org/10.1007/s10479-018-2924-x

ORIGINAL RESEARCH

Exact analysis for multiserver queueing systems with cross selling

Mor Armony[1] · Efrat Perel[2] · Nir Perel[3] · Uri Yechiali[4]

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Exact analysis of a multi-server Markovian queueing system with cross selling in steady-state is presented. Cross selling attempt is initiated at the end of a customer's service every time the number of customers in the system is below a threshold. Both probability generating functions (PGFs) and matrix geometric methods are employed. The relation between the methods is revealed by explicitly calculating the entries of the matrix geometric rate-matrix R. Those entries are expressed in terms of the roots of a determinant of a matrix related to the set of linear equations involving the PGFs. This is a further step towards understanding of the analytical relationship between the two methods. Numerical results are presented, showing the effect of the cross selling intensity and of the threshold level on the systems performance measures. Finally, for a given set of parameters, the optimal threshold level is determined.

Keywords Cross-selling · Probability generating functions · Matrix geometric

Mathematics Subject Classification 60K25 · 90B22

✉ Nir Perel
perelnir@gmail.com

Mor Armony
marmony@stern.nyu.edu

Efrat Perel
efratp@afeka.ac.il

Uri Yechiali
uriy@post.tau.ac.il

1 Stern School of Business, New York University, New York, USA

2 School of Industrial Engineering and Management, Afeka College of Engineering, Tel-Aviv, Israel

3 School of Industrial Engineering and Management, Shenkar - Engineering, Design and Art, Ramat Gan, Israel

4 Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel

Published online: 08 June 2018

1 Introduction

Imagine standing in line for a cashier at a supermarket with a few people in front of you. After waiting for a while, as you get closer to the cashier, you notice that she offers the person in front of you a long list of special deals, all of this while you are waiting your turn. You understand that the supermarket would like to maximize revenue by offering these special deals, but you also realize that without these promotion attempts, service times would be shorter and the line would move faster. Considering this tradeoff, the supermarket might change its policy to only offer special deals when the lines are short.

This policy that extends customer service time to increase the value of the service to the firm whenever the number in the system is less than a threshold has gotten some attention in the literature. Specifically, in the context of cross selling, being applied extensively in call centers, see Aksin and Harker (1999), Armony and Gurvich (2010), Byers and So (2007), Gurvich et al. (2009) and Lerzan Örmeci and Zeynep Akşin (2010). Similar policies are used in systems with higher and lower priority customers, in which lower priority customers are served only when the number in system is less than a threshold [see Armony and Maglaras (2004a, b), in the context of call-back option, or Gans and Zhou (2003), Bhulai and Koole (2003), Pang and Perry (2014) in the context of blended call centers, and Gurvich et al. (2008), Perry and Whitt (2009) in multiclass call centers]. A similar policy has also been proposed to handle the tradeoff between accurate diagnosis and timely service in healthcare, see Wang et al. (2010), Mills et al. (2013) and Alizamir et al. (2013). Recent literature on dynamic control in queueing systems with speed-quality tradeoff includes, see e.g. Anand et al. (2011), Kostami and Rajagopalan (2013), Zhan and Ward (2015) and Xu et al. (2015).

In the context of cross-selling it has been shown in Armony and Gurvich (2010) that in order to maximize expected steady-state revenue from cross-selling (e.g. offering special deals in the super market example) while keeping expected steady-state waiting time low, it is asymptotically optimal to initiate cross-selling activity only if the number of customers in the system is less than a threshold. The asymptotic optimality is in a many-server heavy-traffic regime and the asymptotically optimal threshold is calculated via limiting approximated performance expressions. Exact expressions for the steady-state distribution and performance measures have been unavailable so far. In the current paper, we set out to perform exact analysis of such a threshold policy in a queueing system with a fixed number of servers. Assuming the system is Markovian, we compute the steady-state distribution of the two-dimensional cross-selling system using the methods of (i) generating functions, see Litvak and Yechiali (2003), Perel and Yechiali (2008, 2013a, b, 2014), and of (ii) matrix geometric, see e.g. Neuts (1981), Kim et al. (2008). We calculate the system's performance measures and use the results to find the optimal cross-selling threshold.

Our contributions are both in methodology and application. From a methodological point of view, while we solve an open problem using existing techniques (i.e. probability generating functions and matrix geometric analysis), the specifics of this problem required inventivness in how these methods should be used. These two methods are commonly used to find the steady-state probabilities of two-dimensional Markov chains. Typically, one or the other method is used. However, the exact connection between the two methods is not yet well understood. Therefore, our revealed connection between the two methods is important from a methodological point of view. It entails the explicit calculation of the entries of the rate matrix R, associated with the matrix geometric method in terms of the roots of a determinant of a matrix $A(z)$, which is the cornerstone of the PGFs method. In most cases, the rate matrix R is calculated numerically via successive substitutions as a solution of a quadratic matrix equation, $A_0 + RA_1 + R^2A_2 = 0$, where A_0, A_1 and A_2 are known matrices composed of the model parameters, see Neuts (1981) and Latouche and Ramaswami (1999). Some advances in this direction have been made recently by Paz and Yechiali (2014), Perel and Yechiali (2013b, 2017) and Hanukov et al. (2017), in which a special structure of the matrices A_0, A_1 and A_2 enables an explicit derivation of the matrix R. In the current paper, when describing the cross-selling model as a quasi-birth and death process (see Sect. 3), the matrices A_0 and A_1 are diagonal, while A_2 consists only of the main diagonal and the one below it, resulting in a lower triangular matrix R, which is calculated explicitly. However, in spite of all the advances mentioned above, the full analytic connection between the two methods is not known yet. Our results serve as an additional step towards a complete understanding of the inherent relationship between the two methods.

For cross-selling applications, our findings are prescriptive in nature and also provide managerial insights in terms of identifying the main drivers of the dynamic cross-selling decision. We further show that the stability condition for the cross selling system is the same as for an identical system with no cross selling at all.

The paper continues as follows: The model is described and formulated in Sect. 2, where steady-state equations are derived, probability generating functions are calculated, and mean queue sized are obtained. In Sect. 3 the matrix geometric method is employed, and the rate matrix R is explicitly derived, while the relationship with the roots of the basic matrix of the PGFs is established. Numerical results are presented in Sect. 4, as well as few examples where the optimal value of the cross selling threshold T, and the optimal arrival rate, are determined. Section 5 concludes the paper.

2 The model

2.1 Model description

Consider a multiserver queueing system with N parallel servers and unlimited waiting room, to which customers arrive according to a Poisson process with rate λ. A customer service has two potential phases. Phase 1 is experienced by every customer with exponential duration having mean $1/\mu$. After a completion of Phase 1 service a customer is identified as a cross-selling candidate with probability p, or the customer completes service and leaves the system with the complementary probability $q = 1-p$. If the customer is a cross-selling candidate and the system manager decides to go ahead and discuss a cross-selling deal with the customer, Phase 2 of the service begins, having exponential duration with mean $1/\xi$. Let $L_1(t)$ denote the number of customers in Phase 1 service at time t, and let $L_2(t)$ be the number of customers in Phase 2 of the service (cross-selling). Let $L_q(t)$ denote the queue length, and $L(t) = L_q(t) + L_1(t) + L_2(t)$ is the total number of customers in the system. Assuming non-idling policies in which a new arrival is admitted to service immediately upon arrival whenever there is an available server, then (suppressing t) it is sufficient to describe the system with a two-dimensional state-space (L_2, L). To see this, note that the number of customers in phase 1 service satisfies $L_1 = \min\{L, N\} - L_2$, and the queue length satisfies $L_q = [L - N]^+$.

We consider the following cross selling policy: Given a threshold $T \geq 1$, upon service completion of phase 1, initiate cross selling (phase 2 service) to a cross-selling candidate if and only if the total number of customers in the system is less than or equal to the threshold, i.e., iff $L \leq T$. Notice that the value of T may be greater than the number of servers N, in which case cross-selling will only be initiated when $L_q \leq T - N$. Similarly, the value of

T might be less than or equal to N, in which case cross-selling will be initiated whenever the number of idle servers exceeds $N - T$. This model is precisely the model considered in Armony and Gurvich (2010), where examples were given in which in (asymptotic) optimality the threshold value is greater or less than the number of servers, respectively.
Under a threshold policy the process $(L_2(t), L(t))$ is a continuous time Markov chain whose states and transition rates for the case that $T > N$ are depicted in Fig. 1. Note that the analysis of the case where $T < N$ is quite similar and is briefly given in the Appendix. The analysis for $T = N$ is omitted from the presentation. Nevertheless, it is taken into consideration when seeking for an optimal threshold level in Sect. 4.

2.2 Steady-state analysis for the case $T > N$

In this section we derive the steady-state distribution of the two-dimensional process defining the states of the system. Again, at time t, let $L_2(t)$ denote the number of customers in cross-selling, and $L(t)$ denote the total number of customers in the system. Let $L_2 = \lim_{t\to\infty} L_2(t)$, $L = \lim_{t\to\infty} L(t)$, and set $P_{jm} = \mathbb{P}(L_2 = j, L = m)$ denote the steady state distribution of the system's state. We assume that $\lambda < N\mu$, which we will show in Sect. 3 that it is a necessary and sufficient condition for the system's stability.

2.2.1 Balance equations and generating functions

The balance equations for the steady-state probabilities are the following: For $j = 0$,

$$\lambda P_{00} = q\mu P_{01} + \xi P_{11}, \tag{1}$$

$$(\lambda + m\mu)\, P_{0m} = \lambda P_{0,m-1} + (m+1)q\mu P_{0,m+1} + \xi P_{1,m+1}, \quad 1 \le m \le N-1, \tag{2}$$

$$(\lambda + N\mu)\, P_{0m} = \lambda P_{0,m-1} + Nq\mu P_{0,m+1} + \xi P_{1,m+1}, \quad N \le m \le T-1, \tag{3}$$

$$(\lambda + N\mu)\, P_{0m} = \lambda P_{0,m-1} + N\mu P_{0,m+1} + \xi P_{1,m+1}, \quad m \ge T. \tag{4}$$

For $1 \le j \le N-1$,

$\underline{m = j}$:

$$(\lambda + j\xi)\, P_{jm} = p\mu P_{j-1,m} + q\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1}, \tag{5}$$

$\underline{j+1 \le m \le N-1}$:

$$\begin{aligned}(\lambda + j\xi + (m-j)\mu)\, P_{jm} = {} & \lambda P_{j,m-1} + (m-j+1)p\mu P_{j-1,m} + (m+1-j)q\mu P_{j,m+1} \\ & + (j+1)\xi P_{j+1,m+1},\end{aligned} \tag{6}$$

$\underline{N \le m \le T-1}$:

$$\begin{aligned}(\lambda + j\xi + (N-j)\mu)\, P_{jm} = {} & \lambda P_{j,m-1} + (N-j+1)p\mu P_{j-1,m} + (N-j)q\mu P_{j,m+1} \\ & + (j+1)\xi P_{j+1,m+1},\end{aligned} \tag{7}$$

$\underline{m = T}$:

$$\begin{aligned}(\lambda + j\xi + (N-j)\mu)\, P_{jm} = {} & \lambda P_{j,m-1} + Np\mu P_{j-1,m} + (N-j)\mu P_{j,m+1} \\ & + (j+1)\xi P_{j+1,m+1},\end{aligned} \tag{8}$$

$\underline{m > T}$:

$$(\lambda + j\xi + (N-j)\mu)\, P_{jm} = \lambda P_{j,m-1} + (N-j)\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1}. \tag{9}$$

Finally, for $j = N$,

$$(\lambda + N\xi)\, P_{Nm} = p\mu P_{N-1,m}, \quad m = N, \tag{10}$$

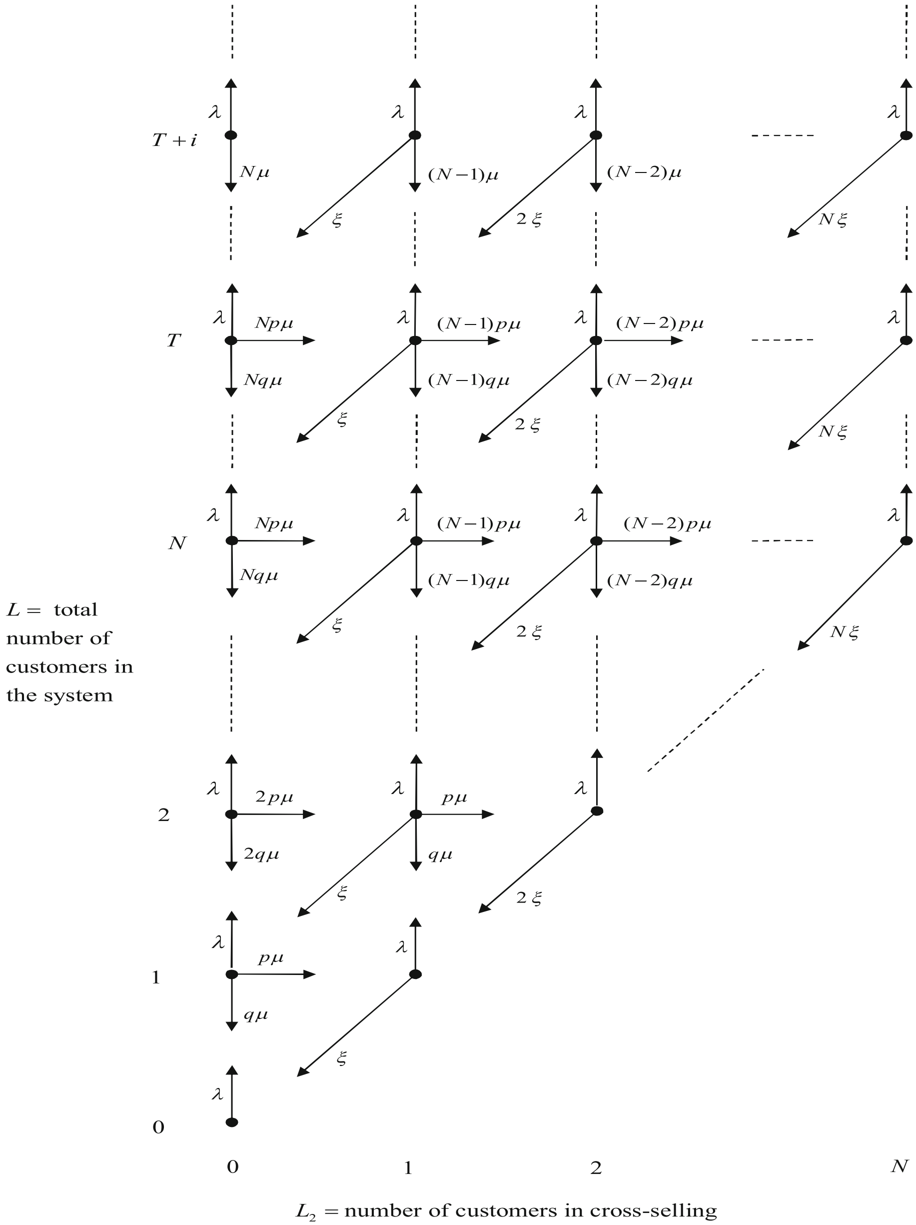

Fig. 1 States and transition-rate diagram of (L_2, L)

$$(\lambda + N\xi)\, P_{Nm} = \lambda P_{N,m-1} + p\mu P_{N-1,m}, \quad N+1 \le m \le T, \tag{11}$$

$$(\lambda + N\xi)\, P_{Nm} = \lambda P_{N,m-1}, \quad m > T. \tag{12}$$

Now, for each $j = 0, 1, 2, \ldots, N$, define the conditional (conditioned on the number of customers in cross selling) probability generating function of the number of customers in the system:

$$G_j(z) = \sum_{m=j}^{\infty} P_{jm} z^m.$$

Then, multiplying each equation for m in Eqs. (2)–(4) by z^m and summing over m together with Eq. 1), we get

$$\begin{aligned}
&((\lambda z - N\mu)(1-z))\, G_0(z) - \xi G_1(z) \\
&\quad = \mu \sum_{m=0}^{N-1} (N-m) P_{0m} z^{m+1} + q\mu \sum_{m=1}^{N} m P_{0m} z^m + Nq\mu \sum_{m=N+1}^{T} P_{0m} z^m - N\mu \sum_{m=0}^{T} P_{0m} z^m.
\end{aligned} \tag{13}$$

Repeating this process for $j = 1, \ldots, N-1$, while using Eqs. (5)–(9), we obtain,

$$\begin{aligned}
&((\lambda z - (N-j)\mu)(1-z)) + j\xi z)\, G_j(z) - (j+1)\xi G_{j+1}(z) \\
&\quad = \mu \sum_{m=j}^{N-1} (N-m) P_{jm} z^{m+1} + p\mu \sum_{m=j}^{N-1} (m-j+1) P_{j-1,m} z^{m+1} \\
&\qquad + (N-j+1) p\mu \sum_{m=N}^{T} P_{j-1,m} z^{m+1} \\
&\qquad + q\mu \sum_{m=j}^{N-1} (m-j+1) P_{j,m+1} z^{m+1} + (N-j) q\mu \sum_{m=N}^{T-1} P_{j,m+1} z^{m+1} \\
&\qquad - (N-j)\mu \sum_{m=j}^{T} P_{jm} z^m.
\end{aligned} \tag{14}$$

Last, from Eqs. (10)–(12) we derive

$$(\lambda(1-z) + N\xi)\, G_N(z) = p\mu \sum_{m=N}^{T} P_{N-1,m} z^m. \tag{15}$$

The set of Eqs. (13)–(15) can be written in a matrix form as

$$A(z) \cdot \mathbf{G}(z) = \mathbf{b}(z), \tag{16}$$

where

$$A(z) = \begin{pmatrix}
\alpha_0(z) & -\xi & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \alpha_1(z) & -2\xi & 0 & \cdots & \cdots & \vdots \\
0 & 0 & \alpha_2(z) & -3\xi & 0 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 0 & \alpha_{N-1}(z) & -N\xi \\
0 & \cdots & \cdots & 0 & 0 & 0 & \alpha_N(z)
\end{pmatrix},$$

$$\alpha_j(z) = (\lambda z - (N-j)\mu)(1-z) + j\xi z, \quad j = 0, 1, \ldots, N-1,$$
$$\alpha_N(z) = \lambda(1-z) + N\xi,$$

and $\mathbf{G}(z) = (G_0(z), G_1(z), \ldots, G_N(z))^T$ is a $(N+1)$-dimensional column vector of the desired PGF's. The elements of the vector $\mathbf{b}(z)$ are

$$b_0(z) = \mu \sum_{m=0}^{N-1} (N-m) P_{0m} z^{m+1} + q\mu \sum_{m=1}^{N} m P_{0m} z^m + Nq\mu \sum_{m=N+1}^{T} P_{0m} z^m$$
$$- N\mu \sum_{m=0}^{T} P_{0m} z^m;$$

For $j = 1, \ldots, N-1$,

$$b_j(z) = \mu \sum_{m=j}^{N-1} (N-m) P_{jm} z^{m+1} + p\mu \sum_{m=j}^{N-1} (m-j+1) P_{j-1,m} z^{m+1}$$
$$+ (N-j+1)p\mu \sum_{m=N}^{T} P_{j-1,m} z^{m+1}$$
$$+ q\mu \sum_{m=j}^{N-1} (m-j+1) P_{j,m+1} z^{m+1} + (N-j)q\mu \sum_{m=N}^{T-1} P_{j,m+1} z^{m+1}$$
$$- (N-j)\mu \sum_{m=j}^{T} P_{jm} z^m,$$

and,

$$b_N(z) = p\mu \sum_{m=N}^{T} P_{N-1,m} z^m.$$

To obtain $G_j(z)$ we use Cramer's rule and write $G_j(z) = \frac{|A_j(z)|}{|A(z)|}$, $j = 0, 1, \ldots, N$, where $|A|$ is the determinant of a matrix A, and $A_j(z)$ is the matrix obtained from $A(z)$ by replacing its jth column by $\mathbf{b}(z)$. The functions $G_j(z)$ are expressed in terms of $\frac{1}{2}(N+1)(2T+2-N)$ unknown 'boundary probabilities', P_{jm}, $j = 0, 1, \ldots, N$ and $m = j, j+1, \ldots, T$ [appearing in $\mathbf{b}(z)$]. In order to derive these boundary probabilities we need to create a set of equations that those probabilities satisfy. First, we consider the following balance equations:

- Equations (1)–(3), which give us T equations.
- Equations (5)–(7), which give us $(T-1) + (T-2) + \cdots + (T-(N-1)) = \frac{1}{2}(N-1)(2T-N)$ equations.
- Equation (10), which gives us 1 equation.
- Equation (11), that gives us $T-N$ equations.

Overall, the above equations provide $\frac{1}{2}(N+1)(2T-N)+1$ equations, and we need another $N = \left[\frac{1}{2}(N+1)(2T+2-N) - \left(\frac{1}{2}(N+1)(2T-N)+1\right)\right]$ equations in order to obtain a unique solution.

Further equations arise from the marginal probabilities $P_{j\bullet} = \sum_{m=j}^{\infty} P_{jm}$, for $j = 0, 1, \ldots, N$, as follows. By applying vertical cuts between every two adjacent columns in Fig. 1 we get

$$p\mu \sum_{m=j+1}^{N} (m-j)P_{jm} + (N-j)p\mu \sum_{m=N+1}^{T} P_{jm}$$
$$= (j+1)\xi P_{j+1,\bullet}, \quad j = 0, 1, \ldots, N-1, \tag{17}$$

which gives us N more equations, but also increases the number of variables by $N+1$. However, we can also use the normalization equation, i.e. $\sum_{j=0}^{N} P_{j\bullet} = 1$. So again, we need N additional equations. To this aim we utilize the roots of $|A(z)|$ as described in the following proposition.

Proposition 1 *For any* $\lambda > 0, \mu > 0, \xi > 0$ *and* $0 < p \leq 1$, $|A(z)|$ *is a polynomial of degree* $2N+1$ *possessing* $N-1$ *distinct roots in the open interval* $(0, 1)$, $N-1$ *distinct roots in the open interval* $(1, \infty)$*, and three more roots at 1,* $\frac{N\mu}{\lambda}$ *and* $1 + \frac{N\xi}{\lambda}$.

Proof Clearly, $|A(z)| = \prod_{j=0}^{N} \alpha_j(z)$. Note that for $j = 1, \ldots, N-1$, the term $\alpha_j(z)$ is a quadratic polynomial with roots $z_{j,1} = \frac{\lambda+(N-j)\mu+j\xi-\sqrt{(\lambda+(N-j)\mu+j\xi)^2-4\lambda(N-j)\mu}}{2\lambda}$ and $z_{j,2} = \frac{\lambda+(N-j)\mu+j\xi+\sqrt{(\lambda+(N-j)\mu+j\xi)^2-4\lambda(N-j)\mu}}{2\lambda}$. Since $\alpha_j(0) = -(n-j)\mu < 0, \alpha_j(1) = j\xi > 0$ and $\alpha_j(\infty) < 0$, we have that $0 < z_{j,1} < 1 < z_{j,2}$. As for $\alpha_0(z)$, its roots are $z_{0,1} = 1$ and $z_{0,2} = \frac{N\mu}{\lambda} > 1$, and the single root of $\alpha_N(z)$ is $z_N = 1 + \frac{N\xi}{\lambda} > 1$. This completes the proof. □

We now utilize the roots of $|A(z)|$. Since $G_j(z)$ is a (partial) probability generating function defined for all $0 \leq z \leq 1$, each root of $|A(z)|$ in that interval is a root of $|A_j(z)|$. Thus, writing $|A_j(z_{j,1})| = 0$ for $j = 1, 2, \ldots, N-1$ provides us with $N-1$ more relations between the desired probabilities. Therefore, only 1 additional equation is needed, and we derive it by an innovative set of diagonal cuts as follows:

For each $m \geq T-N$ we consider cuts along the diagonal, namely, cuts separating states $\{(0, m), (1, m+1), \ldots, (N, m+N)\}$ from states $\{(0, m+1), (1, m+2), \ldots, (N, m+N+1)\}$. If $T \leq 2N-2$, the equations are

$$\lambda \sum_{j=0}^{N} P_{j,m+j} = \mu \sum_{j=0}^{N-m-1} (m+1)P_{j,m+j+1} + \mu \sum_{j=N-m}^{N-1} (N-j)P_{j,m+j+1},$$
$$T-N \leq m \leq N-2, \tag{18}$$

$$\lambda \sum_{j=0}^{N} P_{j,m+j} = \mu \sum_{j=0}^{N-1} (N-j)P_{j,m+j+1}, \quad m \geq N-1. \tag{19}$$

Else, the equations are

$$\lambda \sum_{j=0}^{N} P_{j,m+j} = \mu \sum_{j=0}^{N-1} (N-j)P_{j,m+j+1}, \quad m \geq T-N. \tag{20}$$

Summing Eqs. (18) and (19) over $m \geq T-N$ and changing the order of summation eventually gives

$$\lambda \sum_{j=0}^{N}\left(P_{j\bullet} - \sum_{m=j}^{T-N+j-1} P_{jm}\right) = \mu \sum_{j=0}^{N-1}(N-j)\left(P_{j\bullet} - \sum_{m=j}^{T-N+j} P_{jm}\right)$$
$$- \mu \sum_{m=T-N}^{N-2} \sum_{j=0}^{N-m-2} (N-j-m-1) P_{j,m+j+1}, \quad (21)$$

where $\sum_a^b (\bullet) \triangleq 0$ if $b < a$.

Together with Eq. (21) we obtain the required set of linear equations by which the boundary probabilities are calculated.

Direct sequential calculation of the marginal probabilities Typically, the marginal probability of having j customers in cross-selling, i.e. $P_{j\bullet}$, is obtained by setting $G_j(1)$, where $G_j(z) = \frac{|A_j(z)|}{|A(z)|}$. However, in our case, the set $\{P_{j\bullet}\}$ can be calculated directly in a sequential manner as follows. For $j = 0$, substituting $z = 1$ in Eq. (13) yields

$$P_{1\bullet} = -\frac{b_0(1)}{\xi}.$$

Similarly, for $j = 1$ we get from (14)

$$\xi P_{1\bullet} - 2\xi P_{2\bullet} = b_1(1),$$

or equivalently,

$$P_{2\bullet} = -\frac{1}{2\xi}\left(b_0(1) + b_1(1)\right).$$

In the same manner, we get for all $1 \le j \le N$,

$$P_{j\bullet} = -\frac{1}{j\xi}\sum_{i=0}^{j-1} b_i(1), \quad (22)$$

and $P_{0\bullet} = 1 - \sum_{j=1}^{N} P_{j\bullet}$. Note that, by summing Eqs. (13)–(15) over all j and substituting $z = 1$, the LHS vanishes, implying that the RHS equals zero. That is $\sum_{i=0}^{N} b_i(1) = 0$. Therefore, Eq. (22) can be rewritten as

$$P_{j\bullet} = \frac{1}{j\xi}\sum_{i=j}^{N} b_i(1), \quad j = 1, \ldots, N. \quad (23)$$

Calculation of the PGFs As mentioned earlier, the PGFs can be calculated by $G_j(z) = \frac{|A_j(z)|}{|A(z)|}$, for $j = 0, 1, \ldots, N$. However, in our case, after deriving the unknown boundary probabilities, we can obtain the PGFs sequentially, as follows: From the linear system (16) we have that

$$G_N(z) = \frac{b_N(z)}{\alpha_N(z)}, \quad (24)$$

$$G_j(z) = \frac{1}{\alpha_j(z)}\Big(b_j(z) + (j+1)\xi G_{j+1}(z)\Big), \quad j = 0, 1, \ldots, N-1. \quad (25)$$

Iterating Eq. (25) gives, for $j = 0, 1, \ldots, N$,

$$G_{N-j}(z) = \left(\prod_{i=N-j}^{N} \frac{1}{\alpha_i(z)} \right) \left(\sum_{i=0}^{j} b_{N-i}(z) \frac{(N-i)!}{(N-j)!} \xi^{j-i} \prod_{k=N-i+1}^{N} \alpha_k(z) \right),$$

where we define $\prod_{k=N+1}^{N} (\cdot) = 1$.

Mean queue length Define $P_{\bullet m}$ to be the marginal probability of having m customers in the system, i.e. $P_{\bullet m} = \mathbb{P}(L = m)$. We have: $P_{\bullet m} = \sum_{j=0}^{m} P_{jm}$ for $m = 0, 1, \ldots, N-1$; and $P_{\bullet m} = \sum_{j=0}^{N} P_{jm}$ for $m \geq N$.
The mean number of customers in cross-selling is

$$\mathbb{E}[L_2] = \sum_{j=1}^{N} j P_{j\bullet},$$

and the mean proportion of servers busy in cross-selling is $\frac{\mathbb{E}[L_2]}{N}$.

Furthermore, let $\mathbb{E}[L^{(j)}] := \sum_{m=j}^{\infty} m P_{jm}$ for $j = 0, 1, \ldots, N$. Clearly,

$$\mathbb{E}[L^{(j)}] = \frac{d}{dz} G_j(z)|_{z=1},$$

and the mean total number of customers in the system is, therefore,

$$\mathbb{E}[L] = \sum_{j=0}^{N} \mathbb{E}[L^{(j)}] = \sum_{j=0}^{N} \frac{d}{dz} G_j(z)|_{z=1} = \sum_{m=0}^{\infty} m P_{\bullet m}.$$

Note that the $\mathbb{E}[L^{(j)}]$ may be derived recursively, by differentiation of Eqs. (24) and (25) and using the definitions of $\alpha_j(z)$ and $b_j(z)$. Specifically, by differentiating (24) at $z = 1$ we get

$$\mathbb{E}[L^{(N)}] = \frac{1}{(N\xi)^2} \left(N\xi p\mu \sum_{m=N}^{T} m P_{N-1,m} + \lambda p\mu \sum_{m=N}^{T} P_{N-1,m} \right).$$

Differentiating (25) at $z = 1$ gives, for $j = 0, 1, \ldots, N-1$,

$$\mathbb{E}[L^{(j)}] = \frac{1}{(j\xi)^2} \Big(j\xi \left(b_j'(1) + (j+1)\xi \mathbb{E}[L^{(j+1)}] \right) - \big((N-j)\mu - \lambda + j\xi \big) \big(b_j(1) \\ + s(j+1)\xi P_{j+1,\bullet} \big) \Big).$$

Let L_q denote the number of waiting customers. Then,

$$\begin{aligned} \mathbb{E}[L_q] &= \sum_{m=N+1}^{\infty} (m-N) P_{\bullet m} = \sum_{m=N+1}^{\infty} m P_{\bullet m} - \sum_{m=N+1}^{\infty} N P_{\bullet m} \\ &= \mathbb{E}[L] - \sum_{m=0}^{N} m P_{\bullet m} - N \left(1 - \sum_{m=0}^{N} P_{\bullet m} \right) \\ &= \mathbb{E}[L] - N + \sum_{m=0}^{N} (N-m) P_{\bullet m}. \end{aligned} \tag{26}$$

3 Matrix geometric method

An alternative approach to analyze this model is by constructing a Quasi Birth and Death (QBD) process, with $N+1$ phases, where phase j corresponds to $L_2 = j$ ($j = 0, 1, \ldots, N$), and with an infinite number of levels, where each level corresponds to L, the total number of customers in the system. For $m \geq 0$ we define S_m to be the set of states $\{(j, m) : j = 0, 1, \ldots, nim\{m, N\}\}$, and we arrange the system's states under the order S_0, S_1,..., S_N, S_{N+1},..., S_{N+i},..., for $i \geq 0$. The infinitesimal generator of the QBD is denoted by Q, and is given by

$$Q = \begin{pmatrix}
B_1^{(0)} & B_0^{(0)} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
B_2^{(1)} & B_1^{(1)} & B_0^{(1)} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\mathbf{0} & B_2^{(2)} & B_1^{(2)} & B_0^{(2)} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & B_2^{(N)} & B_1^{(N)} & B_0^{(N)} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & B_2^{(N+1)} & B_1^{(N+1)} & B_0^{(N+1)} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & B_2^{(T)} & B_1^{(T)} & B_0^{(T)} & \mathbf{0} & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & A_2 & A_1 & A_0 & \mathbf{0} & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & A_2 & A_1 & A_0 & \mathbf{0} & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots
\end{pmatrix},$$

where, $B_0^{(n)} = [\lambda\mathbf{I}, \mathbf{0}]$ is an $(n+1) \times (n+2)$ matrix for $n = 0, \ldots, N-1$, $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is a vector of zeros. $B_0^{(N)} = B_0^{(N+1)} = \cdots = B_0^{(T)} = \lambda\mathbf{I}$ is a square matrix of order $(N+1) \times (N+1)$, $B_1^{(n)}$ is an $(n+1) \times (n+1)$ matrix for $n = 0, \ldots, N$ and is given by

$$B_1^{(n)} = \begin{pmatrix}
-(\lambda + n\mu) & np\mu & 0 & \cdots & \cdots & 0 \\
0 & -(\lambda + (n-1)\mu + \xi) & (n-1)p\mu & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & -(\lambda + \mu + (n-1)\xi) & p\mu \\
0 & \cdots & \cdots & \cdots & 0 & -(\lambda + n\xi)
\end{pmatrix},$$

and $B_1^{(N)} = B_1^{(N+1)} = \cdots = B_1^{(T)}$.

$B_2^{(n)}$ is an $(n+1) \times n$ matrix for $n = 1, \ldots, N-1$ and is given by

$$B_2^{(n)} = \begin{pmatrix}
nq\mu & 0 & 0 & \cdots\cdots & 0 \\
\xi & (n-1)q\mu & 0 & \cdots\cdots & 0 \\
0 & 2\xi & (n-2)q\mu & 0 \ \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots \ \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots \ \ddots & q\mu \\
0 & \cdots & \cdots & \cdots \ 0 & n\xi
\end{pmatrix},$$

$B_2^{(N)} = B_2^{(N+1)} = \cdots = B_2^{(T)}$ is an $(N+1) \times (N+1)$ square matrix and is given by

$$B_2^{(N)} = \begin{pmatrix} Nq\mu & 0 & \cdots & \cdots\cdots & 0 \\ \xi & (N-1)q\mu & 0 & \cdots\cdots & 0 \\ 0 & 2\xi & (N-2)q\mu & 0 \quad \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots \quad \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \quad q\mu & 0 \\ 0 & \cdots & \cdots & \cdots \quad N\xi & 0 \end{pmatrix},$$

A_2 is an $(N+1) \times (N+1)$ square matrix given by

$$A_2 = \begin{pmatrix} N\mu & 0 & \cdots & \cdots\cdots & 0 \\ \xi & (N-1)\mu & 0 & \cdots\cdots & 0 \\ 0 & 2\xi & (N-2)\mu & 0 \quad \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots \quad \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \quad \mu & 0 \\ 0 & \cdots & \cdots & \cdots \quad N\xi & 0 \end{pmatrix},$$

$A_0 = B_0^{(N)} = \lambda\mathbf{I}$, and A_1 is an $(N+1) \times (N+1)$ square matrix given by

$$A_1 = \begin{pmatrix} -(\lambda+N\mu) & 0 & \cdots\cdots & \cdots & 0 \\ 0 & -(\lambda+(N-1)\mu+\xi) & 0 \quad \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots \quad \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots \quad 0 & -(\lambda+\mu+(N-1)\xi) & 0 \\ 0 & \cdots & \cdots\cdots & 0 & -(\lambda+N\xi) \end{pmatrix},$$

Define the matrix $A = A_0 + A_1 + A_2$. We get

$$A = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \xi & -\xi & 0 & \cdots & \cdots & 0 \\ 0 & 2\xi & -2\xi & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & N\xi & -N\xi \end{pmatrix}.$$

This matrix is the infinitesimal generator of the death process of the customers in cross-selling, when there are at least $T+1$ customers in the system (and therefore no new customers are referred to cross-selling). Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_N)$ be the stationary vector of the matrix A, i.e. $\boldsymbol{\pi} A = \mathbf{0}$ and $\boldsymbol{\pi} \cdot \mathbf{e} = 1$ (where $\mathbf{e}$ is a column vector with all entries equal to 1). Then, an immediate result is that $\boldsymbol{\pi} = \Bigg(1, \underbrace{0, 0, \ldots, 0}_{N \text{ times}}\Bigg)$. The stability condition is (see Neuts (1981))

$$\boldsymbol{\pi} A_0 \mathbf{e} < \boldsymbol{\pi} A_2 \mathbf{e},$$

which translates here into $\lambda < N\mu$ (as indicated at the beginning of Sect. 2.2).
Define for all $m \geq 0$ the steady-state probability vector $\mathbf{P}_m$, as follows:

$$\mathbf{P}_m = \begin{cases} (P_{0m}, P_{1m}, \ldots, P_{mm}), & 0 \leq m \leq N, \\ (P_{0m}, P_{1m}, \ldots, P_{Nm}), & m > N, \end{cases} \tag{27}$$

Then,

$$\mathbf{P}_m = \mathbf{P}_T R^{m-T}, \quad m \geq T,$$

where R is the minimal non-negative solution of the matrix quadratic equation

$$A_0 + RA_1 + R^2 A_2 = 0. \tag{28}$$

The vectors $\mathbf{P}_0, \mathbf{P}_1,\ldots, \mathbf{P}_T$, can be found by solving the following linear system of equations:

$$\begin{aligned}
&\mathbf{P}_0 B_1^{(0)} + \mathbf{P}_1 B_2^{(1)} = \mathbf{0},\\
&\mathbf{P}_0 B_0^{(0)} + \mathbf{P}_1 B_1^{(1)} + \mathbf{P}_2 B_2^{(2)} = \mathbf{0},\\
&\mathbf{P}_1 B_0^{(1)} + \mathbf{P}_2 B_1^{(2)} + \mathbf{P}_3 B_2^{(3)} = \mathbf{0},\\
&\vdots\\
&\mathbf{P}_{N-2} B_0^{(N-2)} + \mathbf{P}_{N-1} B_1^{(N-1)} + \mathbf{P}_N B_2^{(N)} = \mathbf{0},\\
&\mathbf{P}_{N-1} B_0^{(N-1)} + \mathbf{P}_N B_1^{(N)} + \mathbf{P}_{N+1} B_2^{(N)} = \mathbf{0},\\
&\mathbf{P}_N B_0^{(N)} + \mathbf{P}_{N+1} B_1^{(N)} + \mathbf{P}_{N+2} B_2^{(N)} = \mathbf{0},\\
&\vdots\\
&\mathbf{P}_{T-2} B_0^{(N)} + \mathbf{P}_{T-1} B_1^{(N)} + \mathbf{P}_T B_2^{(N)} = \mathbf{0},\\
&\sum_{m=0}^{N-1} \mathbf{P}_m \mathbf{e}_m + \sum_{m=N}^{T-1} \mathbf{P}_m \mathbf{e}_N + \mathbf{P}_T [\mathbf{I} - R]^{-1} \mathbf{e}_N = 1,
\end{aligned}$$

where $\mathbf{e}_m$ is a vector of 1's of dimension $m+1$. The mean total number of customers in the system $\mathbb{E}[L]$ is given by

$$\begin{aligned}
\mathbb{E}[L] &= \sum_{m=0}^{\infty} m\mathbf{P}_m \mathbf{e}_m = \sum_{m=0}^{N-1} m\mathbf{P}_m \mathbf{e}_m + \sum_{m=N}^{T-1} m\mathbf{P}_m \mathbf{e}_N + \sum_{m=T}^{\infty} m\mathbf{P}_m \mathbf{e}_N\\
&= \sum_{m=0}^{N-1} m\mathbf{P}_m \mathbf{e}_m + \sum_{m=N}^{T-1} m\mathbf{P}_m \mathbf{e}_N + \sum_{m=T}^{\infty} m\mathbf{P}_T R^{m-T} \mathbf{e}_N\\
&= \sum_{m=0}^{N-1} m\mathbf{P}_m \mathbf{e}_m + \sum_{m=N}^{T-1} m\mathbf{P}_m \mathbf{e}_N + T\mathbf{P}_T (\mathbf{I} - R)^{-1} \mathbf{e}_N + \mathbf{P}_T R(\mathbf{I} - R)^{-2} \mathbf{e}_N.
\end{aligned} \tag{29}$$

Calculation of the rate matrix R We denote the elements of R as r_{ij}, for $i, j = 0, 1, \ldots, N$. Since A_0 and A_1 are diagonal matrices, and A_2 is comprised of the main diagonal and the one below it (all other elements are 0), it follows that R is a lower triangular matrix. Algorithms for the computation of the matrix R in Eq. (28) are suggested in various works, see e.g. Artalejo and Gómez-Corral (2008), Latouche and Ramaswami (1999) and Neuts (1981). However, in our model, the elements of R can be derived explicitly, as follows: We first calculate the main diagonal of R. We then derive the elements on the diagonal below it, and so on. It turns out, by the structure of R and from Eq. (28), that each entry of R depends only on entries that are on diagonals above it. As a result, we can calculate sequentially the diagonals of R, starting from the main diagonal and ending at the bottom-left entry. By using this procedure we bypass the need to numerically solve the system of non-linear equations given in (28) and convergence issues become irrelevant.

Before we show the derivation of R, recall that in Sect. 2.2.1 we introduced the matrix $A(z)$ and the roots of its determinant. Specifically, we recall the roots $z_{j,2} = \frac{\lambda+(N-j)\mu+j\xi+\sqrt{(\lambda+(N-j)\mu+j\xi)^2-4\lambda(N-j)\mu}}{2\lambda}$ for $j = 0, 1, \ldots, N-1$, and $z_N = 1 + \frac{N\xi}{\lambda}$, which are all larger than 1. Then, the elements r_{ij} of matrix R are calculated by the following Theorem.

Proposition 2 *The elements of the matrix R are given by:*

$$r_{ii} = \frac{1}{z_{i,2}}, \quad i = 0, 1, \ldots, N-1, \tag{30}$$

$$r_{NN} = \frac{1}{z_N} = \frac{\lambda}{\lambda + N\xi}, \tag{31}$$

$$r_{i+n,i} = \frac{(N-i)\mu \sum_{k=i+1}^{i+n-1} r_{i+n,k} r_{ki} + (i+1)\xi \sum_{k=i+1}^{i+n} r_{i+n,k} r_{k,i+1}}{\lambda + (N-i)\mu + i\xi - (N-i)\mu(r_{ii} + r_{i+n,i+n})},$$
$$\textit{for } i = 0, 1, \ldots, N-1, \quad n = 1, 2, \ldots, N-i. \tag{32}$$

Otherwise, for $i < j$, $r_{ij} = 0$.

Proof Rewriting Eq. (28) for the elements on the main diagonal for $i = 0, 1, \ldots, N-1$ leads to

$$\lambda - (\lambda + (N-i)\mu + i\xi) r_{ii} + (N-i)\mu r_{ii}^2 = 0, \quad i = 0, 1, \ldots, N-1. \tag{33}$$

Since R is the minimal non-negative solution of (28), from (33) we get

$$r_{ii} = \frac{\lambda+(N-i)\mu + i\xi - \sqrt{(\lambda + (N-i)\mu + i\xi)^2 - 4\lambda(N-i)\mu}}{2(N-i)\mu}, \quad i = 0, 1, \ldots, N-1. \tag{34}$$

Note that from Eq. (34) we have that $r_{ii} = \frac{\lambda}{(N-i)\mu} z_{i,1}$. Since $z_{i,1}$ and $z_{i,2}$ are the roots of the polynomial $\alpha_i(z) = (\lambda z - (N-i)\mu)(1-z) + i\xi z$, we have that $z_{i,1} \cdot z_{i,2} = \frac{(N-i)\mu}{\lambda}$, and therefore

$$r_{ii} = \frac{\lambda}{(N-i)\mu} z_{i,1} = \frac{\lambda}{(N-i)\mu} \frac{(N-i)\mu}{\lambda z_{i,2}} = \frac{1}{z_{i,2}}, \tag{35}$$

which proves Eq. (30). Now, writing Eq. (28) for r_{NN} results in

$$\lambda - (\lambda + N\xi) r_{NN} = 0,$$

that is,

$$r_{NN} = \frac{\lambda}{\lambda + N\xi}, \tag{36}$$

which proves (31). Once the main diagonal is calculated, we can sequently calculate all the diagonals below it, as follows. Writing Eq. (28) for $r_{i+n,i}$, for $i = 0, 1, \ldots, N-1$ and $n = 1, 2, \ldots, N-i$ gives

$$-r_{i+n,i}(\lambda + (N-i)\mu + i\xi) + (N-i)\mu \sum_{k=0}^{N} r_{i+n,k} r_{k,i} + \sum_{k=0}^{N} r_{i+n,k} r_{k,i+1} = 0. \tag{37}$$

Table 1 Numerical results for $N = 3, T = 5, \lambda = 2, \mu = 1, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	11.1	4.49	3.61	3.22	3.01	2.95	2.89
$\mathbb{E}[L_2]$	0.889	0.492	0.288	0.154	0.063	0.032	0
$\mathbb{E}[L_q]$	8.21	1.99	1.33	1.07	0.95	0.92	0.89
$\xi\mathbb{E}[L_2]$	0.0889	0.246	0.288	0.308	0.315	0.32	0.32

Table 2 Numerical results for $N = 3, T = 5, \lambda = 2, \mu = 1, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	13.7	6.1	4.59	3.71	3.19	3.03	2.89
$\mathbb{E}[L_2]$	0.987	0.833	0.604	0.359	0.155	0.079	0
$\mathbb{E}[L_q]$	10.71	3.27	1.99	1.35	1.04	0.96	0.89
$\xi\mathbb{E}[L_2]$	0.0987	0.4165	0.604	0.718	0.775	0.79	0.802

Table 3 Numerical results for $N = 3, T = 5, \lambda = 2, \mu = 1, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	14.52	7.01	5.38	4.18	3.38	3.13	2.89
$\mathbb{E}[L_2]$	0.99	0.953	0.801	0.532	0.243	0.125	0
$\mathbb{E}[L_q]$	11.53	4.06	2.57	1.65	1.14	1.00	0.89
$\xi\mathbb{E}[L_2]$	0.099	0.4765	0.801	1.064	1.215	1.25	1.284

Since R is a lower triangular matrix, the lower and upper limits in the summations of Eq. (37) above are $k = i$, and $k = i + n$, respectively. Thus,

$$-r_{i+n,i}(\lambda + (N-i)\mu + i\xi) + (N-i)\mu \sum_{k=i}^{i+n} r_{i+n,k} r_{k,i} + \sum_{k=i+1}^{i+n} r_{i+n,k} r_{k,i+1} = 0. \quad (38)$$

Extracting $r_{i+n,i}$ from (38) results in (32), in which $r_{i+n,i}$ is expressed in terms of all r_{ij} calculated before. This completes the proof. □

4 Numerical results and optimization over T and over λ

Section 4.1 bellow presents numerical results for a set of parameter values. Section 4.2 follows when we first calculate the optimal value of the threshold T for a given range of arrival intensities, and then, assuming that the arrival rate can be controlled, we calculate the optimal value of λ for various values of the threshold T.

4.1 Numerical results

The following three tables (Tables 1, 2, 3) present our first set of numerical results, where the calculated performance measures in all tables are $\mathbb{E}[L]$, $\mathbb{E}[L_2]$, $\mathbb{E}[L_q]$ and $\xi\mathbb{E}[L_2]$, the exit rate from cross selling. The three tables maintain the same parameter values: $N = 3$, $T = 5$, $\lambda = 2$ and $\mu = 1$, but they differ by the values of the parameter p, which assumes the values 0.2, 0.5 and 0.8, respectively. In each table the changing parameter is ξ, assuming values between 0.1 and infinity. Notice that, in the limiting case, when ξ approaches infinity (while $\mathbb{E}[L_2]$ tends to zero), the system converges to an $M(\lambda)/M(\mu)/N$ queue. In steady-state, the exit rate from cross-selling equals the entrance rate, so when $\xi \to \infty$ we have (for $T > N$),

$$\text{Entrance rate} = \xi\mathbb{E}[L_2] \to p\mu\left(\sum_{m=1}^{N} mP_{0m} + \sum_{m=N+1}^{T} NP_{0m}\right),$$

where the probabilities P_{0m}, for $1 \le m \le T$ are the steady-state probabilities for the well-known $M(\lambda)/M(\mu)/N$ queue. The numerical results exhibit that this limit is reached quite fast. The calculations are summarized in Tables 1, 2 and 3 with parameter values $N = 3$, $T = 5$, $\lambda = 2$ and $\mu = 1$.

Additional tables appear in Appendix A.2 with various values of the systems parameters. In all tables in Appendix A.2 it is seen that each performance measure, $\mathbb{E}[L]$, $\mathbb{E}[L_2]$ and $\mathbb{E}[L_q]$ is a monotone decreasing function of the cross-selling rate, ξ. When ξ approaches infinity, $\mathbb{E}[L_2]$ tends to 0 (since cross selling is instantaneous, implying that there are no cross-selling customers in the system). However, the mean rate of system cross-selling, $\xi\mathbb{E}[L_2]$, approaches a positive asymptote (i.e., 0.32 in Table 1, 0.802 in Table 2, 1.284 in Table 3, etc.). Moreover, for a given set of parameters, when ξ approaches infinity, $\mathbb{E}[L]$ is unaffected by p. Furthermore, in each set of 3 tables, larger values of p increase the values of each performance measure. The same effect occurs for larger values of the threshold T.

4.2 Optimization over T and over λ

4.2.1 Optimal T

In this section we set $N = 5$ and calculate the optimal value of T, subject to the following cost parameters. Let c denote the penalty the system incurs for one unit of a customer's waiting time, and let r denote the reward gained from each customer going through cross-selling. Let Π be the system's mean profit per unit time, i.e. $\Pi = r\xi\mathbb{E}[L_2] - c\mathbb{E}[L_q]$. Without loss of generality, we can assume that $r = 1$ and write $\Pi = \xi\mathbb{E}[L_2] - c\mathbb{E}[L_q]$. Note that for larger values of T, $\mathbb{E}[L_2]$ increases, causing a higher system's penalty, while for smaller values of T, the queue size drops, but the system loses potential cross-selling customers.

Figures 2, 3, 4 and 5 describe the behavior of the mean profit Π as a function of the Threshold T, for $c = 0.1, 0.25, 0.5$ and 1, respectively. The rest of the parameters are set to $N = 5, \mu = 2, \xi = 1$ and $p = 0.5$, and for each figure we consider five values of λ: $\lambda = 4, 5, 6, 7$ and 8. It is seen from the figures that for larger values of λ, the profit Π decreases drastically as T increases, where for small values of λ the decrease in Π is more moderate. This occurs, since for larger values of T, a high arrival rate results in more customers in cross-selling as well as a higher waiting time. Also, as the cost of waiting time (i.e. c) increases, the optimal value of T decreases. For example, when $c = 0.1$ and $\lambda = 5$, the optimal value of T is 10 customers, and when $c = 0.5$ and $\lambda = 5$, the optimal value of T is 6 customers. However, when $c = 1$ and $\lambda = 8$, it is not profitable to offer any cross-selling.

4.2.2 Optimal λ

Assuming that the arrival rate λ can be controlled, one can find its optimal value, for any given value of T. This is seen in Tables 4, 5 and 6, calculated for $T = 1$, $T = 5$, and $T = 10$, respectively. The optimal λ value appears in bold in each table. This is further illustrated in Fig. 6, where the mean profit rate function Π is concave with a single maximum.

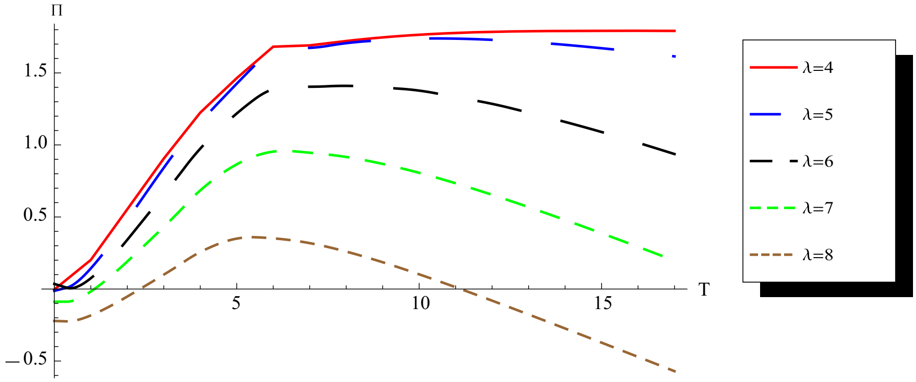

Fig. 2 Mean profit rate, Π, as a function of T, for $N = 5$, $\mu = 2$, $\xi = 1$, $p = 0.5$ and $c = 0.1$

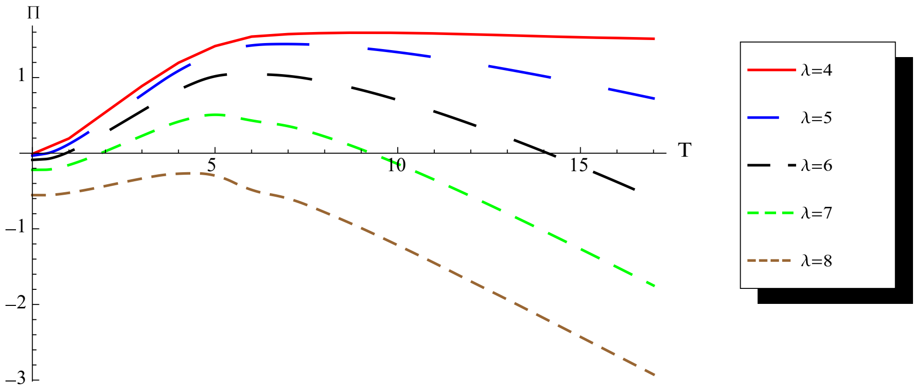

Fig. 3 Mean profit rate, Π, as a function of T, for $N = 5$, $\mu = 2$, $\xi = 1$, $p = 0.5$ and $c = 0.25$

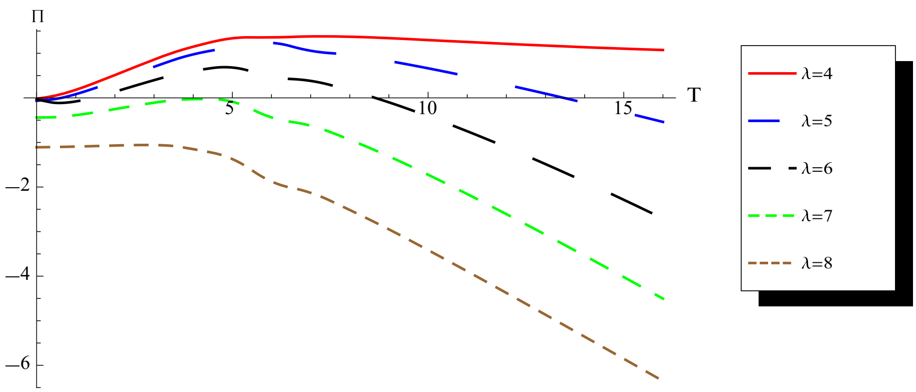

Fig. 4 Mean profit rate, Π, as a function of T, for $N = 5$, $\mu = 2$, $\xi = 1$, $p = 0.5$ and $c = 0.5$

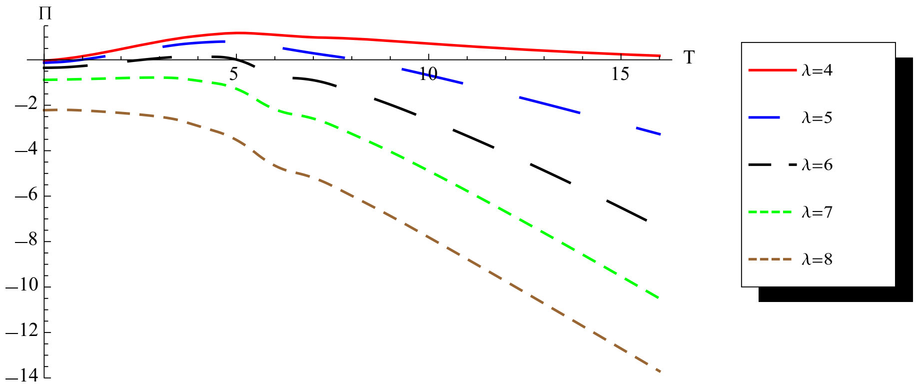

Fig. 5 Mean profit rate, Π, as a function of T, for $N = 5$, $\mu = 2$, $\xi = 1$, $p = 0.5$ and $c = 1$

Table 4 Results for mean profit rate, Π, for $N = 5$, $T = 1$, $\mu = 2$, $\xi = 1$, $p = 0.5$, $c = 0.25$

λ	0.5	1	1.5	2	2.5	3	4
Π	0.1713	0.2476	0.2771	**0.2809**	0.2701	0.2502	0.1942

The optimal profit is marked in bold

Table 5 Results for mean profit rate, Π, for $N = 5$, $T = 5$, $\mu = 2$, $\xi = 1$, $p = 0.5$, $c = 0.25$

λ	0.5	1	2	3	3.5	4	4.5	5
Π	0.2499	0.4987	0.9618	1.2926	1.3823	**1.4160**	1.3942	1.3189

The optimal profit is marked in bold

5 Concluding remarks

This paper deals with a multi-server queueing system where customers, after completing service, may be directed to a cross-selling phase, as long as the total number of customers in the system is below some threshold level, T. We present exact analysis of the system via both the PGFs and the matrix geometric methodologies, and reveal some intrinsic relations between the two methods. Specifically, we show that the elements of the rate matrix R are expressed in terms of the roots a determinant of a matrix related to the set of linear equations involving the PGFs. This result is important, as the full connection between the two solution methodologies is not yet known. We also show that the stability condition of the cross selling system is the same as for an identical system but without cross selling at all, namely $\lambda < N\mu$. Finally, we present numerical results for different sets of parameters and calculate the mean total number of customers in the system, the mean number of waiting customers, the mean number of customers in cross-selling, and the cross selling rate. Our results show monotonicity of these performance measures in the various problem parameters. Interestingly, we observe that in the limit as ξ goes to ∞, the expected cross selling rate converges to a limit that is consistently less than $p\lambda$, and approaches this value when T is large. Furthermore, for different scenarios, we calculate the optimal value of the threshold T which maximizes the system's profit. Our results illustrate that in optimality one may have $T < N$. That is, it may be optimal to stop initiating cross selling even if some of the servers are idle. Also, if

Table 6 Results for mean profit rate, Π, for $N = 5, T = 10, \mu = 2, \xi = 1, p = 0.5, c = 0.25$

λ	0.5	1	2	3	3.5	4	4.5	5
Π	0.2499	0.4998	0.9899	1.4114	1.5464	**1.5903**	1.5202	1.3353

The optimal profit is marked in bold

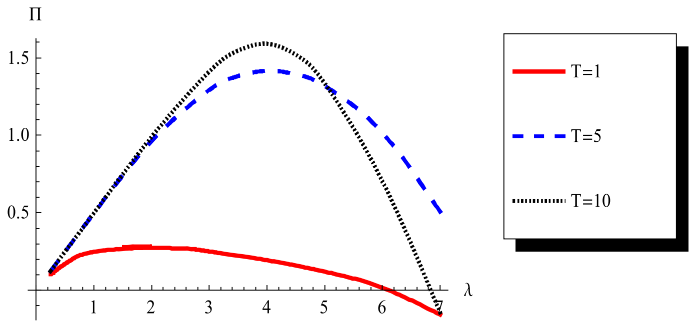

Fig. 6 Π as a function of λ, for $T = 1, 5$ and 10, where $N = 5, \mu = 2, \xi = 1, p = 0.5, c = 0.25$

the arrival rate λ can be controlled, there is an optimal rate that can be calculated, as depicted in Fig. 6.

Our research may be extended in several directions. First, what if the service has more than one discretionary phase? In this case, one might consider employing multiple threshold levels that indicate whether to initiate an additional cross-selling service phase, or completing customer's service at the current phase. This system would be of higher dimension than 2, leading to a more complex analysis. Second, numerical analysis could further investigate what is the optimal number of servers to employ in order to maximize revenue minus holding and staffing costs. Finally, this paper considers a threshold policy for the initiation of cross selling. However, it is possible that the optimal policy takes into account not only the total number of customers in the system, but also how many are in each phase, so that the optimal policy may take the form of a switching curve rather than a one dimensional threshold.

A Appendix

A.1 The case $T < N$

We present here briefly the analysis of the case where $T < N$. We first write the balance equations of the system in steady-state. We then write expressions for the PGF's and construct a system of linear equations in order to calculate the boundary probabilities appearing in the PGF's.

For $j = 0$,

$$\begin{aligned}
\lambda P_{00} &= q\mu P_{01} + \xi P_{11}, \\
(\lambda + m\mu)\, P_{0m} &= \lambda P_{0,m-1} + (m+1)q\mu P_{0,m+1} + \xi P_{1,m+1}, \quad 1 \le m \le T-1, \\
(\lambda + m\mu)\, P_{0m} &= \lambda P_{0,m-1} + (m+1)\mu P_{0,m+1} + \xi P_{1,m+1}, \quad T \le m \le N-1, \\
(\lambda + N\mu)\, P_{0m} &= \lambda P_{0,m-1} + N\mu P_{0,m+1} + \xi P_{1,m+1}, \quad m \ge N.
\end{aligned}$$

For $1 \le j \le T-1$,

$m = j$:

$$(\lambda + j\xi)\, P_{jm} = p\mu P_{j-1,m} + q\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1},$$

$j+1 \le m \le T-1$:

$$(\lambda + j\xi + (m-j)\mu)\, P_{jm} = \lambda P_{j,m-1} + (m-j+1)p\mu P_{j-1,m} + (m+1-j)q\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1},$$

$m = T$:

$$(\lambda + j\xi + (m-j)\mu)\, P_{jm} = \lambda P_{j,m-1} + (m-j+1)p\mu P_{j-1,m} + (m-j+1)\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1},$$

$T+1 \le m \le N-1$:

$$(\lambda + j\xi + (m-j)\mu)\, P_{jm} = \lambda P_{j,m-1} + (m-j+1)\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1},$$

$m \ge N$:

$$(\lambda + j\xi + (N-j)\mu)\, P_{jm} = \lambda P_{j,m-1} + (N-j)\mu P_{j,m+1} + (j+1)\xi P_{j+1,m+1}.$$

For $j = T$,

$$\begin{aligned}
&(\lambda + T\xi)\, P_{Tm} = p\mu P_{T-1,m} + \mu P_{T,m+1}, \quad m = T,\\
&(\lambda + (m-j)\mu + T\xi)\, P_{Tm} = \lambda P_{T,m-1} + (m-j+1)\mu P_{T,m+1}, \quad T+1 \le m \le N-1,\\
&(\lambda + (N-j)\mu + T\xi)\, P_{Tm} = \lambda P_{T,m-1} + (N-j)\mu P_{T,m+1}, \quad m \ge N.
\end{aligned}$$

Recall that $G_j(z) = \sum_{m=j}^{\infty} P_{jm} z^m$. For each $j = 0, 1, \ldots, T$, multiplying each of the balance equation by z^m for the appropriate m and summing over m, results in

$$\begin{aligned}
&((\lambda z - N\mu)(1-z))\, G_0(z) - \xi G_1(z)\\
&\quad = \mu \sum_{m=0}^{N-1} (N-m) P_{0m} z^{m+1} + q\mu \sum_{m=0}^{T-1} (m+1) P_{0,m+1} z^{m+1} + \mu \sum_{m=T}^{N-1} (m+1) P_{0,m+1} z^{m+1}\\
&\qquad - N\mu \sum_{m=0}^{N} P_{0m} z^m,
\end{aligned} \tag{39}$$

$$\begin{aligned}
&((\lambda z - (N-j)\mu)(1-z)) + j\xi z)\, G_j(z) - (j+1)\xi G_{j+1}(z)\\
&\quad = \mu \sum_{m=j}^{N-1} (N-m) P_{jm} z^{m+1} + p\mu \sum_{m=j}^{T} (m-j+1) P_{j-1,m} z^{m+1}\\
&\qquad + \mu \sum_{m=T}^{N-1} (m-j+1) P_{j,m+1} z^{m+1}\\
&\qquad + q\mu \sum_{m=j}^{T-1} (m-j+1) P_{j,m+1} z^{m+1} - (N-j)\mu \sum_{m=j}^{N} P_{jm} z^m, \quad j = 1, \ldots, T-1,
\end{aligned} \tag{40}$$

and

$$
\begin{aligned}
((\lambda z-\mu(N-T))(1-z)+T\xi z)\,G_T(z) &= \mu(N-T)\sum_{m=T}^{N-1} P_{Tm}z^{m+1}\\
&+p\mu P_{T-1,T}z^{T+1}+\mu\sum_{m=T}^{N-1}(m-T+1)P_{T,m+1}z^{m+1}\\
&-(N-T)\mu\sum_{m=T}^{N}P_{Tm}z^{m}-\mu\sum_{m=T+1}^{N-1}(m-T)P_{Tm}z^{m+1}. \qquad (41)
\end{aligned}
$$

The set of Eqs. (39)–(41) can be written as

$$A(z)\cdot\mathbf{G}(z)=\mathbf{b}(z),$$

where

$$
A(z)=\begin{pmatrix}
\alpha_0(z) & -\xi & 0 & \cdots & \cdots & \cdots & 0\\
0 & \alpha_1(z) & -2\xi & 0 & \cdots & \cdots & \vdots\\
0 & 0 & \alpha_2(z) & -3\xi & 0 & \ddots & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0\\
\vdots & \ddots & \ddots & \ddots & 0 & \alpha_{T-1}(z) & -T\xi\\
0 & \cdots & \cdots & 0 & 0 & 0 & \alpha_T(z)
\end{pmatrix},
$$

$\mathbf{G}(z)=(G_0(z),G_1(z),\ldots,G_T(z))^T$ is a $(T+1)$-dimensional column vector of the desired PGF's, and

$$\alpha_j(z)=(\lambda z-(N-j)\mu)(1-z)+j\xi z,\quad j=0,1,...,T.$$

The elements of the vector $\mathbf{b}(z)$ are

$$
\begin{aligned}
b_0(z)=\mu\sum_{m=0}^{N-1}(N-m)P_{0m}z^{m+1}&+q\mu\sum_{m=0}^{T-1}(m+1)P_{0,m+1}z^{m+1}\\
&+\mu\sum_{m=T}^{N-1}(m+1)P_{0,m+1}z^{m+1}-N\mu\sum_{m=0}^{N}P_{0m}z^{m};
\end{aligned}
$$

For $j=1,\ldots,T-1$,

$$
\begin{aligned}
b_j(z)=\mu\sum_{m=j}^{N-1}(N-m)P_{jm}z^{m+1}&+p\mu\sum_{m=j}^{T}(m-j+1)P_{j-1,m}z^{m+1}\\
&+\mu\sum_{m=T}^{N-1}(m-j+1)P_{j,m+1}z^{m+1}\\
&+q\mu\sum_{m=j}^{T-1}(m-j+1)P_{j,m+1}z^{m+1}-(N-j)\mu\sum_{m=j}^{N}P_{jm}z^{m};
\end{aligned}
$$

and,

$$b_T(z) = \mu(N-T)\sum_{m=T}^{N-1} P_{Tm}z^{m+1} + p\mu P_{T-1,T}z^{T+1} + \mu\sum_{m=T}^{N-1}(m-T+1)P_{T,m+1}z^{m+1}$$
$$-(N-T)\mu\sum_{m=T}^{N} P_{Tm}z^m - \mu\sum_{m=T+1}^{N-1}(m-T)P_{Tm}z^{m+1}.$$

To obtain $G_j(z)$ we use Cramer's rule and write $G_j(z) = \frac{|A_j(z)|}{|A(z)|}$, $j = 0, 1, \ldots, T$, where $|A|$ is the determinant of a matrix A, and $A_j(z)$ is the matrix obtained from $A(z)$ by replacing its jth column by $\mathbf{b}(z)$. The functions $G_j(z)$ are expressed in terms of $\frac{1}{2}(T+1)(2N-T)$ unknown 'boundary probabilities', P_{jm}, $j = 0, 1, \ldots, , T$ and $m = j, j+1, \ldots, N-1$ [appearing in $\mathbf{b}(z)$]. In order to derive these boundary probabilities we need to create a set of equations that those probabilities satisfy. First, we consider the balance equations of states P_{jm}, for $0 \le j \le T$ and $j \le m \le N-2$. This gives us $\frac{1}{2}(T+1)(2N-T-2)$ equations.

As in Sect. 2, utilize the roots of $|A(z)|$ which are less than 1, provides us with T more equations relating the boundary probabilities. These roots are

$$z_{j,1} = \frac{\lambda + (N-j)\mu + j\xi - \sqrt{(\lambda + (N-j)\mu + j\xi)^2 - 4\lambda(N-j)\mu}}{2\lambda}, \quad j = 1, 2, \ldots, T.$$

Another equation is the normalization equation, i.e. $\sum_{j=0}^{T} G_j(1) = 1$.

In addition, once all boundary probabilities are calculated, the marginal probabilities of having j customers in cross-selling, i.e. $P_{j\bullet}$, can be derived from

$$p\mu\sum_{m=j+1}^{T}(m-j)P_{jm} = (j+1)\xi P_{j+1,\bullet}, \quad j = 0, 1, \ldots, T-1,$$

and $P_{0\bullet} = 1 - \sum_{j=1}^{T} P_{j\bullet}$.

The mean total number of customers in the system, the mean number of customers in cross-selling and the mean number of waiting customers are given, respectively, by

$$\mathbb{E}[L] = \sum_{j=0}^{T}\frac{d}{dz}G_j(z)|_{z=1} = \sum_{m=0}^{\infty} mP_{\bullet m},$$
$$\mathbb{E}[L_2] = \sum_{j=1}^{T} jP_{j\bullet},$$
$$\mathbb{E}[L_q] = \mathbb{E}[L] - N + \sum_{m=0}^{N}(N-m)P_{\bullet m}.$$

A.2 Numerical results

We present tables to supplement Sect. 4.1, where in each set the changing parameter is the probability p:
In set 1 (Tables 7, 8, 9) the parameter values are $N = 3$, $T = 5$, $\lambda = 4$ and $\mu = 3$.
In set 2 (Tables 10, 11, 12) the parameter values are $N = 3$, $T = 10$, $\lambda = 2$ and $\mu = 1$.
In set 3 (Tables 13, 14, 15) the parameter values are $N = 3$, $T = 10$, $\lambda = 4$ and $\mu = 3$.
In set 4 (Tables 16, 17, 18) the parameter values are $N = 5$, $T = 10$, $\lambda = 2$ and $\mu = 1$.
In set 5 (Tables 19, 20, 21) the parameter values are $N = 5$, $T = 10$, $\lambda = 4$ and $\mu = 3$.

Table 7 Numerical results for $N = 3, T = 5, \lambda = 4, \mu = 3, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	15.53	4.28	2.74	2.05	1.68	1.58	1.48
$\mathbb{E}[L_2]$	1.565	1.07	0.676	0.369	0.152	0.077	0
$\mathbb{E}[L_q]$	12.64	1.89	0.73	0.34	0.19	0.17	0.14
$\xi\mathbb{E}[L_2]$	0.1565	0.535	0.676	0.738	0.76	0.77	0.77

Table 8 Numerical results for $N = 3, T = 5, \lambda = 4, \mu = 3, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	18.88	6.71	4.46	2.94	2.01	1.73	1.48
$\mathbb{E}[L_2]$	1.656	1.543	1.285	0.843	0.374	0.191	0
$\mathbb{E}[L_q]$	15.89	3.835	1.84	0.76	0.32	0.21	0.14
$\xi\mathbb{E}[L_2]$	0.1656	0.7715	1.285	1.686	1.87	1.91	1.93

Table 9 Numerical results for $N = 3, T = 5, \lambda = 4, \mu = 3, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	19.77	7.74	5.61	3.80	2.35	1.89	1.48
$\mathbb{E}[L_2]$	1.666	1.647	1.548	1.198	0.586	0.303	0
$\mathbb{E}[L_q]$	16.78	4.76	2.72	1.27	0.43	0.25	0.14
$\xi\mathbb{E}[L_2]$	0.1666	0.8235	1.548	2.396	2.93	3.03	3.086

Table 10 Numerical results for $N = 3, T = 10, \lambda = 2, \mu = 1, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	16.02	6.18	4.26	3.48	3.10	3.01	2.89
$\mathbb{E}[L_2]$	0.97	0.65	0.37	0.19	0.078	0.04	0
$\mathbb{E}[L_q]$	13.04	3.53	1.89	1.29	1.03	0.97	0.89
$\xi\mathbb{E}[L_2]$	0.097	0.325	0.37	0.38	0.39	0.4	0.4

Table 11 Numerical results for $N = 3, T = 10, \lambda = 2, \mu = 1, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	18.77	10.27	6.88	4.56	3.45	3.19	2.89
$\mathbb{E}[L_2]$	1.00	0.96	0.77	0.46	0.19	0.10	0
$\mathbb{E}[L_q]$	15.77	7.31	4.10	2.10	1.26	1.09	0.89
$\xi\mathbb{E}[L_2]$	0.1	0.48	0.77	0.92	0.95	0.96	0.974

Table 12 Numerical results for $N = 3, T = 10, \lambda = 2, \mu = 1, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	19.53	11.88	9.21	5.87	3.84	3.25	2.89
$\mathbb{E}[L_2]$	1.00	0.99	0.95	0.68	0.31	0.15	0
$\mathbb{E}[L_q]$	16.53	8.88	6.26	3.18	1.53	1.11	0.89
$\xi\mathbb{E}[L_2]$	0.1	0.495	0.95	1.36	1.55	1.55	1.55

Table 13 Numerical results for $N = 3, T = 10, \lambda = 4, \mu = 3, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	21.11	6.35	3.28	2.16	1.71	1.59	1.48
$\mathbb{E}[L_2]$	1.64	1.27	0.78	0.41	0.16	0.08	0
$\mathbb{E}[L_q]$	18.14	3.75	1.17	0.42	0.21	0.18	0.14
$\xi\mathbb{E}[L_2]$	0.164	0.635	0.78	0.8	0.8	0.8	0.8

Table 14 Numerical results for $N = 3, T = 10, \lambda = 4, \mu = 3, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	24.01	11.39	7.55	3.72	2.10	1.76	1.48
$\mathbb{E}[L_2]$	1.67	1.65	1.50	0.97	0.41	0.20	0
$\mathbb{E}[L_q]$	21.01	8.41	4.72	1.42	0.36	0.23	0.14
$\xi\mathbb{E}[L_2]$	0.167	0.825	1.50	1.94	2.0	2.0	2.0

Table 15 Numerical results for $N = 3, T = 10, \lambda = 4, \mu = 3, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	10	∞
$\mathbb{E}[L]$	24.77	12.72	10.24	6.08	2.58	1.95	1.48
$\mathbb{E}[L_2]$	1.67	1.67	1.65	1.40	0.65	0.33	0
$\mathbb{E}[L_q]$	21.77	9.72	7.25	3.34	0.60	0.29	0.14
$\xi\mathbb{E}[L_2]$	0.167	0.835	1.65	2.8	3.2	3.2	3.2

Table 16 Numerical results for $N = 5, T = 10, \lambda = 2, \mu = 1, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	9.05	3.03	2.49	2.26	2.13	2.04
$\mathbb{E}[L_2]$	2.555	0.793	0.399	0.199	0.079	0
$\mathbb{E}[L_q]$	4.49	0.24	0.10	0.06	0.05	0.04
$\xi\mathbb{E}[L_2]$	0.2555	0.3965	0.399	0.399	0.399	0.3997

Table 17 Numerical results for $N = 5, T = 10, \lambda = 2, \mu = 1, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	13.34	5.05	3.28	2.61	2.28	2.04
$\mathbb{E}[L_2]$	2.978	1.883	0.993	0.499	0.206	0
$\mathbb{E}[L_q]$	8.35	1.17	0.29	0.11	0.06	0.04
$\xi\mathbb{E}[L_2]$	0.2978	0.9415	0.993	0.998	0.999	0.999

Table 18 Numerical results for $N = 5, T = 10, \lambda = 2, \mu = 1, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	14.47	7.42	4.23	2.98	2.55	2.04
$\mathbb{E}[L_2]$	2.991	2.619	1.566	0.796	0.522	0
$\mathbb{E}[L_q]$	9.47	2.80	0.67	0.40	0.10	0.04
$\xi\mathbb{E}[L_2]$	0.2991	1.3095	1.566	1.592	1.599	1.599

Table 19 Numerical results for $N = 5, T = 10, \lambda = 4, \mu = 3, p = 0.2$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	13.39	3.29	2.19	1.75	1.52	1.34
$\mathbb{E}[L_2]$	3.451	1.571	0.799	0.399	0.159	0
$\mathbb{E}[L_q]$	8.61	0.39	0.06	0.02	0.008	0.004
$\xi\mathbb{E}[L_2]$	0.3451	0.7855	0.799	0.799	0.799	0.7998

Table 20 Numerical results for $N = 5, T = 10, \lambda = 4, \mu = 3, p = 0.5$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	17.33	7.39	3.83	2.41	1.75	1.34
$\mathbb{E}[L_2]$	3.656	3.182	1.964	0.999	0.399	0
$\mathbb{E}[L_q]$	12.34	2.88	0.54	0.08	0.02	0.004
$\xi\mathbb{E}[L_2]$	0.3656	1.591	1.964	1.997	1.997	1.997

Table 21 Numerical results for $N = 5, T = 10, \lambda = 4, \mu = 3, p = 0.8$

Values of ξ	0.1	0.5	1	2	5	∞
$\mathbb{E}[L]$	18.28	10.15	6.02	3.17	2.00	1.34
$\mathbb{E}[L_2]$	3.661	3.601	2.930	1.595	0.639	0
$\mathbb{E}[L_q]$	13.28	5.20	1.76	0.25	0.03	0.004
$\xi\mathbb{E}[L_2]$	0.3661	1.8005	2.930	3.19	3.195	3.1995

References

Aksin, O. Z., & Harker, P. T. (1999). To sell or not to sell determining the trade-offs between service and sales in retail banking phone centers. *Journal of Service Research*, *2*(1), 19–33.

Alizamir, S., de Véricourt, F., & Sun, P. (2013). Diagnostic accuracy under congestion. *Management Science*, *59*(1), 157–171.

Anand, K. S., Pac, M. F., & Veeraraghavan, S. (2011). Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Science*, *57*(1), 40–56.

Armony, M., & Gurvich, I. (2010). When promotions meet operations: Cross-selling and its effect on call center performance. *Manufacturing & Services Operations Management*, *12*(3), 470–488.

Armony, M., & Maglaras, C. (2004a). Contact centers with a call-back option and real-time delay information. *Operations Research*, *52*(4), 527–545.

Armony, M., & Maglaras, C. (2004b). On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Operations Research*, *52*(2), 271–292.

Artalejo, J. R., & Gómez-Corral, A. (2008). *Retrial queueing systems: A computational approach*. Berlin: Springer.

Bhulai, S., & Koole, G. (2003). A queueing model for call blending in call centers. *IEEE Transactions on Automatic Control*, *48*(8), 1434–1438.

Byers, R. E., & So, K. C. (2007). Note—A mathematical model for evaluating cross-sales policies in telephone service centers. *Manufacturing & Services Operations Management*, *9*(1), 1–8.

Gans, N., & Zhou, Y. P. (2003). A call-routing problem with service-level constraints. *Operations Research*, *51*(2), 255–271.

Gurvich, I., Armony, M., & Maglaras, C. (2009). Cross-selling in a call center with a heterogeneous customer population. *Operations research*, *57*(2), 299–313.

Gurvich, I., Armony, M., & Mandelbaum, A. (2008). Service-level differentiation in call centers with fully flexible servers. *Management Science*, *54*(2), 279–294.

Hanukov, G., Avinadav, T., Chernonog, T., Spiegel, U., & Yechiali, U. (2017). A queueing system with decomposed service and inventoried preliminary services. *Applied Mathematical Modelling*, *47*, 276–293.

Kim, C. S., Lee, M. H., Dudin, A., & Klimenok, V. (2008). Multi-server queueing systems with cooperation of the servers. *Annals of Operations Research*, *162*(1), 57–68.
Kostami, V., & Rajagopalan, S. (2013). Speed-quality trade-offs in a dynamic model. *Manufacturing & Services Operations Management*, *16*(1), 104–118.
Latouche, G., & Ramaswami, V. (1999). Introduction to matrix analytic methods in stochastic modeling. SIAM
Lerzan Örmeci, E., & Zeynep Akşin, O. (2010). Revenue management through dynamic cross selling in call centers. *Production and Operations Management*, *19*(6), 742–756.
Litvak, N., & Yechiali, U. (2003). Routing in queues with delayed information. *Queueing Systems*, *43*(1), 147–165.
Mills, A., Argon, N. T., & Ziya, S. (2013). Resource-based patient prioritization in mass-casualty incidents. *MSOM*, *15*, 361–377.
Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models: An algorithmic approach*. Baltimore: The Johns Hopkins University Press.
Pang, G., & Perry, O. (2014). A logarithmic safety staffing rule for contact centers with call blending. *Management Science*, *61*(1), 73–91.
Paz, N., & Yechiali, U. (2014). An m/m/1 queue in random environment with disasters. *Asia-Pacific Journal of Operational Research*, *31*(03), 1450,016.
Perel, E., & Yechiali, U. (2008). Queues where customers of one queue act as servers of the other queue. *Queueing Systems*, *60*(3), 271–288.
Perel, E., & Yechiali, U. (2013a). On customers acting as servers. *Asia-Pacific Journal of Operational Research*, *30*(5)
Perel, N., & Yechiali, U. (2013b). The Israeli queue with retrials. *Queueing Systems*, *78*(1), 31–56.
Perel, N., & Yechiali, U. (2014). The Israeli queue with priorities. *Stochastic Models*, *29*(3), 353–379.
Perel, E., & Yechiali, U. (2017). Two-queue polling systems with switching policy based on the queue that is not being served. *Stochastic Models*, *33*(3), 430–450. https://doi.org/10.1080/15326349,1312446.
Perry, O., & Whitt, W. (2009). Responding to unexpected overloads in large-scale service systems. *Management Science*, *55*(8), 1353–1367.
Wang, X., Debo, L. G., Scheller-Wolf, A., & Smith, S. F. (2010). Design and analysis of diagnostic service centers. *Management Science*, *56*(11), 1873–1890.
Xu, Y., Scheller-Wolf, A., & Sycara, K. (2015). The benefit of introducing variability in single-server queues with application to quality-based service domains. *Operations Research*, *63*(1), 233–246.
Zhan, D., & Ward, A. R. (2015). Compensation and staffing to trade off speed and quality in large service systems. Available at SSRN.