

TRANSIENT ANALYSIS FOR EXPONENTIAL TIME-LIMITED POLLING MODELS UNDER THE PREEMPTIVE REPEAT RANDOM POLICY

ROLAND DE HAAN,* CQM

AHMAD AL HANBALI,** King Fahd University of Petroleum and Minerals

RICHARD J. BOUCHERIE,*** AND

JAN-KEES VAN OMMEREN,*** University of Twente

Abstract

Polling systems are queueing systems consisting of multiple queues served by a single server. In this paper we analyze two types of preemptive time-limited polling systems, the so-called pure and exhaustive time-limited disciplines. In particular, we derive a direct relation for the evolution of the joint queue length during the course of a server visit. The analysis of the pure time-limited discipline builds on and extends several known results for the transient analysis of an M/G/1 queue. For the analysis of the exhaustive discipline we derive several new results for the transient analysis of the M/G/1 queue during a busy period. The final expressions for both types of polling systems that we obtain generalize previous results by incorporating customer routing, generalized service times, batch arrivals, and Markovian polling of the server.

Keywords: Queueing; polling systems; transient analysis

2010 Mathematics Subject Classification: Primary 60K25

Secondary 90B22

1. Introduction

Polling systems are queueing systems consisting of multiple queues served by a single server. There are many applications of polling models. For instance, traffic light systems, multiple-access protocols for communication networks (e.g. IEEE 802.11), and product assembly systems can be modeled as a polling model. A more recent application field of polling systems is the area of wireless communication systems with mobile stations; see [8, Section 1.3]. The autonomous movements of such stations, hereby dynamically changing the network, create a specific need for studying time-limited polling models in which the visit time to a queue can be bounded by the server's mobility and not necessarily by its queue length. Another consequence of this mobility is that data packet transmissions may be preempted and must be repeated once connections are re-established. In wireless communications the transmission rate of data packets depends on the level of interference in the channel, which is typically changing with time [21]. Therefore, every time a connection is re-established the transmission rate of the data packets can be different. This makes a *preemptive-repeat-random*

Received 26 July 2018; revision received 14 July 2019.

* Postal address: CQM, The Netherlands

** Corresponding author. Postal address: Department of Systems Engineering, College of Computer Science and Engineering, King Fahd University of Petroleum and Minerals, PO Box 5063, Dhahran 31261, Kingdom of Saudi Arabia.

Email address: ahmad.alhanbali@kfupm.edu.sa

*** Postal address: Department of Stochastic Operations Research, University of Twente, The Netherlands.

policy more appropriate in this context. According to this latter policy, if a service is interrupted, then at the next server visit a new service time will be drawn from the original service time distribution. Well-known surveys of a broad class of polling models include [19], [20], and [24], for example. For more recent surveys, we refer to [3], [4], and [23].

A key relation in the analysis of polling systems relates the joint queue length at the end of a server visit to queue i , denoted by Q_i , to the joint queue length at the start of the server visit to Q_i . This relation can be written in the following general form:

$$\beta^i(\mathbf{z}) = \mathcal{F}(\alpha^i(\mathbf{z})), \quad (1)$$

where $\beta^i(\mathbf{z})$ is the probability generating function (p.g.f.) of the joint queue length at the end of a server visit to Q_i , $\alpha^i(\mathbf{z})$ is the p.g.f. of the joint queue length at the start of a server visit to Q_i , and \mathcal{F} is an operator representing the mapping between the queue lengths at these epochs and depends on the assumed service discipline.

In the analysis of polling systems a fundamental role is played by the so-called branching property; see [14] and [18]. Polling systems which operate under service disciplines satisfying this branching property (e.g. the exhaustive and gated disciplines) are amenable to a tractable analysis. De Haan, Boucherie, and van Ommeren [9] established the key relation, (1), for the pure exponential time-limited discipline with preemptive service in an indirect, recursive manner. According to the pure exponential time-limited discipline, the server visiting a queue continues to service this queue for a period of time that is exponentially distributed. Under the assumption of phase-type service times, a direct relation between $\beta^i(\mathbf{z})$ and $\alpha^i(\mathbf{z})$ is derived for this discipline using a matrix analytical approach in [2]. De Haan *et al.* [9] also rederived a result of [11] for the exhaustive exponential time-limited discipline for the special case of phase-type service times. According to the exhaustive exponential time-limited service discipline, the server visiting a queue continues to service a queue for an exponentially distributed period of time or until the queue becomes empty, whichever occurs first. Eliazar and Yechiali [11] studied the exhaustive exponential time-limited discipline for preemptive service; see also [12]. Observing that upon successful service completion at a queue the busy period in fact regenerates, Eliazar and Yechiali obtained a closed-form relation between the joint queue length at the end and the start of a server visit. Leung [16] analyzed the key relation, (1), for the exhaustive exponential time-limited discipline with non-preemptive service. This was done in a recursive manner by conditioning on specific intermediate events during a server visit. De Souza e Silva, Gail, and Muntz [10] studied the key relation for the exhaustive deterministic time-limited discipline for both preemptive and non-preemptive service. Under the assumption of exponential service times, they analyzed the transient behavior of the system by applying uniformization techniques to find the joint queue-length distribution $\beta^i(\mathbf{z})$. Note that for the special case of a two-queue polling model with exponential service times, Coffman, Fayolle, and Mitrani [6] found a closed-form solution of the generating function of the joint queue length by solving a boundary value problem. Unfortunately, this solution method appears to be inapplicable for more than two queues with general service time distribution.

In the present paper we derive a direct relation for the evolution of the joint queue length during the course of a server visit, that is, we relate $\beta^i(\mathbf{z})$ and $\alpha^i(\mathbf{z})$ in a direct manner. This is done for both the pure and the exhaustive exponential time-limited discipline with general service time requirements and preemptive service. More specifically, the service of individual customers is according to the *preemptive-repeat-random* policy. Moreover, we incorporate customer routing into our analysis, such that it may be applied to a large variety of queueing networks with a single server operating under one of the above-mentioned time-limited service disciplines. We further extend our model by considering batch Poisson arrivals and

Markovian polling of the server. The analysis of the pure time-limited discipline builds on Cohen's derivation in [7, Section II.4.3] and extends several of his results for the transient analysis of the M/G/1 queue. In addition, for the analysis of the exhaustive discipline we will derive several new results for the transient analysis of the M/G/1 queue during a busy period. The final expressions (for both the exhaustive and pure case) that we obtain for the key relations are of the form

$$\beta^i(\mathbf{z}) = d_1(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_2(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*),$$

where

$$\alpha^i(\mathbf{z}_i^*) := \alpha^i(z_1, \dots, z_{i-1}, l_i(\mathbf{z}), z_{i+1}, \dots, z_M),$$

$d_1(\mathbf{z})$, and $d_2(\mathbf{z})$ are functions of $\mathbf{z} = (z_1, \dots, z_M)$ which are largely determined by the Laplace–Stieltjes transform (LST) of the service-time distribution, and $l_i(\mathbf{z})$ is related to the length of the busy period of a customer at Q_i . These relations generalize the results in [2] by relaxing the phase-type distribution assumption on the service times and by incorporating customer routing and Markovian polling of the server. Compared to [11], our results include the following model extensions for the pure time-limited discipline: batch arrivals, customer routing, and Markovian polling of the server. This is done by extending several results of the transient analysis of the M/G/1 queue.

Note that complementary to these key relations, there exists a relation between the p.g.f.s $\beta^i(\mathbf{z})$ and $\alpha^i(\mathbf{z})$ which couples the queue length at the start of a server visit to Q_j to the queue length at the end of a server visit to Q_j . Together, these relations for all queues in the system give rise to a system of equations which may be solved numerically in an iterative fashion to derive the joint queue-length distributions. In this respect, the key relation for the time-limited disciplines derived in this paper not only presents a more elegant counterpart of the recursive relation obtained in [9] but also offers a computationally more attractive alternative. For more details of the incorporation of our relation into a computational scheme we refer to [2, Section 5]. Moreover, in [2, Sections 6 and 7], this scheme was evaluated from a computational point of view and compared to an existing numerical algorithm to show its benefits.

In summary, our contributions in this paper are as follows.

- We enrich the existing results on exponential time-limited models by incorporating batch Poisson arrivals, generalized service times, customer routing, and Markovian polling of the server.
- We derive new results for the transient probabilities of the M/G/1 queue and of the batch Poisson arrivals M/G/1 queue.
- Our derived key relations allow a computationally attractive procedure to find the joint queue-length distribution.

The rest of this paper is organized as follows. We describe the model and the notation in Section 2. The key relations for the pure and the exhaustive exponential time-limited discipline are presented in Sections 3 and 4, respectively. For clarity of exposition, in these latter sections we shall restrict ourselves to Poisson arrivals and fixed cyclic routing of the server, that is, upon visit completion to a queue the server switches to another queue according to a predetermined routing table. The model extensions to batch Poisson arrivals and Markovian polling of the server are separately analyzed in Section 5. We wrap up with a discussion of the final results for the key relations in Section 6. In Appendix A we give the transient analysis for

the M/G/1 queue during a busy period. The complete proofs of the key relations are given in Appendices B and C.

2. Model and notation

Consider a polling system with cyclic routing of the server and with M queues, $M \geq 1$. Each queue has Poisson arrivals, and independent and identically distributed service times. The model extension with batch Poisson arrivals or Markovian polling of the server is deferred to Section 5. In this paper we consider two service disciplines, that is,

- pure exponential time-limited discipline (P-TL),
- exhaustive exponential time-limited discipline (E-TL).

According to both disciplines, the server will visit a queue for at most an amount of time Y_i , which is exponentially distributed with rate ξ_i and independent of other visit times and of the service times. Under the E-TL discipline the server will move to the next queue as soon as the (currently visited) queue becomes empty or when the timer Y_i expires, whichever occurs first. Under the P-TL discipline the server remains at the queue until the timer expires. Under both disciplines, the customer receiving service will be preempted at the expiration of the timer and in such a case a new service time will be drawn from the original distribution at the start of the next visit; that is, we assume the preemptive-repeat with resampling strategy.

Customers who have completed their service at Q_i , $i = 1, \dots, M$, will join Q_j , $j = 1, \dots, M$, with probability $r_{ij} \geq 0$, and with probability $r_{i0} \geq 0$ they will leave the system. Clearly, these routing probabilities r_{ij} must satisfy $\sum_{j=0}^M r_{ij} = 1$, $i = 1, \dots, M$. We assume henceforth that $r_{ii} = 0$ and let $r^i(\mathbf{z})$ denote the p.g.f. of the number of arrivals to all queues generated by a single departing customer at Q_i , i.e. $r^i(\mathbf{z}) = r_{i0} + \sum_j r_{ij} z_j$. Note that $r_{ii} \neq 0$ may be incorporated if we redefine the service time at Q_i as the sum of a geometric number (with expectation $1/(1 - r_{ii})$) of independent original service times, and adjust the routing probabilities as follows: $r'_{ij} = r_{ij}/(1 - r_{ii})$ and $r'_{ii} = 0$.

The random variables I_i and X_i , $i = 1, \dots, M$, refer to the interarrival time of new customers and the service time of customers at Q_i . The three families of interarrival times, service times, and visit times are assumed to be independent.

The switch-over times are not considered in this paper. The only exceptions are in Lemmas 1 and 2, and in Section 5.2. This is because we focus purely on the key relation, equation (1). We refer to [2] and [9] for the incorporation of this relation into the framework for the computation of the joint queue-length probabilities. Namely, in [2 Section 5–7] an iterative algorithm was proposed. In this latter paper the performance of this algorithm was evaluated from a computational point of view, and compared with an existing algorithm.

It is assumed that the queues of the polling system are stable. In the following lemmas we shall state the stability condition for both the pure time-limited and the exhaustive time-limited systems. The proofs of these lemmas are straightforward extensions of those of Theorems 3.1 and 3.2 in [8]. Note that the stability proof in [8] relied largely on the stability proof of Fricker and Jaibi [13] for a class of polling systems with non-preemptive and work-conserving service disciplines.

Lemma 1. (Stability of the pure time-limited discipline.) *Let c_i denote the total expected switch-over time during a polling cycle. The stability of the pure time-limited discipline is as follows:*

$$\text{System is stable} \iff \rho_i < \kappa_i, \quad i = 1, \dots, M,$$

where

$$\rho_i = \lambda_i^* \cdot \frac{1 - \tilde{X}_i(\xi_i)}{c_t + \xi_i \tilde{X}_i(\xi_i)}, \quad \kappa_i = \frac{1/\xi_i}{c_t + \sum_{j=1}^M 1/\xi_j}, \quad \lambda_i^* = \lambda_i + \sum_{j=1}^M \lambda_j^* r_{ji}.$$

Note that $\tilde{X}_i(s)$ is the LST of the service time and $(1 - \tilde{X}_i(\xi_i))/(\xi_i \tilde{X}_i(\xi_i))$ is the expected value of the *effective service time* of a job in Q_i which includes the work lost due to service preemptions; κ_i is the availability fraction of the server at Q_i .

Lemma 2. (Stability of the exhaustive time-limited discipline.) *Let c_t denote the total expected switch-over time during a polling cycle. The stability of the exhaustive time-limited discipline is as follows:*

$$\text{System is stable} \iff \rho + \max_{i=1, \dots, M} \left(\lambda_i^* \frac{1 - \tilde{X}_i(\xi_i)}{\tilde{X}_i(\xi_i)} \right) \cdot c_t < 1,$$

where

$$\rho = \sum_{j=1}^M \frac{\lambda_j^* (1 - \tilde{X}_j(\xi_j))}{\xi_j \tilde{X}_j(\xi_j)}, \quad \lambda_i^* = \lambda_i + \sum_{j=1}^M \lambda_j^* r_{ji}.$$

Observe that ρ represents the total offered load to the system and $\tilde{X}_i(\xi_i)/(1 - \tilde{X}_i(\xi_i))$ is the mean number of served jobs at Q_i during a cycle when Q_i is saturated, that is, Q_i has infinitely many jobs waiting for service.

Here we introduce the notation that will be used below.

- Q : an arbitrary queue in the system.
- x_t : number of customers at time t at Q .
- $z_{(n)}$: number of customers left behind by the n th departing customer from Q .
- r'_n : time of the n th departure from Q .
- $D(t)$: number of departures from Q in $[0, t)$.
- $A(t)$: number of arrivals to Q in $[0, t)$.
- I : exponentially distributed random variable with parameter λ denoting the interarrival time to Q .
- X : generally distributed random variable denoting the service time at Q .
- $\mathbf{1}_A$: indicator function of event A .
- $\tilde{X}(\cdot)$: LST of random variable X .
- $\hat{\mu}(s, y)$: root x with the smallest absolute value less than one of $x = y \cdot \tilde{X}(s + \lambda(1 - x))$.
- \mathbf{N}_i^s : number of customers at all queues at the start of a server visit to Q_i .
- \mathbf{N}_i^e : number of customers at all queues at the end of a server visit to Q_i .
- $N_{i,j}(t)$: number of customers at Q_j at time t during a server visit to Q_i .
- $\alpha^i(\mathbf{z})$: p.g.f. of \mathbf{N}_i^s .

- $\beta^i(\mathbf{z})$: p.g.f. of \mathbf{N}_i^e .
- $r^i(\mathbf{z})$: p.g.f. of the number of arrivals to all queues generated by a single departing customer at Q_i .

3. Analysis of the pure time-limited service discipline

In this section we analyze the P-TL discipline. Under this discipline, the server will only depart from the queue when the time limit is reached. It should be stressed that the server will not leave the queue when it becomes empty. We will derive an expression for $\beta^i(\mathbf{z})$, the p.g.f. of the number of customers at all queues at the instant that the server leaves Q_i , in terms of the number of customers present at the start of the visit, $\alpha^i(\mathbf{z})$. Here, we present only the essential analytical steps and the main result. The detailed proofs are given in Appendix B.

Consider a visit of the server to an arbitrary queue Q . During this visit, the queue-length process at Q is a birth-and-death process, while the queue-length process at the other queues is a pure birth process. Note also that arrivals to these other queues may be both exogenous and endogenous (from Q). Our interest is in the number of customers at time t given an initial number of customers at the queues at the start of the server visit to Q . Moreover, to include customer routing in the analysis, we need to keep track of the number of departures during a visit. To record this number of departures, it is not sufficient to know the number of customers at Q at the beginning and end of a visit. To keep track of the number of departures, we will focus on the transient probabilities $p_{hk}^{(n)}(t)$, which are defined as follows:

$$p_{hk}^{(n)}(t) := \begin{cases} \mathbb{P}(x_t = k, D(t) = n \mid x_0 = h) & h, k, n = 0, 1, \dots, \\ 0 & \text{otherwise,} \end{cases}$$

where x_t refers to the number of customers at Q at time t and $D(t)$ refers to the number of departures from Q until time t . We relate these probabilities to the transient probabilities for the standard M/G/1 queue, which we denote by $P_{hj}^{(n)}(t)$. These time-dependent conditional probabilities, which also incorporate the number of departures until time t , are defined for $n = 1, 2, \dots, h, j = 0, 1, \dots$, and $t > 0$ as [7, page 239]

$$P_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t \mid z_{(0)} = h),$$

where $z_{(n)}$ refers to the number of customers left behind by the n th departure, and r'_n refers to the epoch of the n th departure. Further, it is assumed that at time $t = 0$ the 0th customer left the queue. We consider the function $\pi_h(r, s, y)$, which is defined in terms of $P_{hj}^{(n)}(t)$ as follows:

$$\pi_h(r, s, y) := \sum_{n=1}^{\infty} y^n \sum_{j=0}^{\infty} r^j \int_0^{\infty} e^{-st} dP_{hj}^{(n)}(t), \quad h = 0, 1, \dots,$$

and which is explicitly provided in [7, page 240] for $h = 0, 1, \dots$, as

$$\pi_h(r, s, y) = \frac{y \cdot \tilde{X}(s + \lambda(1 - r))}{r - y \cdot \tilde{X}(s + \lambda(1 - r))} \cdot \left\{ r^h - \frac{\lambda(1 - r) + s}{\lambda(1 - \hat{\mu}(s, y)) + s} \cdot \hat{\mu}^h(s, y) \right\}, \quad (2)$$

where $\hat{\mu}(s, y)$ is the root x with the smallest absolute value less than one of $x = y \cdot \tilde{X}(s + \lambda(1 - x))$. Note that $\hat{\mu}(s, y)$ is the joint generating function of the busy period and the number of customers served during this period; see e.g. [7, page 250].

To take advantage of this explicit result, we will first present an explicit expression for the transient probabilities $p_{hk}^{(n)}(t)$ in terms of $P_{hj}^{(n)}(t)$. For convenience, we define

$$F_k^{(0)}(t) = \mathbf{1}_{\{k=0\}}\mathbb{P}(A(t) = 0, I > t) + \mathbf{1}_{\{k \geq 1\}}\mathbb{P}(A(t) = k, I + X > t), \quad k = 0, 1, \dots,$$

$$F_k(t) = \mathbb{P}(A(t) = k, X > t), \quad k = 0, 1, \dots$$

That is, $F_k(t)$ refers to k exogenous arrivals to Q during period $[0, t]$ which is shorter than the service time X of the job that started service at time $t = 0$, that is, assuming a non-empty queue at $t = 0$ and the job under service is not yet finished. In the special case of an empty queue at $t = 0$ (see $F_k^{(0)}(t)$), we need to account for the fact that first an arrival should occur before any service may start. Then, we can relate $p_{hk}^{(n)}(t)$ to $P_{hj}^{(n)}(t)$ for $n = 1, 2, \dots, h, k = 0, 1, \dots$, and $t > 0$ as follows.

Lemma 3. *We have*

$$p_{hk}^{(n)}(t) = \int_{u=0}^t F_k^{(0)}(t-u) dP_{h0}^{(n)}(u) + \sum_{j=1}^k \int_{u=0}^t F_{k-j}(t-u) dP_{hj}^{(n)}(u).$$

To retrieve the terms $\pi_h(r, s, y)$, we first take the LST of $p_{hk}^{(n)}(t)$ (see Remark 2 below). Next, we take the generating function of this expression with respect to the number of customers at the end of a server visit. Note that our interest here is specifically in this number rather than in the number at the time of the n th departure, since the server only leaves upon expiration of the timer. In a final step, we take the generating function with respect to the number of departures until time t to obtain an expression for $p_{hk}^{(n)}(t)$ in terms of $\pi_h(r, s, y)$. These consecutive steps provide us with the following result for $h = 0, 1, \dots$.

Lemma 4. *We have*

$$\sum_{n=1}^{\infty} y^n \sum_{k=0}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dp_{hk}^{(n)}(t)$$

$$= \frac{s}{\lambda(1-r) + s} \cdot \frac{\lambda(1-r \cdot \tilde{X}(\lambda(1-r) + s)) + s}{\lambda + s} \cdot \pi_{h0}(s, y)$$

$$+ \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)) \cdot (\pi_h(r, s, y) - \pi_{h0}(s, y)), \tag{3}$$

where the terms $\pi_{h0}(s, y)$ are given by (see [7, page 240])

$$\pi_{00}(s, y) = \frac{\lambda}{\lambda(1 - \hat{\mu}(s, y)) + s} \cdot \hat{\mu}(s, y), \tag{4}$$

$$\pi_{h0}(s, y) = \frac{\lambda + s}{\lambda(1 - \hat{\mu}(s, y)) + s} \cdot \hat{\mu}^h(s, y), \quad h = 1, 2, \dots \tag{5}$$

The right-hand side of (3) can be interpreted as follows. The first part refers to the case that upon the n th departure zero customers are left behind, while the second part refers to a strictly positive number left behind by the n th departing customer. Moreover, the second part can be

decomposed into two independent components: $\pi_h(r, s, y) - \pi_{h0}(s, y)$ accounts for the queue-length evolution until the n th departure and the other component accounts for the queue-length evolution during the final, interrupted service. A similar reasoning holds for the first part.

Thus, we have related the transient probabilities of our interest to known results for the M/G/1 queue. To incorporate these results into the polling model, we refer to a specific queue Q_i by adding an index i to the generic variables. Next, by unconditioning on the system state at the start of a visit and incorporating the expressions above into the definition of $\beta^i(\mathbf{z})$, we obtain the main result of this section for the p.g.f. of the joint queue length at the end of a server visit under the P-TL discipline.

Theorem 1. (Pure exponential time-limited discipline.) *We have*

$$\beta^i(\mathbf{z}) = d_1^P(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_2^P(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*),$$

where

$$\begin{aligned} d_1^P(\mathbf{z}) &= \frac{\xi_i}{z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*)} \cdot \frac{z_i \cdot (1 - \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*))}{\lambda_i(1 - z_i) + \xi_i^*}, \\ d_2^P(\mathbf{z}) &= d_1^P(\mathbf{z}) + \frac{\xi_i}{z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*)} \cdot \frac{(z_i - r^i(\mathbf{z})) \cdot \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*)}{\lambda_i(1 - \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))) + \xi_i^*}, \\ \xi_i^* &= \xi_i + \sum_{j \neq i} \lambda_j(1 - z_j), \\ \alpha^i(\mathbf{z}_i^*) &:= \alpha^i(z_1, \dots, z_{i-1}, \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z})), z_{i+1}, \dots, z_M). \end{aligned} \tag{6}$$

We refer to Section 6 for a detailed interpretation of the results shown in Theorem 1.

Remark 1. (Phase-type service times.) We note that Theorem 1 generalizes the result for the special case $r^i(\mathbf{z}) = 1$ (i.e. no customer routing) given in [2] for the case of phase-type service times and in [1] for the exponential service times.

Remark 2. (Exponential time limit.) The step of taking the LST of $p_{hk}^{(n)}(t)$ in fact corresponds to unconditioning over the exponentially distributed visit time. This shows that our assumption on the visit time plays a crucial role in the analysis.

Remark 3. The results of Theorem 1 can be used in a numerical framework to compute the joint queue-length distribution. We refer the reader to [2, Section 5] for details of this computational algorithm which uses an iterative method. Moreover, in [2, Section 6] the algorithm’s implementation is described in detail and a test of its performance from a computational point of view is presented. Finally, in [2, Section 7] a comparison of the algorithm with the numerical approach proposed in [22] is shown and the method’s benefits are reported.

4. Analysis of the exhaustive time-limited service discipline

Let us next consider the E-TL discipline. Under this discipline the server will depart from a queue not only when the time limit is reached but also when the queue becomes empty. Again, we will derive an expression for $\beta^i(\mathbf{z})$, the p.g.f. of the number of customers at all queues at the instant that the server leaves Q_i . This will be done in terms of the number of customers

present at the start of the visit, $\alpha^i(\mathbf{z})$. As in the previous section, we present here only the main analytical steps and the final result. The proofs are given in Appendix C.

Under the E-TL discipline, the server may leave a queue for two reasons, namely, the server departs due to the queue being empty or due to the timer expiring. Let {empty} and {timer} denote the corresponding server events. Recall that \mathbf{N}_i^s and \mathbf{N}_i^e denote the multi-dimensional random variable of the number of customers at all queues at the start and the end of a server visit to Q_i , respectively. The p.g.f. of \mathbf{N}_i^e , $\beta^i(\mathbf{z})$, can be decomposed into two parts depending on the reason for a server departure, as the server departs only if the queue is empty or if the timer expires. Moreover, these events are readily seen to be mutually exclusive (the service-time distribution and timer distribution are both continuous distributions, so that the probability of the given events occurring simultaneously is zero). Hence, the p.g.f. of the number of customers at the end of a server visit period to Q_i satisfies

$$\beta^i(\mathbf{z}) = \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e}] = \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{empty}\}}] + \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}}].$$

Next, in Sections 4.1 and 4.2, we will derive the conditional p.g.f.s

$$\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{empty}\}} \mid \mathbf{N}_i^s = \mathbf{n}], \quad \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}],$$

where \mathbf{n} denotes the vector (n_1, \dots, n_M) . Finally, we will uncondition these expressions to get our main result in Section 4.3.

4.1 $\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{empty}\}} \mid \mathbf{N}_i^s = \mathbf{n}]$

If the {empty} event occurs, the queue may be empty upon arrival of the server or become empty upon departure of a customer. If the server finds an empty queue upon arrival, then clearly $\mathbf{N}_i^e = \mathbf{N}_i^s$. Else, if the queue is non-empty upon server arrival, then the evolution of queue-length process during the visit is closely related to the length of a busy period in a standard M/G/1 queue. This is formalized in the following lemma.

Lemma 5. *The joint conditional p.g.f. of the number of customers at the end of a visit period to Q_i and the server departure due to the queue being empty satisfies*

$$\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{empty}\}} \mid \mathbf{N}_i^s = \mathbf{n}] = \hat{\mu}_i^{n_i}(\xi_i^*, r^i(\mathbf{z})) \cdot \prod_{j \neq i} z_j^{n_j}, \tag{7}$$

where

$$\xi_i^* = \xi_i + \sum_{j \neq i} \lambda_j (1 - z_j).$$

For the case when the server departs due to an empty Q_i , Lemma 5 shows that every customer waiting in Q_i at the beginning of server visit is replaced in an independent and identically distributed manner. These replacement jobs do not only concern Q_i but all other queues as well. This latter is captured via the term $r^i(\mathbf{z})$, which is defined as the p.g.f. of the number of arrivals to all queues generated by a single departing customer at Q_i .

4.2 $\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}]$

If the {timer} event occurs, the queue must be non-empty upon arrival of the server, and it remains non-empty during the course of the visit and is still non-empty at the expiration of the timer. The analysis of this case builds on the work of Cohen for the transient analysis of the

M/G/1 queue. However, contrary to the analysis for the P-TL discipline, we cannot directly apply the formulae derived in [7]. This is due to the fact that we need specifically to account for not entering the state with zero customers at Q_i during the course of a server visit. Below, we state the transient probabilities of interest and several related expressions. Next, using these expressions, we will derive $\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}]$.

We first consider the conditional joint queue-length distribution at time $t > 0$ given an initial number of customers at time $t = 0$ and given that the server is at an arbitrary queue Q . It is good to note that during a server visit to Q the queue-length process at the other queues is simply a pure-birth process. Hence, we neglect the other queues for the moment and concentrate on the marginal queue-length probabilities for Q , denoted by $q_{hk}^{(n)}(t)$, which we define as

$$q_{hk}^{(n)}(t) := \begin{cases} \mathbb{P}(x_t = k, D(t) = n, x_v > 0, 0 < v < t \mid x_0 = h) & n = 0, 1, \dots, h, k = 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

For convenience, let us recall the definition of the probabilities $P_{hj}^{(n)}(t)$, for $n = 1, 2, \dots, h, j = 0, 1, \dots$ and $t > 0$,

$$P_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t \mid z_{(0)} = h).$$

Analogously, we define $R_{hj}^{(n)}(t)$, for $h, j, n = 1, 2, \dots$, and $t > 0$,

$$R_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t, z_{(k)} > 0, 0 < k < n \mid z_{(0)} = h),$$

where it is assumed that at time $t = 0$ a new service starts. We note that $R_{hj}^{(n)}(t)$ is only defined for $h, j = 1, 2, \dots$. This is due to the fact that the event of a server arriving at an empty queue (i.e. $h = 0$) and the event of the n th customer leaving an empty queue behind (i.e. $j = 0$) are not considered as {timer} events, but always as {empty} events.

We consider the function $\gamma_h(r, s, y)$, which is defined in terms of $R_{hj}^{(n)}(t)$ as

$$\gamma_h(r, s, y) := \sum_{n=1}^{\infty} y^n \sum_{j=1}^{\infty} r^j \int_0^{\infty} e^{-st} dR_{hj}^{(n)}(t), \quad h = 1, 2, \dots,$$

and which is explicitly given (see Appendix A for the derivation) for $h = 1, 2, \dots$, as

$$\gamma_h(r, s, y) = \frac{r(-\hat{\mu}^h(s, y) + y \cdot \tilde{X}(\lambda(1-r) + s) \cdot r^{h-1})}{r - y \cdot \tilde{X}(\lambda(1-r) + s)}.$$

Analogous to the approach in the previous section, we intend to utilize the explicit expressions for $\gamma_h(r, s, y)$. To this end, we will start by relating the transient probabilities $q_{hk}^{(n)}(t)$ to the time-dependent probabilities $R_{hj}^{(n)}(t)$ at embedded epochs of service completion. For convenience, we recall that

$$F_k(t) = \mathbb{P}(A(t) = k, X > t), \quad k = 0, 1, \dots,$$

that is, $F_k(t)$ refers to the number of arrivals to Q during a period which is shorter than a service time X . The specific relation between $q_{hk}^{(n)}(t)$ and $R_{hj}^{(n)}(t)$ is then given in the following lemma.

Lemma 6. *We have*

$$q_{hk}^{(n)}(t) = \int_{u=0}^t \sum_{j=1}^k F_{k-j}(t-u) dR_{hj}^{(n)}(u), \quad n = 1, 2, \dots, h, k = 1, 2, \dots$$

Again, to obtain the terms $\gamma_h(r, s, y)$, we take the LST of $q_{hk}^{(n)}(t)$ (see Remark 2). Next, we take the generating function with respect to the number of customers at the end of the server visit of the resulting expression, and finally we take the generating function with respect to the number of departures. Hence, we obtain the following result.

Lemma 7. *We have*

$$\sum_{n=1}^{\infty} y^n \sum_{k=1}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dq_{hk}^{(n)}(t) = \gamma_h(r, s, y) \cdot \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)), \quad h = 1, 2, \dots \tag{8}$$

The right-hand side of (8) can be recognized as a convolution of two independent parts. The first part, $\gamma_h(r, s, y)$, refers to the queue length at the instant of the final (successful) service completion during the visit, while the other part refers to the number of arrivals during an interrupted service.

Next, we present the explicit expression for the joint conditional p.g.f. of the number of customers at all queues at the end of a visit to a specific queue Q_i when the server departure is due to the timer expiration. The condition is on the number of customers present at the start of the visit.

Lemma 8. *We have*

$$\mathbb{E}[\mathbf{z}^{N_i^c} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}] = \frac{\xi_i \cdot z_i \cdot (1 - \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*)) (z_i^{n_i} - \hat{\mu}_i^{n_i}(\xi_i^*, r^i(\mathbf{z})))}{[\lambda_i(1 - z_i) + \xi_i^*] \cdot [z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*)]} \cdot \prod_{j \neq i} z_j^{n_j}, \tag{9}$$

where

$$\xi_i^* = \xi_i + \sum_{j \neq i} \lambda_j(1 - z_j).$$

Compared to Lemma 5, expression (9) is much more complex, which is due to the fact that when the timer expires the remaining jobs at Q_i are not replaced in an independent and identical manner.

4-3 $\mathbb{E}[\mathbf{z}^{N_i^c}]$

Combining the two conditional results of (7) and (9), we obtain our main result of this section for the E-TL service discipline.

Theorem 2. (Exhaustive exponential time-limited discipline). *We have*

$$\beta^i(\mathbf{z}) = d_1^E(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_2^E(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*), \tag{10}$$

where

$$\begin{aligned} d_1^E(\mathbf{z}) &= d_1^P(\mathbf{z}), \\ d_2^E(\mathbf{z}) &= 1, \\ \alpha^i(\mathbf{z}_i^*) &= \alpha^i(z_1, \dots, z_{i-1}, \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z})), z_{i+1}, \dots, z_M), \\ \xi_i^* &= \xi_i + \sum_{j \neq i} \lambda_j(1 - z_j). \end{aligned}$$

Observe that (10) generalizes the result for the special case $r^i(\mathbf{z}) = 1$ (i.e. no customer routing) given in [11]. We refer to Section 6 for a detailed interpretation of the results shown in Theorem 2. To find the p.g.f.s of joint queue-length distribution based on results in Theorem 2, we refer to Remark 3.

Remark 4. (*Phase-type service times.*) Theorem 2 generalizes the result for the special case $r^i(\mathbf{z}) = 1$ (i.e. no customer routing) given in [2] for the case of phase-type service times and in [1] for the exponential service times.

Remark 5. (*Exhaustive service discipline.*) In the limit case of $\xi_i \downarrow 0$ the time limit is of infinite length. Hence, in this case (assuming a stable queue), the server will always depart due to Q_i being empty. It can readily be found that for $\xi_i \downarrow 0$ and $r^i(\mathbf{z}) = 1$ the following expression for $\beta^i(\mathbf{z})$ is obtained:

$$\beta^i(\mathbf{z}) = \alpha^i(\mathbf{z}_i^*),$$

where

$$\alpha^i(\mathbf{z}_i^*) := \alpha^i\left(z_1, \dots, z_{i-1}, \hat{\mu}_i\left(\sum_{j \neq i} \lambda_j(1 - z_j), 1\right), z_{i+1}, \dots, z_M\right).$$

This result matches the well-known result for the exhaustive service discipline [14,18].

5. Model extensions

The analysis of the basic polling model may be extended in various directions. Extending the basic model increases the range of applications that can be modeled. Moreover, such extensions are also interesting from a theoretical point of view. The specific extensions that we discuss here are the following: batch Poisson arrivals of customers and Markovian polling of the server.

5.1 Batch Poisson arrivals

In this subsection we concentrate on the time-limited polling systems with batch Poisson arrivals. The analysis of these batch arrival systems closely follows the analysis of single arrival systems. Therefore we will only present the main steps and the final results.

The analysis of the time-limited polling system is based on the transient analysis of the M/G/1 queue, for which closed-form expressions for the joint LST for the queue length and number of departures at time t is available; see e.g. [7, pages 239–240]. In the case of batch arrivals, we need the transient analysis for the $M^X/G/1$ queue, that is a single server queue with compound Poisson arrivals. The transient analysis was partially done in [7, pages 239–240]. Below, we shall derive the required expressions for our analysis.

The batch sizes are assumed to be mutually independent and independent of the arrival and service processes. Let λ denote the arrival rate of batches and $\hat{\psi}(\cdot)$ the generating function of the batch size. Recall that the generating function of the service time of an individual customer is denoted by $\tilde{X}(\cdot)$

Let $z_{(n)}$ denote the number of customers left behind by the n th departing customer after time 0 and let r'_n be its departure time. We assume that at time $t = 0$ customer 0 departs. Let $R_{hj}^{(n)}(t)$ denote the following probability:

$$R_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t, z_{(k)} > 0, 0 < k < n \mid z_{(0)} = h) \quad \text{for } h = 1, 2, \dots, j = \max(0, h - n).$$

Lemma 9. *The generating function of $R_{hj}^{(n)}(t)$ is*

$$\gamma_h(r, s, y) = \frac{r(-\hat{\mu}^h(s, y) + y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))) \cdot r^{h-1}}{r - y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))},$$

for $h = 1, 2, \dots$

The result in Lemma 9 can be seen as an extension of Theorem 5 in Appendix A for the M/G/1 queue. Now we have the ingredients to relate $\beta^i(\mathbf{z})$ and $\alpha^i(\mathbf{z})$ for batch Poisson arrivals. For the system with a P-TL service discipline with batch arrivals, after a similar analysis to Section 3 and Appendix B, we obtain the following result (*cf.* Theorem 1).

Theorem 3. (Pure exponential time-limited discipline with batch arrivals.) *We have*

$$\beta^i(\mathbf{z}) = d_{B1}^P(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_{B2}^P(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*),$$

where

$$d_{B1}^P(\mathbf{z}) = \frac{\xi_i}{z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - \hat{\psi}_i(z_i)) + \xi_i^*)} \cdot \frac{z_i \cdot (1 - \tilde{X}_i(\lambda_i(1 - \hat{\psi}_i(z_i)) + \xi_i^*))}{\lambda_i(1 - \hat{\psi}_i(z_i)) + \xi_i^*},$$

$$d_{B2}^P(\mathbf{z}) = d_{B1}^P(\mathbf{z}) + \frac{\xi_i}{z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - \hat{\psi}_i(z_i)) + \xi_i^*)} \cdot \frac{(z_i - r^i(\mathbf{z})) \cdot \tilde{X}_i(\lambda_i(1 - \hat{\psi}_i(z_i)) + \xi_i^*)}{\lambda_i(1 - \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))) + \xi_i^*},$$

$$\alpha^i(\mathbf{z}_i^*) = \alpha^i(z_1, \dots, z_{i-1}, \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z})), z_{i+1}, \dots, z_M),$$

$$\xi_i^* = \xi_i + \sum_{j \neq i} \lambda_j(1 - \hat{\psi}_j(z_j)).$$

For the system with an E-TL service discipline with batch Poisson arrivals, after a similar analysis to Section 4 and Appendix C, we obtain the following result (*cf.* Theorem 2).

Theorem 4. (Exhaustive exponential time-limited discipline with batch arrivals.) *We have*

$$\beta^i(\mathbf{z}) = d_{B1}^P(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + \alpha^i(\mathbf{z}_i^*),$$

where $d_{B1}^P(\mathbf{z})$ is given in Theorem 3.

5.2 Markovian polling of the server

We have assumed until now that the server polls the queues according to a fixed cyclic schedule. To allow for more general polling schedules, the routing (or polling) of the server is taken to follow a Markovian pattern (see [5] and [17]). Let $s_{i,j} \geq 0, j = 1, \dots, M$, denote the probability upon departure from queue i that the next queue that will be visited is j . This next queue could also be i again. This feature is particularly meaningful in the case of non-zero switch-over times, whereas in the case of zero switch-over times we may simply set $s_{i,i} = 0, i = 1, \dots, M$.

Let us describe the embedded process of visits to the queues (thus neglecting switch-over times for the moment) by a discrete-time Markov chain $X_n \in \{1, \dots, M\}, n \geq 0$, driven by the transition probability matrix $S = \{s_{i,j}\}_{i,j=1,\dots,M}$. We assume that this Markov chain has an equilibrium distribution which we denote by $\tau_i, i = 1, \dots, M$. Our interest is in the fraction

of time that the Markov chain is in a state i (denoted by κ_i) and also in the fraction of times that the process is switching from state i to state j (denoted by $\Delta_{i,j}$). To enrich the analysis here we include the switch-over time from Q_i to Q_j and assume it is independent of all other variables, for example arrival times, service times, and server visit times. Let us define the cycle time of Q_i by C_i as the time between two consecutive polling instants of the server at Q_i . This means that a cycle C_i comprises exactly one visit time to Q_i . The mean cycle time can then be expressed as a weighted sum of mean visit and mean switch-over times as follows:

$$\mathbb{E}[C_i] = \sum_k \frac{\tau_k}{\tau_i} \left(\mathbb{E}[Y_k] + \sum_j s_{k,j} \cdot c_{k,j} \right), \quad i = 1, \dots, M,$$

where $\mathbb{E}[Y_k]$ denotes the mean visit time to queue k and $c_{k,j}$ the mean switch-over time from Q_k to Q_j . Note that the ratio τ_k/τ_i equals the mean number of visits to queue k per visit to i . There exists a simple relation between the mean cycle times of the different queues in the sense that the product $\mathbb{E}[C_i] \cdot \tau_i$ is constant for all queues. This relation immediately explains why, under Markovian polling, in contrast to the cyclic polling case (for which $\tau_i = 1/M, i = 1, \dots, M$), the mean cycle times per queue are not necessarily equal for all queues.

It readily follows for $\kappa_i, i = 1, \dots, M$, which is in fact the long-term fraction of time the server is available at Q_i , that

$$\kappa_i = \frac{\mathbb{E}[Y_i]}{\mathbb{E}[C_i]} = \frac{\mathbb{E}[Y_i]}{\sum_k (\tau_k/\tau_i)(\mathbb{E}[Y_k] + \sum_j s_{k,j} \cdot c_{k,j})} = \frac{\tau_i \mathbb{E}[Y_i]}{\sum_k \tau_k (\mathbb{E}[Y_k] + \sum_j s_{k,j} \cdot c_{k,j})}, \quad (11)$$

where it should be noted that the denominator in the last part of (11) does not depend on i . Similarly, for $\Delta_{i,j}$ we find that

$$\Delta_{i,j} = \frac{\tau_i \cdot s_{i,j} \cdot c_{i,j}}{\sum_k \tau_k (\mathbb{E}[Y_k] + \sum_j s_{k,j} \cdot c_{k,j})}.$$

Again, the adjustments in our modeling framework for the joint queue-length probabilities are relatively simple. We note that routing of the server solely plays a role in the relation between the queue length at a visit start instant and the queue length at the preceding visit completion instant. Since the preceding queue that was served is now random, the expression for $\alpha^i(\mathbf{z})$ becomes

$$\alpha^i(\mathbf{z}) = \sum_j q_{j,i} \hat{C}_{j,i}(\mathbf{z}) \beta^j(\mathbf{z}),$$

where $\hat{C}_{j,i}(\mathbf{z})$ is the joint generating function of the total number of Poisson arrivals to all the queues during $C_{j,i}$, the switch-over time from Q_j to Q_i . Here $q_{j,i}$ is the probability that the preceding queue was Q_j , given that the server is at Q_i , and is given by the transition probability of the time reversed Markov chain (see e.g. [15])

$$q_{j,i} = \frac{\tau_j \cdot s_{j,i}}{\tau_i}.$$

Finally, we conclude that for the P-TL discipline the p.g.f. of the steady-state joint queue-length probabilities is

$$P(\mathbf{z}) = \sum_{i=1}^M \left(\beta^i(\mathbf{z}) \cdot \kappa_i + \sum_{j=1}^M \hat{C}_{i,j}^R(\mathbf{z}) \cdot \Delta_{i,j} \right),$$

where

$$\hat{C}_{i,j}^R(\mathbf{z}) = \beta^i(\mathbf{z}) \cdot \frac{1 - \tilde{C}_{i,j}(\sum_k \lambda_k(1 - z_k))}{c_{i,j} \cdot \sum_k \lambda_k(1 - z_k)}.$$

Note that for the E-TL discipline we can just derive the expressions of the p.g.f. of the steady-state joint queue-length process at polling instances. Based on these results, it is possible to find the p.g.f. of the marginal queue-length distribution in Q_i at an arbitrary time; see e.g. [11].

6. Concluding remarks

The final results for the P-TL discipline and the E-TL discipline have a similar form. More specifically, these results can be written as follows:

$$\text{P-TL: } \beta^i(\mathbf{z}) = d_1^P(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_2^P(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*), \tag{12}$$

$$\text{E-TL: } \beta^i(\mathbf{z}) = d_1^E(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)) + d_2^E(\mathbf{z}) \cdot \alpha^i(\mathbf{z}_i^*), \tag{13}$$

where $d_1^E(\mathbf{z}) = d_1^P(\mathbf{z})$ is given in (6), $d_2^E(\mathbf{z}) = 1$ and $d_2^P(\mathbf{z})$ is given in (6).

Equations (12) and (13) may be interpreted as follows. Consider a visit of the server to Q_i . Regarding the timer, it may occur that (i) the timer expires before Q_i becomes empty for the first time, or (ii) the timer expires only after Q_i becomes empty for the first time. It is readily seen that the queue-length process is identical for both service disciplines in the case (i). This is reflected in the similarity of the terms $d_1^P(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*))$ and $d_1^E(\mathbf{z}) \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*))$. However, in case (ii), the queue-length process is different for the disciplines. Under the exhaustive time-limited discipline, the server immediately leaves upon the queue becoming empty; say this occurs at time t_0 . Under the pure time-limited discipline, at time t_0 the server will remain at the queue and a sequence of idle and busy periods will follow until eventually the timer expires. This latter contribution (after t_0) to the queue-length process is represented in the term $d_2^P(\mathbf{z})$.

The function $d_2^P(\mathbf{z})$ reflects the p.g.f. of the number of customers at all queues at the end of a server visit process, which runs for an exponential amount of time and which starts from an empty queue. This function can be analyzed as follows. First, observe that the timer will interrupt the visit process during either an idle or a busy period. Second, observe that this process is regenerative in the sense that if the timer does not expire before the end of the first busy period, then the process starts like new at that specific time instant. Recall that I_i denotes the length of an idle period at Q_i and Y_i denotes the exponential visit time of the server to Q_i . Further, let U_i denote the length of a busy period at Q_i starting with a single customer. Then we may write the following relation for $d_2^P(\mathbf{z})$:

$$\begin{aligned} d_2^P(\mathbf{z}) &= \mathbb{E}[\mathbf{z}^{\mathbf{N}_i(Y_i)} \mathbf{1}_{\{I_i > Y_i\}} \mid \mathbf{N}_i(0) = \mathbf{n}, N_{i,i}(0) = 0] \\ &\quad + \mathbb{E}[\mathbf{z}^{\mathbf{N}_i(Y_i)} \mathbf{1}_{\{I_i \leq Y_i, I_i + U_i > Y_i\}} \mid \mathbf{N}_i(0) = \mathbf{n}, N_{i,i}(0) = 0] \\ &\quad + \mathbb{E}[\mathbf{z}^{\mathbf{N}_i(I_i + U_i)} \mathbf{1}_{\{I_i + U_i \leq Y_i\}} \mid \mathbf{N}_i(0) = \mathbf{n}, N_{i,i}(0) = 0] \cdot d_2^P(\mathbf{z}) \\ &= \frac{\xi_i}{\lambda_i + \xi_i^*} + \frac{\lambda_i}{\lambda_i + \xi_i^*} \\ &\quad \times \mathbb{E}[\mathbf{z}^{\mathbf{N}_i(Y_i)} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i(0) = (n_1, \dots, n_{i-1}, 1, n_{i+1}, \dots, n_M)] \\ &\quad + \frac{\lambda_i}{\lambda_i + \xi_i^*} \cdot \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z})) \cdot d_2^P(\mathbf{z}), \end{aligned}$$

where $\mathbb{E}[\mathbf{z}^{\mathbf{N}_i(Y_i)} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i(0) = \mathbf{n}]$ is provided in the analysis of the exhaustive time-limited discipline (see Lemma 8). Then, inserting this result of Lemma 8 and reorganizing the terms appropriately, we obtain $d_2^P(\mathbf{z})$ as in (6).

This result indeed confirms the interpretation given above. We remarked that for the common exhaustive discipline (E), i.e. the time-limited version with the time limit set to infinity, the term $d_1^E(\mathbf{z})$ vanishes, that is,

$$E: \beta^i(\mathbf{z}) = \alpha^i(\mathbf{z}_i^*),$$

In this paper we have studied two time-limited service disciplines for polling systems, namely, the pure and the exhaustive exponential time-limited discipline with preemptive service. Specifically, we have obtained relations between the p.g.f.s of the joint queue-length distribution at visit beginning and visit completion epochs. To this end, we have used both known (see Section 3) and novel results (see Section 4) for the transient behavior of the M/G/1 queue. Our final expressions for the key relation, (1), extend earlier results for the pure limited [2] and the exhaustive time-limited discipline [11] by incorporating new model features such as customers routeing among queues; see Section 5 for the other model features. These relations can be used to obtain the joint queue-length distribution at these embedded epochs, for example using the framework presented in [2]. These key relations not only provide us with a more elegant result but have also been shown to significantly alleviate the computational efforts required to compute the joint queue-length distribution of the polling system along the framework of [9]. Finally, we have been able to give a clear interpretation of the final expressions for the key relations, the discussion after (12) and (13).

Appendix A: Transient analysis of the M/G/1 queue during a busy period

In this section we analyze the transient behavior of the M/G/1 queue during a busy period. We follow an approach similar to that used by Cohen [7] to study the transient behavior of the full queue-length process of the M/G/1 queue. To this end, we consider a single queue served by a single server. Customers arrive at the queue according to a Poisson process with rate λ . The service requirement X of a customer is generally distributed.

Our interest is in the queue-length process during a busy period with some initial number of customers. Moreover, we keep track of the number of departures until time t . Therefore, as in the transient transition probabilities $P_{hj}^{(n)}(t)$ that were defined in [7, page 239], we define the transient probabilities $R_{hj}^{(n)}(t)$ which specifically account for the fact that the system is non-empty from time 0 up to time t . More precisely, the transient probabilities $R_{hj}^{(n)}(t)$ are defined for $h, j, n = 1, 2, \dots$, and $t > 0$ as

$$R_{hj}^{(n)}(t) := \mathbb{P}(z(n) = j, r'_n \leq t, z(k) > 0, 0 < k < n \mid z(0) = h),$$

where it is assumed that at time $t = 0$ a new service starts. Note that $R_{hj}^{(n)}(t)$ is only defined for $h, j \geq 1$. Our objective is to find an explicit expression for $\gamma_h(r, s, y)$ which is defined as

$$\begin{aligned} \gamma_h(r, s, y) &:= \sum_{n=1}^{\infty} y^n \mathbb{E}(r^{z(n)} e^{-sr'_n} \mathbf{1}(z(k) > 0, 0 < k < n)), \\ &= \sum_{n=1}^{\infty} y^n \sum_{j=1}^{\infty} r^j \int_0^{\infty} e^{-st} dR_{hj}^{(n)}(t), \quad h = 1, 2, \dots \end{aligned}$$

From the definition of $R_{hj}^{(n)}(t)$ it follows immediately that

$$R_{lj}^{(1)}(t) = \int_{\tau=0}^t e^{-\lambda\tau} \frac{(\lambda\tau)^j}{j!} dX(\tau), \quad j = 1, 2, \dots,$$

$$R_{hj}^{(1)}(t) = \int_{\tau=0}^t e^{-\lambda\tau} \frac{(\lambda\tau)^{j+1-h}}{(j+1-h)!} dX(\tau), \quad j = h-1, h, \dots, h=2, 3, \dots,$$

$$R_{hj}^{(1)}(t) = 0, \quad \text{otherwise.}$$

Also, analogously to (4.20) of [7, page 239], we have the following recursive relation for $R_{hj}^{(n)}(t)$ for $t > 0, h, j = 1, 2, \dots, n = 2, 3, \dots$:

$$R_{hj}^{(n)}(t) = \sum_{l=1}^{\infty} \int_{u=0}^t R_{hl}^{(n-1)}(t-u) du R_{lj}^{(1)}(u). \tag{14}$$

The following definitions will be used below:

$$\gamma_{hj}^{(n)}(s) := \int_0^{\infty} e^{-st} dR_{hj}^{(n)}(t), \quad h, j, n = 1, 2, \dots,$$

$$\gamma_h^{(n)}(r, s) := \sum_{j=1}^{\infty} r^j \gamma_{hj}^{(n)}(s), \quad h, n = 1, 2, \dots,$$

$$\gamma_{hj}(s, y) := \sum_{n=1}^{\infty} y^n \gamma_{hj}^{(n)}(s), \quad h, j = 1, 2, \dots,$$

$$\gamma_h(r, s, y) := \sum_{n=1}^{\infty} y^n \gamma_h^{(n)}(r, s), \quad h = 1, 2, \dots$$

As an immediate consequence of (14), we obtain the following result.

Lemma 10. *We have*

$$\gamma_h^{(n)}(r, s) = \sum_{l=1}^{\infty} \gamma_{hl}^{(n-1)}(s) \cdot \gamma_l^{(1)}(r, s), \quad h = 1, 2, \dots, n = 2, 3, \dots \tag{15}$$

The final term of the right-hand side of (15), $\gamma_l^{(1)}(r, s)$, refers to the number of arrivals during a service time starting with l customers. We have to distinguish between starting with one or with two or more customers, since in the former case the queue might be empty upon service completion and this situation should be excluded as we restrict the analysis to the evolution within a busy period. A closed-form expression for this term is then given in the following lemma.

Lemma 11. *We have*

$$\gamma_1^{(1)}(r, s) = \tilde{X}(\lambda(1-r) + s) - \tilde{X}(\lambda + s),$$

and for $h \geq 2$,

$$\gamma_h^{(1)}(r, s) = r^{h-1} \cdot \tilde{X}(\lambda(1-r) + s).$$

Proof. Let us first consider the case $h \geq 2$:

$$\begin{aligned}
 \gamma_h^{(1)}(r, s) &= \sum_{j=1}^{\infty} r^j \gamma_{hj}^{(1)}(s) \\
 &= \sum_{j=h-1}^{\infty} r^j \gamma_{hj}^{(1)}(s) \\
 &= \sum_{j=h-1}^{\infty} r^j \int_{t=0}^{\infty} e^{-st} dR_{hj}^{(1)}(t) \\
 &= \int_{t=0}^{\infty} s e^{-st} \sum_{j=h-1}^{\infty} r^j R_{hj}^{(1)}(t) dt \\
 &= \int_{t=0}^{\infty} s e^{-st} \int_{\tau=0}^t e^{-\lambda\tau} \sum_{j=h-1}^{\infty} r^{h-1} \cdot \frac{(r\lambda\tau)^{j+1-h}}{(j+1-h)!} dX(\tau) dt \\
 &= r^{h-1} \int_{\tau=0}^{\infty} e^{-\lambda\tau(1-r)} \int_{t=\tau}^{\infty} s \cdot e^{-st} dt dX(\tau) \\
 &= r^{h-1} \cdot \tilde{X}(\lambda(1-r) + s).
 \end{aligned}$$

For $h = 1$, we should have at least one arrival before the first departure, otherwise the queue would become empty. Hence, in the derivation of $\gamma_1^{(1)}(r, s)$, we do not encounter the complete power series representation of the exponential function, so that the final expression will consist of two parts. More precisely,

$$\begin{aligned}
 \gamma_1^{(1)}(r, s) &= \sum_{j=1}^{\infty} r^j \gamma_{hj}^{(n)}(s) \\
 &= \int_{t=0}^{\infty} s e^{-st} \int_{\tau=0}^t e^{-\lambda\tau} \sum_{j=1}^{\infty} \frac{(r\lambda\tau)^j}{j!} dX(\tau) dt \\
 &= \int_{\tau=0}^{\infty} e^{-\lambda\tau} \cdot (e^{-\lambda\tau r} - 1) \int_{t=\tau}^{\infty} s e^{-st} dt dX(\tau) \\
 &= \tilde{X}(\lambda(1-r) + s) - \tilde{X}(\lambda + s). \quad \square
 \end{aligned}$$

Next, we are ready to present our main result of this section, i.e. a closed-form expression for $\gamma_h(r, s, y)$.

Theorem 5. *We have*

$$\gamma_h(r, s, y) = \frac{r}{r - y \cdot \tilde{X}(\lambda(1-r) + s)} (-\hat{\mu}^h(s, y) + y \cdot \tilde{X}(\lambda(1-r) + s) \cdot r^{h-1}),$$

where $h = 1, 2, \dots$, and $\hat{\mu}(s, y)$ is the smallest root in x with absolute value smaller than one of the function $x = y \cdot \tilde{X}(\lambda(1-x) + s)$.

Proof. Starting from the definition of $\gamma_h(r, s, y)$ and applying Lemmas 10 and 11, we obtain the following relations after some manipulations:

$$\gamma_1(r, s, y) \cdot \left(1 - \frac{y}{r} \cdot \tilde{X}(\lambda(1-r) + s)\right) = y \cdot (\tilde{X}(\lambda(1-r) + s) - \tilde{X}(\lambda + s) \cdot (1 + \gamma_{11}(s, y))), \tag{16}$$

and for $h = 2, 3, \dots$,

$$\gamma_h(r, s, y) \cdot \left(1 - \frac{y}{r} \cdot \tilde{X}(\lambda(1-r) + s)\right) = y \cdot (\tilde{X}(\lambda(1-r) + s) \cdot r^{h-1} - \tilde{X}(\lambda + s) \cdot \gamma_{h1}(s, y)). \tag{17}$$

Let $\hat{\mu}(s, y)$ denote the smallest root of the function $x = y \cdot \tilde{X}(\lambda(1-x) + s)$ in x with absolute value smaller than one. Since the functions $\gamma_h(r, s, y)$ should be analytic for $|r| \leq 1$, it follows that $\hat{\mu}(s, y)$ is a zero of the right-hand side of the expressions above. Thus, we immediately obtain for $\gamma_{h1}(s, y)$

$$\gamma_{11}(s, y) = \frac{\hat{\mu}(s, y) - y \cdot \tilde{X}(\lambda + s)}{y \cdot \tilde{X}(\lambda + s)}, \quad \gamma_{h1}(s, y) = \frac{\hat{\mu}^h(s, y)}{y \cdot \tilde{X}(\lambda + s)}, \quad h = 2, 3, \dots$$

Note that inserting $h = 1$ in the latter expression, which we denote by $(\gamma_{h1}(s, y))|_{h=1}$, shows that $\gamma_{11}(s, y) + 1 = (\gamma_{h1}(s, y))|_{h=1}$. Finally, inserting these expressions into (16) and (17) completes the proof. □

Appendix B: Proofs of results in Section 3

For convenience, let us recall the following definitions for $t > 0$:

$$p_{hk}^{(n)}(t) := \begin{cases} \mathbb{P}(x_t = k, D(t) = n \mid x_0 = h) & h, k, n = 0, 1, \dots, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t \mid z_{(0)} = h), \quad n = 1, 2, \dots, h, j = 0, 1, \dots,$$

$$F_k^{(0)}(t) := \mathbf{1}_{\{k=0\}} \mathbb{P}(A(t) = 0, I > t) + \mathbf{1}_{\{k \geq 1\}} \mathbb{P}(A(t) = k, I + X > t), \quad k = 0, 1, \dots,$$

$$F_k(t) := \mathbb{P}(A(t) = k, X > t), \quad k = 0, 1, \dots$$

Proof of Lemma 3

Proof. The proof of Lemma 3 is carried out as follows:

$$p_{hk}^{(n)}(t) := \mathbb{P}(x_t = k, D(t) = n \mid x_0 = h) = \mathbb{P}(x_t = k, r'_n \leq t, r'_{n+1} > t \mid z_{(0)} = h) \tag{18}$$

$$= \int_{u=0}^t \sum_{j=0}^k \mathbb{P}(x_t = k, r'_{n+1} > t \mid r'_n = u, z_{(0)} = h, z_{(n)} = j) \times d_u \mathbb{P}(r'_n \leq u, z_{(n)} = j \mid z_{(0)} = h) \tag{19}$$

$$= \int_{u=0}^t F_k^{(0)}(t-u) dP_{h0}^{(n)}(u) + \sum_{j=1}^k \int_{u=0}^t F_{k-j}(t-u) dP_{hj}^{(n)}(u). \tag{20}$$

That is, first we rewrite the event $D(t) = n$ and use the assumption that at time 0 the 0th customer departed from the queue, so that we obtain (18). Next, we condition on the number of

customers present at the n th departure, $z(n)$, and on the time this departure occurs, r'_n , which leads to (19). Finally, observing that r_{n+1} , $n = 0, 1, \dots$, depends in fact only on r_n and $z(n)$, using that the arrival process is stationary and applying the definitions of $F_k^{(0)}(t)$, $F_k(t)$, and $P_{hj}^{(n)}(t)$ provides us with (20).

Let us define the following LSTs:

$$\begin{aligned} \tilde{F}_k^{(0)}(s) &:= \int_{0-}^{\infty} e^{-st} dF_k^{(0)}(t), \quad k = 0, 1, \dots, \\ \tilde{F}_k(s) &:= \int_{0-}^{\infty} e^{-st} dF_k(t), \quad k = 0, 1, \dots, \\ \pi_{hj}^{(n)}(s) &:= \int_{0-}^{\infty} e^{-st} dP_{hj}^{(n)}(t), \quad n = 1, 2, \dots, h, j = 0, 1, \dots \end{aligned}$$

Then, we may present the following result as an immediate consequence of Lemma 3.

Corollary 1. *We have*

$$\int_{t=0-}^{\infty} e^{-st} dP_{hk}^{(n)}(t) = \tilde{F}_k^{(0)}(s)\pi_{h0}^{(n)}(s) + \sum_{j=1}^k \tilde{F}_{k-j}(s)\pi_{hj}^{(n)}(s). \quad \square$$

Proof of Lemma 4

Before we get to the actual proof of Lemma 4, we present another lemma. Let us introduce the auxiliary functions $G^{(0)}(r, s)$ and $G(r, s)$. These functions refer to the number of customers that arrive in the system during a period which starts at a service completion instant at an arbitrary queue Q and ends at a timer expiration which occurs before the next service is completed. More specifically, the function $G^{(0)}(r, s)$ refers to the case with zero customers present after a service completion, while $G(r, s)$ refers to the case with a strictly positive number of customers present at a service completion instant.

Lemma 12. *We have*

$$\begin{aligned} G^{(0)}(r, s) &:= \sum_{k=0}^{\infty} r^k \tilde{F}_k^{(0)}(s) = \frac{s}{\lambda(1-r) + s} \cdot \frac{\lambda(1-r \cdot \tilde{X}(\lambda(1-r) + s)) + s}{\lambda + s}, \\ G(r, s) &:= \sum_{k=0}^{\infty} r^k \tilde{F}_k(s) = \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)). \end{aligned}$$

Proof. First we will prove the expression for $G^{(0)}(r, s)$ as follows:

$$\begin{aligned} G^{(0)}(r, s) &:= \sum_{k=0}^{\infty} r^k \tilde{F}_k^{(0)}(s) \\ &= s \cdot \int_{t=0}^{\infty} e^{-st} \mathbb{P}(A(t) = 0) dt + r \cdot \sum_{k=1}^{\infty} r^{k-1} \cdot s \cdot \int_{t=0}^{\infty} e^{-st} \mathbb{P}(A(t) = k, \quad I + X > t) dt \end{aligned} \tag{21}$$

$$= \frac{s}{\lambda(1-r) + s} \cdot \frac{\lambda(1-r \cdot \tilde{X}(\lambda(1-r) + s)) + s}{\lambda + s}. \tag{22}$$

That is, we separate the terms for $k = 0$ and $k \geq 1$, insert the expression for $\tilde{F}_k^{(0)}(s)$ and perform some simple calculations, yielding (21). Next, we condition on the interarrival time (for the case $k \geq 1$) and use the fact that for a given time t the events $\{A(t) = k\}$ and $\{X > t\}$ are independent. The final expression, (22), then readily follows from the Poisson arrival assumption and some simple manipulations. Analogously, we find for $G(r, s)$ that

$$G(r, s) := \sum_{k=0}^{\infty} r^k \tilde{F}_k(s) = \sum_{k=0}^{\infty} r^k \cdot s \int_{t=0}^{\infty} e^{-st} \mathbb{P}(A(t) = k, X > t) dt$$

$$= \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)). \quad \square$$

Let us give several definitions which will be used in the proof of Lemma 4:

$$\pi_h^{(n)}(r, s) := \sum_{j=0}^{\infty} r^j \pi_{hj}^{(n)}(s), \quad h = 0, 1, \dots, n = 1, 2, \dots,$$

$$\pi_{h0}(s, y) := \sum_{n=1}^{\infty} y^n \pi_{h0}^{(n)}(s), \quad h = 0, 1, \dots,$$

$$\pi_h(r, s, y) := \sum_{n=1}^{\infty} y^n \pi_h^{(n)}(r, s), \quad h = 0, 1, \dots$$

Proof of Lemma 4. The proof of Lemma 4 in fact consists of the following three main steps:

$$\sum_{n=1}^{\infty} y^n \sum_{k=0}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dp_{hk}^{(n)}(t) \tag{23}$$

$$= \sum_{n=1}^{\infty} y^n \sum_{k=0}^{\infty} r^k \left(\tilde{F}_k^{(0)}(s) \pi_{h0}^{(n)}(s) + \sum_{j=1}^k \tilde{F}_{k-j}(s) \pi_{hj}^{(n)}(s) \right) \tag{24}$$

$$= \sum_{n=1}^{\infty} y^n (G^{(0)}(r, s) \cdot \pi_{h0}^{(n)}(s) + G(r, s) (\pi_h^{(n)}(r, s) - \pi_{h0}^{(n)}(s))) \tag{25}$$

$$= \frac{s}{\lambda(1-r) + s} \cdot \frac{\lambda(1-r) \cdot \tilde{X}(\lambda(1-r) + s) + s}{\lambda + s} \cdot \pi_{h0}(s, y)$$

$$+ \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)) \cdot (\pi_h(r, s, y) - \pi_{h0}(s, y)).$$

In the first step, we substitute the result of Corollary 1 into (23) leading to (24). Next, we derive the generating function with respect to the number of customers at the end of a visit. After some manipulations and using the definitions of $G^{(0)}(r, s)$, $G(r, s)$, $\pi_{h0}^{(n)}(s)$, and $\pi_h^{(n)}(r, s)$, we arrive at (25). In the final step, we use the definitions of $\pi_h(r, s, y)$ and $\pi_{h0}(s, y)$ and insert the explicit expressions for $G^{(0)}(r, s)$ and $G(r, s)$ which were derived in Lemma 12. \square

Proof of Theorem 1

We prove the expression for $\beta^i(\mathbf{z})$ for a specific queue Q_i as given in Theorem 1 by first deriving the conditional p.g.f. $\beta_{\mathbf{n}}^i(\mathbf{z}) := \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} | \mathbf{N}_i^s = \mathbf{n}]$ and then unconditioning on \mathbf{N}_i^s , the

number of customers present at the start of a visit to Q_i . For convenience, let us define ξ_i^* as follows:

$$\xi_i^* := \xi_i + \sum_{j \neq i} \lambda_j (1 - z_j).$$

We recall that we refer to a specific queue Q_i by adding an index i to a generic variable. Next, $\beta_{\mathbf{n}}^i(\mathbf{z})$ can be expressed as follows.

Lemma 13. *We have*

$$\begin{aligned} \beta_{\mathbf{n}}^i(\mathbf{z}) &= \frac{\xi_i}{\xi_i^*} \cdot (G_i^{(0)}(z_i, \xi_i^*) \cdot (\pi_{n_i,0}(\xi_i^*, r^i(\mathbf{z})) + \mathbf{1}_{\{n_i=0\}}) + G_i(z_i, \xi_i^*) \cdot (\pi_{n_i}(z_i, \xi_i^*, r^i(\mathbf{z})) + z_i^{n_i} \\ &\quad - \pi_{n_i,0}(\xi_i^*, r^i(\mathbf{z})) - \mathbf{1}_{\{n_i=0\}})) \cdot \prod_{j \neq i} z_j^{n_j}. \end{aligned} \tag{26}$$

Proof. Let $A_{i,j}(t)$ denote the number of arrivals to Q_j (both external and internal arrivals) during a visit to Q_i . Recall further that $D_i(t)$ denotes the number of departures at Q_i from time 0 to t . Starting from the definition of the p.g.f., we condition on the timer Y_i and introduce the number of departures from Q_i until time t :

$$\begin{aligned} \beta_{\mathbf{n}}^i(\mathbf{z}) &= \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1} \cdots z_M^{m_M} \mathbb{P}(\mathbf{N}_i^e = \mathbf{m} \mid \mathbf{N}_i^s = \mathbf{n}) \\ &= \int_0^{\infty} \xi_i e^{-\xi_i t} \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1} \cdots z_M^{m_M} \sum_n \mathbb{P}(\mathbf{N}_i(t) = \mathbf{m}, D_i(t) = n \mid \mathbf{N}_i(0) = \mathbf{n}) dt. \end{aligned}$$

After some simple rearrangements and using that, given t and $D_i(t)$, the queue-length process at Q_i is independent of the aggregate arrival process to the other queues, we obtain the following:

$$\begin{aligned} &\int_0^{\infty} \xi_i e^{-\xi_i t} \sum_n \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1 - n_1} \cdots z_M^{m_M - n_M} \\ &\quad \times \mathbb{P}(\{A_{i,j}(t) = m_j - n_j, \text{ for } j \neq i\} \mid D_i(t) = n, \mathbf{N}_i(0) = \mathbf{n}) \\ &\quad \times \sum_{m_i} z_i^{m_i} \mathbb{P}(N_{i,i}(t) = m_i \mid D_i(t) = n, \mathbf{N}_i(0) = \mathbf{n}) \\ &\quad \times \mathbb{P}(D_i(t) = n \mid \mathbf{N}_i(0) = \mathbf{n}) dt \cdot \prod_{j \neq i} z_j^{n_j}. \end{aligned}$$

These aggregate arrivals to $Q_j, j \neq i$, can be decomposed into two independent parts, namely a first part referring to external arrivals at each queue and a second part referring to customers that were served at Q_i and routed to some other queue. The latter is represented by the term $(r^j(\mathbf{z}))^n$. Also noting that $N_{i,i}(t)$ depends only on $\mathbf{N}_i(0)$ through $N_{i,i}(0)$, we retrieve $p_{n_i, m_i}^{(n)}(t)$ and eventually find that

$$\beta_{\mathbf{n}}^i(\mathbf{z}) = \int_0^{\infty} \xi_i e^{-\xi_i^* t} \sum_{n=0}^{\infty} \sum_{m_i=0}^{\infty} z_i^{m_i} (r^i(\mathbf{z}))^n p_{n_i, m_i}^{(n)}(t) dt \cdot \prod_{j \neq i} z_j^{n_j}. \tag{27}$$

Then, we can apply Lemma 4 for $n \geq 1$, while for $n = 0$ we use

$$\begin{aligned} & \sum_{m_i=0}^{\infty} z_i^{m_i} \int_0^{\infty} \xi_i e^{-\xi_i^* t} p_{n_i m_i}^{(0)}(t) dt \\ &= \mathbf{1}_{\{n_i=0\}} \cdot \sum_{m_i=0}^{\infty} z_i^{m_i} \int_0^{\infty} \xi_i e^{-\xi_i^* t} \mathbb{P}(A_i(t) = m_i, I_i + X_i > t) dt \\ & \quad + \mathbf{1}_{\{n_i \geq 1\}} \cdot \sum_{m_i=0}^{\infty} z_i^{m_i} \int_0^{\infty} \xi_i e^{-\xi_i^* t} \mathbb{P}(A_i(t) = m_i - n_i, X_i > t) dt \\ &= \frac{\xi_i}{\xi_i^*} \cdot (\mathbf{1}_{\{n_i=0\}} \cdot G_i^{(0)}(z_i, \xi_i^*) + \mathbf{1}_{\{n_i \geq 1\}} \cdot z_i^{n_i} \cdot G_i(z_i, \xi_i^*)). \end{aligned}$$

The final expression for $\beta_{\mathbf{n}}^i(\mathbf{z})$ follows from inserting this result together with the result from Lemma 4 into (27) and some simple manipulations. \square

Proof of Theorem 1. The proof follows immediately by unconditioning $\beta_{\mathbf{n}}^i(\mathbf{z})$ on the state $\mathbf{n} = (n_1, \dots, n_M)$ at the start of the visit. The result of this operation is shown below. Equation (28) follows by substitution of (26) into the definition of $\beta^i(\mathbf{z})$. We note that the final expression, (29), follows from inserting the explicit expressions for $G_i^{(0)}(r, s)$ and $G_i(r, s)$ (see Lemma 12), inserting the expressions for $\pi_h(z_i, \xi_i^*, r^i(\mathbf{z}))$ and $\pi_{h0}(\xi_i^*, r^i(\mathbf{z}))$, $h \geq 0$, which are given in (2), (4), and (5), and some simple manipulations. That is,

$$\begin{aligned} \beta^i(\mathbf{z}) &= \sum_{n_1=0}^{\infty} \dots \sum_{n_M=0}^{\infty} \beta_{\mathbf{n}}^i(\mathbf{z}) \mathbb{P}(\mathbf{N}_i^s = \mathbf{n}) \\ &= \sum_{n_1=0}^{\infty} \dots \sum_{n_M=0}^{\infty} \mathbb{P}(\mathbf{N}_i^s = \mathbf{n}) \cdot \prod_{j \neq i} z_j^{n_j} \cdot \frac{\xi_i}{\xi_i^*} \\ & \quad \times (G_i(z_i, \xi_i^*) \cdot (\pi_{n_i}(z_i, \xi_i^*, r^i(\mathbf{z})) + z_i^{n_i} - \pi_{n_i,0}(\xi_i^*, r^i(\mathbf{z})) - \mathbf{1}_{\{n_i=0\}}) \\ & \quad + G_i^{(0)}(z_i, \xi_i^*) \cdot (\pi_{n_i,0}(\xi_i^*, r^i(\mathbf{z})) + \mathbf{1}_{\{n_i=0\}})) \end{aligned} \tag{28}$$

$$\begin{aligned} &= \frac{\xi_i}{z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\xi_i + \sum_j \lambda_j(1 - z_j))} \\ & \quad \times \left(\frac{\tilde{X}_i(\xi_i + \sum_j \lambda_j(1 - z_j)) \cdot (z_i - r^i(\mathbf{z}))}{\lambda_i(1 - \hat{\mu}_i(\xi_i, r^i(\mathbf{z})) + \xi_i^*)} \cdot \alpha^i(\mathbf{z}_i^*) \right. \\ & \quad \left. + \frac{(1 - \tilde{X}_i(\xi_i + \sum_j \lambda_j(1 - z_j))) \cdot z_i}{\lambda_i(1 - z_i) + \xi_i^*} \cdot \alpha^i(\mathbf{z}) \right), \end{aligned} \tag{29}$$

where

$$\alpha^i(\mathbf{z}_i^*) := \mathbb{E}[z_1^{N_1^s} \dots \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))^{N_i^s} \dots z_M^{N_M^s}]$$

and $\hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))$ is the root x with the smallest absolute value less than one of $x = r^i(\mathbf{z}) \cdot \tilde{X}_i(\xi_i^* + \lambda_i(1 - x))$. \square

Appendix C: Proofs of results in Section 4

For convenience, let us recall the following definitions for $t > 0$:

$$q_{hk}^{(n)}(t) := \begin{cases} \mathbb{P}(x_t = k, D(t) = n, x_v > 0, 0 < v < t | x_0 = h) & n = 0, 1, \dots, h, k = 1, 2, \dots \\ 0 & \text{otherwise,} \end{cases}$$

$$R_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t, z_{(k)} > 0, 0 < k < n | z_{(0)} = h), \quad h, j, n = 1, 2, \dots,$$

$$F_k(t) := \mathbb{P}(A(t) = k, X > t), \quad k = 0, 1, \dots$$

Proof of Lemma 5

Proof. The first observation is that each customer at $Q_j, j \neq i$, will still be present at the end of the visit, which is accounted for in the term $\prod_{j \neq i} z_j^{n_j}$. Second, each customer present at the start of the visit at Q_i will effectively be replaced by a random population during the course of the visit in an identical fashion. In particular, the size of this population is given by $\hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))$. To see this, recall that $\hat{\mu}_i(s, y)$ refers to the joint generating function of the busy period and the number of customers served during this period. The term ξ_i^* in $\hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))$ accounts for the exogenous arrivals to the other queues in the system during a busy period that ends before the timer expires. Similarly, the term $r^i(\mathbf{z})$ in $\hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))$ accounts for the internal arrivals to the other queues (from Q_i) during this period. As initially there are n_i identical customers present at Q_i , this leads to n_i independent contributions which are recognized in the power of $\hat{\mu}_i(\xi_i^*, r^i(\mathbf{z}))$. □

Proof of Lemma 6

Proof. Lemma 6 is readily proved by using arguments similar to those in the proof of Lemma 3:

$$\begin{aligned} q_{hk}^{(n)}(t) &= \mathbb{P}(x_t = k, D(t) = n, x_v > 0, 0 < v < t | x_0 = h) \\ &= \mathbb{P}(x_t = k, r'_n \leq t, r'_{n+1} > t, x_v > 0, 0 < v < t | z_{(0)} = h) \\ &= \int_{u=0}^t \sum_{j=1}^k \mathbb{P}(x_t = k, r'_{n+1} > t | r'_n = u, z_{(0)} = h, z_{(m)} > 0, 0 \leq m \leq n, z_{(n)} = j) \\ &\quad \times d_u \mathbb{P}(r'_n \leq u, z_{(n)} = j, z_{(m)} > 0, 0 < m < n | z_{(0)} = h) \\ &= \int_{u=0}^t \sum_{j=1}^k F_{k-j}(t-u) dR_{hj}^{(n)}(u). \end{aligned}$$

Let us define the following LSTs:

$$\begin{aligned} \tilde{F}_k(s) &:= \int_{0-}^{\infty} e^{-st} dF_k(t), \quad k = 0, 1, \dots, \\ \gamma_{hj}^{(n)}(s) &:= \int_{0-}^{\infty} e^{-st} dR_{hj}^{(n)}(t), \quad h, j, n = 1, 2, \dots \end{aligned} \quad \square$$

Note that a direct consequence of Lemma 6 is the following.

Corollary 2. *We have*

$$\int_{t=0}^{\infty} e^{-st} dq_{hk}^{(n)}(t) = \sum_{j=1}^k \gamma_{hj}^{(n)}(s) \tilde{F}_k(s), \quad h, k, n = 1, 2, \dots \tag{30}$$

Proof of Lemma 7

Let us give several definitions which will be used in the proof of Lemma 7:

$$\gamma_h^{(n)}(r, s) := \sum_{j=0}^{\infty} r^j \gamma_{hj}^{(n)}(s), \quad h, n = 1, 2, \dots,$$

$$\gamma_h(r, s, y) := \sum_{n=1}^{\infty} y^n \gamma_h^{(n)}(r, s), \quad h = 1, 2, \dots$$

Proof of Lemma 7. The proof consists of three consecutive steps similar to the proof of Lemma 4:

$$\begin{aligned} & \sum_{n=1}^{\infty} y^n \sum_{k=1}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dq_{hk}^{(n)}(t) \\ &= \sum_{n=1}^{\infty} y^n \sum_{k=1}^{\infty} r^k \sum_{j=1}^k \gamma_{hj}^{(n)}(s) \tilde{F}_{k-j}(s). \end{aligned} \tag{31}$$

$$\begin{aligned} &= \sum_{n=1}^{\infty} y^n \gamma_h^{(n)}(r, s) G(r, s) \\ &= \gamma_h(r, s, y) \cdot \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)). \end{aligned} \tag{32}$$

First we substitute (30) into (31). Next, using the definitions of $\gamma_h^{(n)}(r, s)$ and $G(r, s)$ (see Lemma 12) immediately yields (32). The final step follows from the definition of $\gamma_h(r, s, y)$ and the substitution of the explicit expression for $G(r, s)$. □

Proof of Lemma 8

As a preliminary to proving Lemma 8, we present the following result for $h = 1, 2, \dots$, for the special case of $D(t) = 0$, that is, no departures occur before the timer expires.

Lemma 14. *We have*

$$\sum_{k=1}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dq_{hk}^{(0)}(t) = r^h \cdot \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)).$$

Proof.

$$\begin{aligned} & \sum_{k=1}^{\infty} r^k \int_{t=0}^{\infty} e^{-st} dq_{hk}^{(0)}(t) \\ &= r^h \cdot \int_{t=0}^{\infty} s e^{-st} \sum_{k=h}^{\infty} r^{k-h} \mathbb{P}(A(t) = k - h, X > t) dt \end{aligned} \tag{33}$$

$$= r^h \cdot \frac{s}{\lambda(1-r) + s} \cdot (1 - \tilde{X}(\lambda(1-r) + s)). \tag{34}$$

Elaborating on the definition of $q_{hk}^{(0)}(t)$, we may obtain (33) after some simple manipulations. Equation (34) then follows directly from the earlier derivation of $G(r, s)$ (see Lemma 12). \square

Proof of Lemma 8. To consider a specific queue Q_i , we will again add an index i to the generic variables. Let $A_{i,j}(t)$ denote the number of arrivals to Q_j (both external and internal arrivals) during a visit to Q_i . Recall further that $D_i(t)$ denotes the number of departures at Q_i from time 0 to t . Starting from the definition of the p.g.f., we condition on the timer Y_i and introduce the number of departures from Q_i until time t :

$$\begin{aligned} & \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}] \\ &= \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1} \cdots z_M^{m_M} \mathbb{P}(\mathbf{N}_i^e = \mathbf{m}, \{\text{timer}\} \mid \mathbf{N}_i^s = \mathbf{n}) \\ &= \int_0^{\infty} \xi_i e^{-\xi_i t} \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1} \cdots z_M^{m_M} \\ & \quad \times \sum_n \mathbb{P}(\mathbf{N}_i(t) = \mathbf{m}, \{\text{timer}\}, D_i(t) = n \mid \mathbf{N}_i(0) = \mathbf{n}) dt. \end{aligned}$$

Using that given t and $D_i(t)$ the queue-length process at Q_i is independent of the aggregate arrival process to the other queues and working out the event $\{\text{timer}\}$, we obtain

$$\begin{aligned} & \int_0^{\infty} \xi_i e^{-\xi_i t} \sum_n \sum_{m_1=0}^{\infty} \cdots \sum_{m_M=0}^{\infty} z_1^{m_1-n_1} \cdots z_M^{m_M-n_M} \\ & \quad \times \mathbb{P}(\{A_{i,j}(t) = m_j - n_j, \text{ for } j \neq i\} \mid D_i(t) = n, \mathbf{N}_i(0) = \mathbf{n}) \\ & \quad \times \sum_{m_i} z_i^{m_i} \mathbb{P}(N_{i,i}(t) = m_i, N_{i,i}(v) > 0, 0 < v < t \mid D_i(t) = n, \mathbf{N}_i(0) = \mathbf{n}) \\ & \quad \times \mathbb{P}(D_i(t) = n \mid \mathbf{N}_i(0) = \mathbf{n}) dt \cdot \prod_{j \neq i} z_j^{n_j}. \end{aligned}$$

Exactly following the same reasoning that led to (27), we obtain

$$\mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{timer}\}} \mid \mathbf{N}_i^s = \mathbf{n}] = \int_{t=0}^{\infty} \xi_i e^{-\xi_i^* t} \sum_{n=0}^{\infty} \sum_{m_i=1}^{\infty} z_i^{m_i} (r^i(\mathbf{z}))^n q_{n,m_i}^{(n)}(t) dt \cdot \prod_{j \neq i} z_j^{n_j}.$$

Then, we may apply Lemma 7 for $n \geq 1$ and Lemma 14 for $n = 0$. This yields the desired result in (9) after substituting the explicit expressions for $G_i(r, s)$ (see Lemma 12) and $\gamma_h(r, s, y)$ (see (2)) and performing some simple manipulations. \square

Proof of Theorem 2

The final result for $\beta^i(\mathbf{z})$ is obtained by first unconditioning the conditional p.g.f.s of the previous lemmas and then merging these outcomes. Let us define $\beta_e^i(\mathbf{z}) := \mathbb{E}[\mathbf{z}^{\mathbf{N}_i^e} \mathbf{1}_{\{\text{empty}\}}]$ and

$\beta_t^i(\mathbf{z}) := \mathbb{E}[\mathbf{z}^{N_t^e} \mathbf{1}_{\{\text{timer}\}}]$. Recall that $\alpha^i(\mathbf{z}_i^*) = \alpha^i(z_1, \dots, z_{i-1}, \hat{\mu}_i(\xi_i^*, r^i(\mathbf{z})), z_{i+1}, \dots, z_M)$. By unconditioning the expressions in Lemmas 5 and 8, it follows immediately that

$$\beta_e^i(\mathbf{z}) = \alpha^i(\mathbf{z}_i^*)$$

and

$$\beta_t^i(\mathbf{z}) = \frac{\xi_i \cdot z_i \cdot (1 - \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*))}{(\lambda_i(1 - z_i) + \xi_i^*)(z_i - r^i(\mathbf{z}) \cdot \tilde{X}_i(\lambda_i(1 - z_i) + \xi_i^*))} \cdot (\alpha^i(\mathbf{z}) - \alpha^i(\mathbf{z}_i^*)).$$

Proof of Theorem 2. The proof follows directly from $\beta^i(\mathbf{z}) = \beta_e^i(\mathbf{z}) + \beta_t^i(\mathbf{z})$, where the last two terms are given in the latter two equations. □

Proofs of results in Section 5.1

For convenience, let us recall that the batch sizes are assumed to be mutually independent and independent of the arrival and the service processes. Let λ denote the arrival rate of batches and $\hat{\psi}(\cdot)$ the generating function of the batch size. Recall that the generating function of the service time of an individual customer is denoted by $\tilde{X}(\cdot)$

Let $z_{(n)}$ denote the number of customers left behind by the n th departing customer after time 0 and let r'_n be its departure time. We assume that at time $t = 0$ customer 0 departs. Let $R_{hj}^{(n)}(t)$ denote the probability

$$R_{hj}^{(n)}(t) := \mathbb{P}(z_{(n)} = j, r'_n \leq t, z_{(k)} > 0, 0 < k < n \mid z_{(0)} = h),$$

for $h = 1, 2, \dots$

Proof of Lemma 9. Let us denote

$$P_{hk}^{(n)}(t) := \mathbb{P}(z_{(n)} = k, r'_n \leq t \mid z_{(0)} = h).$$

Our objective is to find the joint generating function

$$\pi_h(r, s, y) := \sum_{n=1}^{\infty} y^n \mathbb{E}[r^{z_{(n)}} e^{-sr'_n}].$$

Define

$$\pi_{hk}(s, y) := \sum_{n=1}^{\infty} y^n \int_0^{\infty} e^{-st} dP_{hk}^{(n)}(t).$$

Following [7, pages 239–240, 386–387], we find that

$$\begin{aligned} &\pi_h(r, s, y)(r - y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))) \\ &= \frac{y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))}{\lambda + s} (\lambda\pi_{h0}(s, y)(\hat{\psi}(r) - \lambda - s) + (\lambda + s)\delta_h(r, s)), \end{aligned} \tag{35}$$

where $h = 0, 1, 2, \dots$, and

$$\delta_h(r, s) = \begin{cases} \frac{\lambda \hat{\psi}(r)}{\lambda + s} & h = 0, \\ r^h & h = 1, 2, \dots \end{cases}$$

For $\text{Re}(s) \geq 0$ and $|y| < 1$, the equation $r = y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))$ has a unique root $r = \hat{\mu}(s, y)$ with $|r| < 1$ which can be seen by applying Takács' lemma. Since for $\text{Re}(s) \geq 0$ and $|y| < 1$ the functions $\pi_h(r, s, y)$, $h = 0, 1, \dots$ are analytic in $|r| \leq 1$, it follows that $\hat{\mu}(s, y)$ should also be a zero of the right-hand side of (35). After some algebra we find that

$$\pi_{00}(s, y) = \frac{\lambda}{\lambda - \hat{\psi}(\hat{\mu}(s, y)) + s} \hat{\mu}(s, y),$$

$$\pi_{h0}(s, y) = \frac{\lambda + s}{\lambda - \hat{\psi}(\hat{\mu}(s, y)) + s} \hat{\mu}^h(s, y), \quad h = 1, 2, \dots$$

The latter yields that

$$\pi_h(r, s, y) = \left(y\tilde{X}(s + \lambda(1 - \hat{\psi}(r))) \left(r^h - \frac{s + \lambda(1 - \hat{\psi}(r))}{s + \lambda(1 - \hat{\psi}(\hat{\mu}(s, y)))} \hat{\mu}^h(s, y) \right) \right) / (r - y\tilde{X}(s + \lambda(1 - \hat{\psi}(r)))) ,$$

for $h = 0, 1, \dots$

An analysis similar to that in Appendix A gives the joint generating function $\gamma_h(r, s, y)$ of the probability $R_{hj}^{(n)}(t)$, which completes the proof. \square

References

- [1] AL HANBALI, A., DE HAAN, R., BOUCHERIE, R. J. AND VAN OMMEREN, J.-K. (2008). Time-limited and k -limited polling systems: a matrix analytic solution. In *Proc. 3rd International Conference on Performance Evaluation Methodologies and Tools*, art. 17. Association for Computing Machinery.
- [2] AL HANBALI, A., DE HAAN, R., BOUCHERIE, R. J. AND VAN OMMEREN, J.-K. (2012). Time-limited polling systems with batch arrivals and phase-type service times. *Ann. Operat. Res.* **198** (1), 57–82.
- [3] BOON, M. A., VAN DER MEI, R. AND WINANDS, E. M. (2011). Applications of polling systems. *Surv. Operat. Res. Manag. Sci.* **16** (2), 67–82.
- [4] BORST, S. AND BOXMA, O. (2018). Polling: past, present, and perspective. *TOP* **26**, 335–369.
- [5] BOXMA, O. J. AND WESTSTRATE, J. A. (1989). Waiting times in polling systems with Markovian server routing. In *Messung, Modellierung und Bewertung von Rechensystemen und Netzen Informatik-Fachberichte*, eds G. Stiege and J. S. Lie, pp. 89–104. Springer, Berlin and Heidelberg.
- [6] COFFMAN, E. G., FAYOLLE, G. AND MITRANI, I. (1988). Two queues with alternating service periods. In *Performance '87: Proc. 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*, pp. 227–239.
- [7] COHEN, J. W. (1982). *The Single Server Queue*, revised edn. North-Holland.
- [8] DE HAAN, R. (2009). Queueing models for mobile ad hoc networks. Doctoral thesis, University of Twente, Enschede. <http://doc.utwente.nl/61385>
- [9] DE HAAN, R., BOUCHERIE, R. J. AND VAN OMMEREN, J.-K. (2009). A polling model with an autonomous server. *Queueing Systems* **62** (3), 279–308.
- [10] DE SOUZA E SILVA, E., GAIL, H. R. AND MUNTZ, R. R. (1995). Polling systems with server timeouts and their application to token passing networks. *IEEE/ACM Trans. Networking* **3** (5), 560–575.
- [11] ELIAZAR, I. AND YECHIALI, U. (1998). Polling under the randomly-timed gated regime. *Stoch. Models* **14** (1), 79–93.
- [12] ELIAZAR, I. AND YECHIALI, U. (1998). Randomly timed gated queueing systems. *SIAM J. Appl. Math.* **59** (2), 423–441.
- [13] FRICKER, C. AND JAIBI, M. (1994). Monotonicity and stability of periodic polling models. *Queueing Systems* **15** (1–4), 211–238.
- [14] FUHRMANN, S. W. (1992). A decomposition result for a class of polling models. *Queueing Systems* **11** (1–2), 109–120.
- [15] KELLY, F. P. (2011). *Reversibility and Stochastic Networks*. Cambridge University Press.
- [16] LEUNG, K. (1994). Cyclic-service systems with non-preemptive time-limited service. *IEEE Trans. Commun.* **42** (8), 2521–2524.

- [17] LEVY, H. AND SIDI, M. (1990). Polling systems: applications, modeling, and optimization. *IEEE Trans. Commun.* **38** (10), 1750–1760.
- [18] RESING, J. (1993). Polling systems and multitype branching processes. *Queueing Systems* **13** (4), 409–426.
- [19] TAKAGI, H. (1997). Queueing analysis of polling models: progress in 1990–1994. In *Frontiers In Queueing: Models and Applications in Science and Engineering* (Probability and Stochastics Series), ed. J. H. Dshalalow, pp. 119–146. CRC Press.
- [20] TAKAGI, H. (2000). Analysis and application of polling models. In *Performance Evaluation: Origins and Directions* (Lect. Notes Comput. Sci. **1769**), pp. 423–442. Springer, Berlin.
- [21] TSE, D. AND VISWANATH, P. (2005). *Fundamentals of Wireless Communication*. Cambridge University Press.
- [22] VAN HOUTD, B. (2010). Numerical solution of polling systems for analyzing networks onchips. In *Proc. NSMC 2010, Williamsburg, VA*, pp. 90–93.
- [23] VISHNEVSKII, V. M. AND SEMENOVA, O. V. (2006). Mathematical methods to study the polling systems. *Automat. Remote Control* **67** (2), 173–220.
- [24] YECHIALI, U. (1993). Analysis and control of polling systems. In *Performance Evaluation of Computer and Communication Systems* (Lect. Notes Comput. Sci. **729**), pp. 630–650. Springer.