



# An exact root-free method for the expected queue length for a class of discrete-time queueing systems

A. Oblakova<sup>1</sup> · A. Al Hanbali<sup>2</sup> · R. J. Boucherie<sup>1</sup> ·  
J. C. W. van Ommeren<sup>1</sup> · W. H. M. Zijm<sup>1</sup>

Received: 18 June 2018 / Revised: 24 April 2019 / Published online: 17 May 2019  
© The Author(s) 2019

## Abstract

For a class of discrete-time queueing systems, we present a new exact method of computing both the expectation and the distribution of the queue length. This class of systems includes the bulk-service queue and the fixed-cycle traffic-light (FCTL) queue, which is a basic model in traffic-control research and can be seen as a non-exhaustive time-limited polling system. Our method avoids finding the roots of the characteristic equation, which enhances both the reliability and the speed of the computations compared to the classical root-finding approach. We represent the queue-length expectation in an exact closed-form expression using a contour integral. We also introduce several realistic modifications of the FCTL model. For the FCTL model for a turning flow, we prove a decomposition result. This allows us to derive a bound on the difference between the bulk-service and FCTL expected queue lengths, which turns out to be small in most of the realistic cases.

**Keywords** Bulk-service queue · Fixed-cycle traffic-light model · Roots · Contour integrals

**Mathematics Subject Classification** 60K25 · 90B22

---

This paper is part of the project Dynafloat which was supported by Topconsortia voor Kennis en Innovatie Logistiek and Nederlandse Organisatie voor Wetenschappelijk Onderzoek [Grant Number 438-13-206].

---

✉ A. Oblakova  
a.oblakova@utwente.nl

<sup>1</sup> University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands

<sup>2</sup> King Fahd University of Petroleum and Minerals, University Blvd, Dhahran 34463, Saudi Arabia

## 1 Introduction

In the analysis of many queues, the probability-generating function (pgf) of the queue length is represented as a rational function with several unknown variables. The classical way of finding the unknowns is to use the analyticity of the pgf inside the unit disk; see [3,6]. The analyticity implies that each zero of the denominator, i.e., each root of the corresponding characteristic equation, inside the closed unit disk is also a zero of the numerator, which yields a system of linear equations for the unknowns. In the case of a stable system and distinct zeros, it is possible to find all the required variables. The main drawback of this approach is finding the zeros, since, in general, no closed-form expression exists. Moreover, the precision of the found zeros also influences the result. To avoid these problems, some authors suggest approximations of the queue length; see, for example, [11,16,20]. Another approach is to use the matrix-analytic method; see [13]. However, this also requires solving a (matrix) equation.

For a certain class of discrete-time queueing systems, we present an exact root-free approach to find the expected queue length and the queue-length distribution. The key point of our method is that the unknown probabilities depend on the zeros of the denominator in a symmetric way and that, as we show, the actual values of the zeros are not important. We apply the residue theorem and find these unknown probabilities in terms of contour integrals. The mean queue length is, in fact, a function of only one such contour integral. This makes our method computationally efficient in comparison with the standard root-finding method. Moreover, our method gives a closed-form solution without implicitly defined variables such as the roots of the characteristic equation in the root-finding method or the rate matrix in the matrix-analytic approach. Thus, our solution can be used, for example, to find asymptotic results for heavy load.

The considered class of queueing systems includes, among others, the bulk-service queue, see [3], and the fixed-cycle traffic-light (FCTL) queue, see [6]. The server in the bulk-service queue serves a fixed-size batch of customers each time unit. If there are fewer customers in the queue, all of them are served. In each time unit, a random number of customers arrives, where the arrivals during different time units are independent of each other. The bulk-service queue also appears as an embedded Markov chain for the continuous-time multi-server  $M/D/s$  queue, see [10], or  $M^X/G^B/1$  queue with batch arrivals and batch service. The FCTL model is the basic and key model in traffic-control research and can be seen as a polling system with non-exhaustive time-limited service discipline. The service, visiting and switching times of the server are fixed and deterministic. Unlike most polling systems, the server continues to serve even if the queue empties, as it has no knowledge of the queue length (there are no detectors), cf. [2,7]. However, if the queue empties during the green (service) time, then, during the rest of the green time, all arriving customers are served instantly, i.e., arriving vehicles proceed without delay. We can also apply our method for queueing systems, where the pgf of the analyzed random variable depends on the *minimal polynomial* of the unknown zeros, i.e., the polynomial that has the same zeros (with the same multiplicity) and has the smallest power. An example of such a system is the discrete-time  $G/G/1$  queue; see [18].

In this paper, we also consider a modification of the FCTL model for a lane with turning flow, where, unlike the FCTL model, the customers arriving during green time

and finding an empty queue require the same service time as the queued vehicles. For this new model, we prove a decomposition result similar to [5]. Moreover, we prove that given the same arrival process and server capacity, the expected bulk-service queue is bounded from below by the expected *overflow queue* for the FCTL model and from above by the expected overflow queue for the FCTL model for turning flow, where the overflow queue is the queue just after service. These results give us a bound on the difference between these models. We propose several other extensions of the FCTL queue, which include the effects of traffic interruption, for example, due to actuated control for pedestrians or uncertain departure times.

In the numerical results, we compare our method with the root-finding and matrix-analytic approaches. We observe that the root-finding approach can be numerically quite unreliable, while the matrix-analytic approach is much slower than our method. We also apply our method to the FCTL queue and its extensions. In particular, we investigate the impact of the variability of the arrival process on the expected queue length and consider the difference between straight-going and turning flows. In addition, we study effects of traffic disruption.

The paper is structured as follows: The method in its general form is presented in Sect. 2. In Sect. 3, we apply our method to the bulk-service queue and compare it with the root-finding and matrix-analytic approaches in terms of speed and reliability. In Sect. 4, we present several realistic extensions of the FCTL model. Finally, in Sect. 5, we present the conclusions and propose future research directions.

## 2 Method

In this section, we propose a generic method to analyze various discrete-time queueing systems, such as the bulk-service queue and the FCTL queue; see Sects. 3 and 4. For these systems, the pgf of the queue length is a rational function with unknown probabilities in the numerator. The classical way of determining the unknowns requires finding the roots of the characteristic equation. In our method, we use the special structure of the numerator for these systems to find an alternative root-free solution. Let us consider the function  $X(z)$  of the following form:

$$X(z) = \frac{\sum_{k=0}^{g-1} x_k z^k (B(z))^{g-1-k}}{D(z)} f(z), \tag{1}$$

where  $x_k$  are unknown coefficients that we want to determine;  $D(z)$ ,  $B(z)$  and  $f(z)$  are known functions such that  $B(1) = 1$ ,  $f(1) = 0$ , and, for each zero  $z_l \in \{z: |z| \leq 1, z \neq 1\}$  of the denominator,  $f(z_l) \neq 0$ ; and  $X(1) = 1$ . For the analysis in this section, we do not need  $X(z)$  to be a pgf. In what follows, we suppose that the following assumptions on the functions  $B(z)$ ,  $D(z)$ ,  $f(z)$  and  $X(z)$  hold:

**Assumption 1** (*Analyticity assumption*) For some  $\alpha > 0$ , the functions  $B(z)$  and  $D(z)$  are analytic and the function  $f(z)$  is two times differentiable in the disk  $D_{1+\alpha} = \{z: |z| < 1 + \alpha\}$ . The function  $X(z)$  is continuous in the closed unit disk.

**Assumption 2** (*Roots assumption*) The equation

$$D(z) = 0 \tag{2}$$

has exactly  $g$  roots  $z_0 = 1, z_1, \dots, z_{g-1}$  in the closed unit disk  $\bar{D}_1 = \{z: |z| \leq 1\}$ . For any  $z \neq w, z, w \in \bar{D}_1$ , the function  $B(z)$  satisfies

$$B(z)w \neq B(w)z.$$

Note that from Assumption 2, the following statement is deduced immediately.

**Corollary 1** *The equation  $z = B(z)$  has only one root in  $\bar{D}_1$ , namely 1.*

**Remark 1** Without loss of generality, we can assume that  $D(0) \neq 0$ . Otherwise, both denominator and numerator can be divided by  $z^n$  for some  $n > 0$  such that the new  $D(z)$  satisfies the condition  $D(0) \neq 0$ .

**2.1 Coefficients expressed in terms of roots**

In this subsection, we discuss how  $x_k, k = 0, \dots, g - 1$ , depend on the roots of (2).

Let  $y_k = B(z_k)/z_k$  for  $k = 0, \dots, g - 1$ . Since we know that  $D(0) \neq 0$ , then  $z_k \neq 0$  for each  $k = 0, \dots, g - 1$ . Rewrite  $X(z)$  in the following way:

$$X(z) = \frac{\sum_{k=0}^{g-1} x_k \left(\frac{B(z)}{z}\right)^{g-1-k}}{D(z)} f(z)z^{g-1}.$$

Since  $z_k \neq 0$  and  $f(z_k) \neq 0, k = 1, \dots, g - 1$ , the numerator is equal to 0 for  $z = z_k$  and

$$\sum_{l=0}^{g-1} x_l y_k^{g-1-l} = 0.$$

Consider the following polynomial:

$$h(y) = \sum_{l=0}^{g-1} x_l y^{g-1-l} = \sum_{l=0}^{g-1} x_{g-1-l} y^l.$$

The function  $h(y)$  is a polynomial of degree  $g - 1$ , and  $y_k, \text{ for } k = 1, \dots, g - 1$ , are zeros of this polynomial. Note that whenever  $z_k$  is a multiple root of the equation  $D(z) = 0, y_k$  is a multiple root of the polynomial  $h(y)$  with the same multiplicity. Moreover, according to Assumption 2,  $y_k \neq y_l$  if  $z_k \neq z_l$ . Thus,  $h(y)$  can be written as

$$h(y) = x_0 \prod_{k=1}^{g-1} (y - y_k). \tag{3}$$

By applying Vieta’s formulas (see [19]) to the polynomial  $h(y)$ , we get, for  $k = 1, \dots, g - 1$ , that

$$\frac{x_k}{x_0} = (-1)^k \sigma_k(y_1, \dots, y_{g-1}), \tag{4}$$

where

$$\sigma_k(y_1, \dots, y_{g-1}) = \sigma_k = \sum_{1 \leq i_1 < \dots < i_k \leq g-1} y_{i_1} \dots y_{i_k}$$

are elementary symmetric polynomials. Here, for notational convenience, we assume that

$$\sigma_0(y_1, \dots, y_{g-1}) = \sigma_0 = 1.$$

Eq. (4) gives us  $x_k$  up to a normalization constant, which we derive from the equation  $X(1) = 1$  by applying L’Hôpital’s rule:

$$X(1) = \frac{h(1)}{D'(1)} f'(1) = \frac{\sum_{k=0}^{g-1} x_k}{D'(1)} f'(1) = 1. \tag{5}$$

### 2.2 Coefficients expressed in terms of contour integrals

Using (4), we can find the coefficients  $x_k$  if we know the values of the elementary symmetric polynomials  $\sigma_k$ . We will represent these symmetric polynomials as functions of the following symmetric polynomials:

$$\eta_k = \eta_k(y_1, \dots, y_{g-1}) = \sum_{l=1}^{g-1} y_l^k, \tag{6}$$

for  $k = 1, \dots, g - 1$ . Below, in Theorem 1, we provide a way to find  $\eta_k$  without computing the roots of (2). The proof of the theorem requires the following auxiliary result, which can be obtained using L’Hôpital’s rule.

**Lemma 1** Consider a function  $F(z)$  that is analytic in a neighborhood of  $z_k$ , where  $k = 0, \dots, g - 1$ . Then, the residue of the function  $\frac{D'(z)}{D(z)} F(z)$  at  $z_k$  is equal to

$$\lim_{z \rightarrow z_k} \frac{D'(z)F(z)(z - z_k)}{D(z)} = m_{z_k} F(z_k),$$

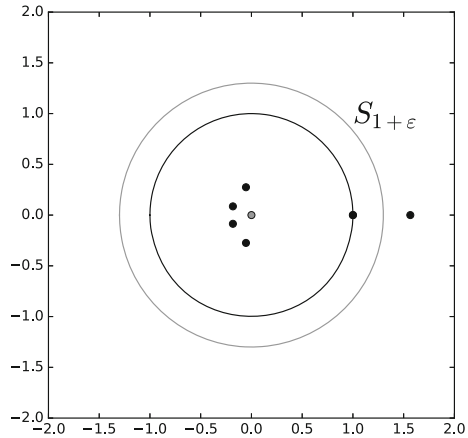
where  $m_{z_k}$  is the multiplicity of the zero  $z_k$ .

**Theorem 1** Consider  $\varepsilon > 0$  and  $\delta > 0$  such that there are no roots of Eq.(2) in  $\bar{D}_{1+\varepsilon} \setminus \bar{D}_1$  and  $\bar{D}_\delta$ . Then,

$$\eta_k = \frac{1}{2\pi i} \oint_{S_{1+\varepsilon}} \frac{D'(z)}{D(z)} \left(\frac{B(z)}{z}\right)^k dz - \frac{1}{2\pi i} \oint_{S_\delta} \frac{D'(z)}{D(z)} \left(\frac{B(z)}{z}\right)^k dz - 1, \tag{7}$$

where  $S_r = \{z: |z| = r\}$ .

**Fig. 1** Circle  $S_{1+\varepsilon}$  such that there are no roots of Eq. (2) in  $\bar{D}_{1+\varepsilon} \setminus \bar{D}_1$ . The roots of the equation  $D(z) = z^5 - (z/2 + 1/2)^9 = 0$  are represented as bold black points, and 0 is given by the bold gray point



**Proof** Note that the only singularities of the function  $\frac{D'(z)}{D(z)} \left(\frac{B(z)}{z}\right)^k$  in  $D_{1+\varepsilon}$  are  $0, z_0 = 1, z_1, \dots, z_{g-1}$ ; see Fig. 1. Thus, by the residue theorem and Lemma 1, the difference of the integrals on the right-hand side of (7) is equal to  $\sum_{l=0}^{g-1} y_l^k + r_0 - r_0 = y_0^k + \eta_k = \eta_k + 1$ , where  $r_0$  is the residue of the function  $\frac{D'(z)}{D(z)} \left(\frac{B(z)}{z}\right)^k$  at 0.  $\square$

**Remark 2** Strictly speaking, we also need  $\varepsilon < \alpha$ ; see Assumption 1. However, in many applications,  $\alpha$  is very big or infinite, and, therefore, we omit this requirement hereafter. In Appendix B, we elaborate on how to choose  $\varepsilon$  and  $\delta$ .

The following lemma provides a way of finding  $\sigma_k, k = 1, \dots, g - 1$ , using  $\eta_k, k = 1, \dots, g - 1$ . We omit the proof of this lemma since it requires only careful computation of the monomials’ coefficients on the right side of Eq. (8).

**Lemma 2** For each  $k = 1, \dots, g - 1$ , the following recurrence equation holds:

$$\sigma_k = \frac{1}{k} \sum_{l=1}^k (-1)^{l+1} \sigma_{k-l} \eta_l. \tag{8}$$

Thus, we can use formulas (4), (5) and (7), together with Lemma 2, to find the coefficients  $x_k, k = 0, \dots, g - 1$ .

### 2.3 Expectation expressed in terms of roots

Now, consider  $X'(1)$ . For queueing systems, this represents the expected queue length.

$$\begin{aligned} X'(1) &= \left( \frac{\sum_{k=0}^{g-1} x_k z^k (B(z))^{g-1-k}}{D(z)} f(z) \right)' \Big|_{z=1} \\ &= \frac{f'(1)}{D'(1)} \cdot \left( h \left( \frac{B(z)}{z} \right) z^{g-1} \right)' \Big|_{z=1} + h(1) \cdot \left( \frac{f(z)}{D(z)} \right)' \Big|_{z=1} \end{aligned}$$

$$= \frac{f'(1)}{D'(1)} \left( h'(1)(B'(1) - 1) + (g - 1)h(1) \right) + h(1) \frac{f''(1)D'(1) - f'(1)D''(1)}{2(D'(1))^2}.$$

From (5), we observe that  $f'(1)/D'(1) = 1/h(1)$ . Thus,

$$X'(1) = (B'(1) - 1) \frac{h'(1)}{h(1)} + g - 1 + \frac{f''(1)D'(1) - f'(1)D''(1)}{2D'(1)f'(1)}. \tag{9}$$

The remaining term we need to find here is  $h'(1)/h(1)$ . By using representation (3), we get that

$$\frac{h'(1)}{h(1)} = \frac{\left( \prod_{k=1}^{g-1} (y - y_k) \right)'}{\prod_{k=1}^{g-1} (y - y_k)} \Bigg|_{y=1} = \sum_{k=1}^{g-1} \frac{1}{1 - y_k}. \tag{10}$$

Thus, in terms of roots, we have that

$$X'(1) = (B'(1) - 1) \sum_{k=1}^{g-1} \frac{z_k}{z_k - B(z_k)} + g - 1 + \frac{f''(1)}{2f'(1)} - \frac{D''(1)}{2D'(1)}. \tag{11}$$

### 2.4 Expectation expressed in terms of contour integrals

We can use a contour integral to find  $\sum_{k=1}^{g-1} z_k/(z_k - B(z_k))$ . Let  $\varepsilon$  be as above. Then

$$\sum_{k=1}^{g-1} \frac{z_k}{z_k - B(z_k)} = \frac{1}{2\pi i} \oint_{S_{1+\varepsilon}} \frac{D'(z)}{D(z)} \frac{z}{z - B(z)} dz - res_1, \tag{12}$$

where

$$res_1 = \frac{1}{1 - B'(1)} \left( 1 + \frac{B''(1)}{2(1 - B'(1))} + \frac{D''(1)}{2D'(1)} \right) \tag{13}$$

is the residue of the function  $\frac{D'(z)}{D(z)} \frac{z}{z - B(z)}$  at 1. By Corollary 1, we know that there are no roots of the equation  $z = B(z)$  inside  $\bar{D}_1$  except 1. Thus, we do not need to subtract any other residues of the function inside the integral.

**Remark 3** When  $X(z)$  is a pgf of the queue length, it is also possible to find the variance of the queue length, i.e.,  $X''(1) + X'(1) - (X'(1))^2$ , in the same way, but since it is a lengthy expression, we omit it here. The application of contour integrals goes further. Under certain conditions, the function  $X(z)$  can be written as an exponent of a contour integral, yielding a Pollaczek integral; see [4], where authors use our factorization result (3).

Note that the expectation  $X'(1)$  represents the average queue length and is a real number. Thus, for  $B'(1) \in \mathbb{R}$ , only the real part of the integral (12) should be computed.

### 2.5 Algorithms

In this subsection, we summarize our contour-integral algorithms to compute  $X'(1)$  and the coefficients  $x_k$  using the results of Sects. 2.1, 2.2, 2.3 and 2.4. For computational questions such as how to choose the parameters  $\varepsilon$  and  $\delta$  or how to compute a complex integral, we refer to Appendix B.

**Algorithm 1** (Computation of  $X'(1)$ )

1. Find  $\varepsilon > 0$  such that there are only  $g$  roots of Eq. (2) in the disk  $\bar{D}_{1+\varepsilon}$ , using one of two ways described in Sect. B.3.
2. Compute (the real part of) the integral

$$I = \int_{-\pi}^{\pi} z(\varphi) \frac{D'(z(\varphi))}{D(z(\varphi))} \frac{z(\varphi)}{z(\varphi) - B(z(\varphi))} d\varphi,$$

where  $z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ .

3. Compute  $X'(1)$  by

$$X'(1) = \frac{(B'(1) - 1)}{2\pi} I + \frac{B''(1)}{2(1 - B'(1))} + g + \frac{f''(1)}{2f'(1)}. \tag{14}$$

**Remark 4** Note that we parametrized  $S_{1+\varepsilon}$  by  $z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ . Therefore,  $dz = izd\varphi$  and

$$\oint_{S_{1+\varepsilon}} F(z) dz = i \int_{-\pi}^{\pi} F(z(\varphi)) z(\varphi) d\varphi, \tag{15}$$

where  $F(z)$  represents the integration function.

**Algorithm 2** (Computation of  $x_k, k = 0, \dots, g - 1$ )

1. Find  $\varepsilon > 0$  such that there are only  $g$  roots of Eq. (2) in the disk  $\bar{D}_{1+\varepsilon}$ , using one of two ways described in Sect. B.3.
2. Find  $\delta > 0$  such that there are no roots of Eq. (2) in the disk  $\bar{D}_\delta$ , using one of two ways described in Sect. B.2.
3. Compute  $\eta_k$ , using

$$\eta_k = -1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} z \frac{D'(z)}{D(z)} \left( \frac{B(z)}{z} \right)^k - w \frac{D'(w)}{D(w)} \left( \frac{B(w)}{w} \right)^k d\varphi,$$

where, for convenience,  $z = z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ ,  $w = w(\varphi) = \delta e^{i\varphi}$ .

4. Compute  $\sigma_k$  by (8) using  $\eta_k$ .
5. Compute  $x_k, k = 0, \dots, g - 1$ , using (4) together with (5).

In Appendix B.1, we consider a simplified algorithm for the case when  $B(z)$  has no roots inside  $\bar{D}_1$ , which is equivalent to  $B(-1) > 0$  if  $B(z)$  is a pgf. Fortunately, in many applications, the function  $B(z)$  has this property. Therefore, it is often better to use the simplified algorithm as described in Appendix B.1.



### 2.6 Method application in queueing theory

In this subsection, we give some remarks concerning the application of our method. First, we study a special subclass of functions  $X(z)$  of the form (1) for which it is easy to check Assumption 2. This subclass generalizes the form of the queue-length pgf for the bulk-service and FCTL-type queues considered in Sects. 3 and 4. Then, we describe the application to queueing systems that do not have a pgf of the form (1).

Suppose that

$$D(z) = z^g - A(z) \tag{16}$$

and the functions  $A(z), B(z)$  are pgfs. Consider the following assumption.

**Assumption 3** (*Stability assumption*) The functions  $A(z)$  and  $B(z)$  are pgfs and satisfy  $A'(1) < g$  and  $B'(1) < 1$ .

The following two lemmas show that Assumptions 1 and 3 imply Assumption 2. Therefore, if the denominator of  $X(z)$  has the form (16), and Assumptions 1 and 3 hold, we can apply our method.

**Lemma 3** Consider a pgf  $A(z)$  that is analytic in the disk  $D_{1+\rho}$  for some  $\rho > 0$ . Suppose that  $A'(1) < g$ . Then, the equation  $z^g - A(z) = 0$  has exactly  $g$  roots  $z_0 = 1, z_1, \dots, z_{g-1}$  in  $D_1$ .

**Lemma 4** Consider a pgf  $B(z)$  that is analytic in the disk  $D_{1+\rho}$  for some  $\rho > 0$ . Suppose that  $B'(1) < 1$ . If  $z \neq w$  and  $z, w \in \bar{D}_1$ , then

$$zB(w) \neq wB(z).$$

For the proof of Lemma 3, we refer to [1]. The proof of Lemma 4 is given in Appendix A.

**Remark 5** In our examples, the second part,  $B'(1) < 1$ , of the stability assumption will follow from the first part,  $A'(1) < g$ . However, for the general case, we have to assume that  $B'(1) < 1$  to prove the second part of Assumption 2; see the proof of Lemma 4.

Note that for many queueing systems, it was proven that the unknown part of the pgf can be expressed in terms of the function

$$\tilde{h}(z) = \prod_{k=1}^{g-1} \frac{z - z_k}{1 - z_k},$$

where  $z_k, k = 1, \dots, g - 1$  are the only roots of a certain equation in  $\bar{D}_1 \setminus \{1\}$ . For such systems, our approach can be used to find the pgf, see Eqs. (4), (7) and (8), or the expectation, see (10) and (12). One of the examples is a discrete-time  $G/G/1$  queue; see [18]. For the  $G/G/1$  queue, the generating function of the limiting waiting-time distribution depends either on  $\tilde{h}(z)$  or on  $1/(z^g \tilde{h}(1/z))$ , depending on whether the pgf of the inter-arrival or service time is a rational function, i.e., is equal to  $A_1(z)/A_2(z)$ ,

where  $A_1(z)$  and  $A_2(z)$  are polynomials of (at most) power  $g$ . Similar results hold for the queue-length pgf in the case of a bulk-service queue with a random bulk size. The reason for this is that these two systems are governed by the same equation:

$$X_{n+1} = \max\{X_n + A_n - B_n, 0\}.$$

For the bulk-service queue,  $X_n$  is the queue length at time  $n$ , and  $A_n$  and  $B_n$  are the numbers of arrivals and departures at time  $n$ , respectively. For the  $G/G/1$  queue,  $X_n$  and  $A_n$  are the waiting time and the service time of the  $n$ th customer, respectively, and  $B_n$  is the inter-arrival time between the  $n$ th and  $(n + 1)$ st customers.

### 3 Method comparison

In this section, we consider a discrete-time bulk-service queue. For this example, we compare our contour-integral approach in terms of speed and reliability with the root-finding and matrix-analytic methods. In Sect. 3.1, we describe the model. Then, we apply the considered methods in Sects. 3.2, 3.3 and 3.4. Finally, we present the numerical results in Sect. 3.5 and discuss the differences between the approaches in Sect. 3.6.

#### 3.1 Discrete bulk-service queue

Consider a discrete-time queue with bulk service. This means that in each time unit the server serves  $g$  customers in the queue, and during this time  $A_b$  new customers arrive with a pgf  $A_b(z)$ . The server serves only those customers that are present in the queue at the beginning of the time unit. If there are fewer than  $g$  customers in the queue, all of them are served.

Denote by  $X_b(z)$  the pgf of the queue length at the beginning of a time unit in the stationary state. Let  $q_k$  be the stationary probability of having  $k$  customers in the queue at the beginning of a time unit. Then (see, for example, [3]), we obtain

$$X_b(z) = \frac{\sum_{k=0}^{g-1} q_k (z^g - z^k)}{z^g - A_b(z)} A_b(z). \quad (17)$$

The stability condition in this model is  $A'_b(1) < g$ . If we consider the pgf  $X_b^-(z)$  of the queue length after service but before arrivals, we get

$$X_b^-(z) = \frac{\sum_{k=0}^{g-1} q_k (z^g - z^k)}{z^g - A_b(z)}.$$

In the following subsections, we apply our method, the root-finding and matrix-analytic approaches to find the expectation  $(X_b^-)'(1)$ .

### 3.2 Our method

At first glance, the pgf of the queue length, see (17), is not in the form (1), but we can rearrange it as follows:

$$X_b(z) = \frac{\sum_{k=0}^{g-1} q_k(z^g - z^k)}{z^g - A_b(z)} A_b(z) = \frac{\sum_{l=0}^{g-1} (\sum_{k=0}^l q_k) z^l}{z^g - A_b(z)} (z - 1) A_b(z).$$

Hence, we take  $A(z) = A_b(z)$ ,  $B(z) = 1$ ,  $f(z) = (z - 1)A_b(z)$  and  $x_k = \sum_{l=0}^k q_l$ . Note that  $q_k = x_k - x_{k-1}$ ,  $k = 2, \dots, g - 1$ , and  $q_1 = x_1$ . In the case  $A_b(0) = 0$ , we can reduce the complexity as before, since  $q_l$ , as the probability of having  $l$  customers in the queue, is zero for  $l = 0, \dots, n - 1$ , where  $n$  is the least possible number of customers that arrive in one time slot. Thus, we suppose that  $A_b(0) \neq 0$ . Hence,  $z_k \neq 0$  for every  $k = 1, \dots, g - 1$  and

$$f(z_k) = A_b(z_k)(z_k - 1) = z_k^g(z_k - 1) \neq 0.$$

If the system is stable, then  $A'(1) = A'_b(1) < g$ . Since  $B'(1) = 0 < 1$ , the stability assumption holds. Therefore, we can use our method if  $A_b(z)$  is analytic in the disk  $D_{1+\alpha}$  for some  $\alpha > 0$ . For the case of  $X_b^-(z)$ , the only change would be that  $f(z) = z - 1$ .

From (14), we get that the expected queue length is given by the following formula:

$$X'_b(1) = -\frac{I}{2\pi} + g + A'_b(1), \tag{18}$$

where  $I$  is given by

$$I = \int_{-\pi}^{\pi} \frac{gz^g - zA'_b(z)}{z^g - A_b(z)} \frac{z}{z - 1} d\varphi, \tag{19}$$

$z = z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ , and  $\varepsilon$  is chosen as in Theorem 1. Note that (18) and (19) are equivalent to the following formula:

$$X'_b(1) = A'_b(1) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{gz^g - zA'_b(z)}{z^g - A_b(z)} \frac{1}{z - 1} d\varphi.$$

Thus, for the case of  $X_b^-(z)$ , we get

$$(X_b^-)'(1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{gz^g - zA'_b(z)}{z^g - A_b(z)} \frac{1}{1 - z} d\varphi. \tag{20}$$

### 3.3 Root-finding approach

The root-finding method requires a numerical algorithm to find the roots of (2). Then, one needs to pick the roots that are inside the unit circle, namely the roots  $z_1, \dots, z_{g-1}$ .

The coefficients  $x_k$  in (1) then follow from a linear system of equations. For the case of the bulk-service queue, we get

$$\begin{pmatrix} 0 & 1 & \dots & g-1 \\ 1 & z_1 & \dots & z_1^{g-1} \\ \dots & \dots & \dots & \dots \\ 1 & z_{g-1} & \dots & z_{g-1}^{g-1} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \dots \\ x_{g-1} \end{pmatrix} = \begin{pmatrix} g - A'_b(1) \\ 0 \\ \dots \\ 0 \end{pmatrix}. \tag{21}$$

Note that the matrix is singular if two roots coincide. If the roots are close, the resulting values of  $x_l$  depend heavily on the precision of the roots. Given values of  $x_l$ , we compute the probabilities  $q_k = x_k - x_{k-1}$ , where  $x_{-1} = 0$ , and find the expected queue length as follows:

$$(X_b^-)'(1) = \frac{\sum_{k=0}^{g-1} q_k(g(g-1) - k(k-1))}{2(g - A'_b(1))} - \frac{g(g-1) - A''_b(1)}{2(g - A'_b(1))}. \tag{22}$$

Alternatively to solving (21), we can directly use the roots; see Eq. (11). For the bulk-service, we get

$$(X_b^-)'(1) = - \sum_{k=1}^{g-1} \frac{z_k}{z_k - 1} + g - 1 - \frac{g(g-1) - A''_b(1)}{2(g - A'_b(1))}. \tag{23}$$

The possible drawbacks of both approaches are that the root-finding algorithm can fail to find exactly  $g - 1$  roots inside the unit circle and that the precision of the roots can be unsatisfactory.

### 3.4 Matrix-analytic approach

It is possible to represent the bulk-service queue as an  $M/G/1$ -type queue; see [13]. Consider the Markov chain of the queue length at service completion, i.e., before the arrivals. Then, the transition matrix is a block matrix

$$P = \begin{pmatrix} B_0 & A_2 & A_3 & \dots \\ A_0 & A_1 & A_2 & \dots \\ 0 & A_0 & A_1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where

$$B_0 = \begin{pmatrix} \sum_{i=0}^g a_i & a_{g+1} & \dots & a_{2g-1} \\ \sum_{i=0}^{g-1} a_i & a_g & \dots & a_{2g-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^2 a_i & a_2 & \dots & a_g \end{pmatrix}, \quad A_j = \begin{pmatrix} a_{jg} & a_{jg+1} & \dots & a_{jg+g-1} \\ a_{jg-1} & a_{jg} & \dots & a_{jg+g-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{jg-g+1} & a_{jg-g+2} & \dots & a_{jg} \end{pmatrix},$$

and  $a_k$  is the probability of having  $k$  arrivals, with  $a_k = 0$  for  $k < 0$ . Using the notation for  $M/G/1$ -type queues, instead of the queue length  $l$  we consider a two-dimensional process  $(n, m)$ , where the level state  $n$  is equal to the integer part of the quotient  $l/g$ , i.e.,  $\lfloor l/g \rfloor$ , and the background state  $m$  is the remainder of such a division, i.e.,  $l - \lfloor l/g \rfloor g$ .

The steady-state distribution is found using the matrix  $G$  defined as the minimal nonnegative solution of the following equation:

$$G = \sum_{j=0}^{\infty} A_j G^j.$$

For numerical computation, this infinite summation should be truncated. The stationary probabilities  $\pi_n = (\pi_{ng}, \dots, \pi_{ng+g-1})$  for level  $n$  are computed recursively using Ramaswami’s formula, see [8]:

$$\pi_n = \sum_{k=0}^{n-1} \pi_k \bar{A}_{n+1-k} (I - \bar{A}_1)^{-1}, \quad n = 1, 2, \dots,$$

where

$$\bar{A}_n = \sum_{k=n}^{\infty} A_{n+k} G^k,$$

and  $\pi_0$  follows from the following system of linear equations:

$$\begin{aligned} \pi_0 \bar{B}_0 &= \pi_0, \\ \pi_0 \left( \mathbf{e} + \mathbf{e} \left( I - \sum_{i=1}^{\infty} \bar{A}_i \right)^{-1} \sum_{i=1}^{\infty} \bar{A}_i \right)^T &= 1, \end{aligned}$$

with  $\bar{B}_0 = B_0 + \bar{A}_2 G$  and  $\mathbf{e} = (1, \dots, 1)$ . The average queue length is approximated by

$$(X_b^-)'(1) \approx \sum_{k=1}^{gN} k \pi_k, \tag{24}$$

where the value for  $N$  can be chosen so that the contribution of the vector  $\pi_N$  is small, i.e.,

$$\sum_{k=0}^{g-1} (gN + k) \pi_{gN+k} < \varepsilon_M.$$

In [15], an alternative way was proposed to find the average queue length for  $M/G/1$ -type queues. For the bulk-service queue, we can use a simplified version of this approach. The average queue length is computed using  $\pi_0$  and  $\pi_* = \sum_{i=1}^{\infty} \pi_i$ , which can be found from the following system of linear equations:

$$\begin{aligned} \boldsymbol{\pi}_0 \mathbf{e}^T + \boldsymbol{\pi}_* \mathbf{e}^T &= 1, \\ \boldsymbol{\pi}_0 \left( B_0 - \sum_{n=3}^{\infty} \bar{A}_n G \right) + \boldsymbol{\pi}_* \left( A_0 - \sum_{n=2}^{\infty} \bar{A}_n G \right) &= \boldsymbol{\pi}_0, \\ \boldsymbol{\pi}_0 \sum_{n=2}^{\infty} \bar{A}_n + \boldsymbol{\pi}_* \sum_{n=1}^{\infty} \bar{A}_n &= \boldsymbol{\pi}_*. \end{aligned}$$

Note that the average queue length is the sum of the average level and the average background state, i.e.,

$$(X_b^-)'(1) = \sum_{k=1}^{\infty} gk\boldsymbol{\pi}_k \mathbf{e}^T + \boldsymbol{\pi}_0 \mathbf{b}^T + \boldsymbol{\pi}_* \mathbf{b}^T, \tag{25}$$

where  $\mathbf{b} = (0, 1, \dots, g - 1)$ . Thus, it is only left to find  $\mathbf{r} = \sum_{k=1}^{\infty} k\boldsymbol{\pi}_k$  using

$$\begin{aligned} \mathbf{r} \left( I - \sum_{k=0}^{\infty} A_k \right) &= \boldsymbol{\pi}_0 \left( \sum_{k=2}^{\infty} (k - 1) A_k \right) + \boldsymbol{\pi}_* \left( \sum_{k=2}^{\infty} (k - 1) A_k - A_0 \right), \\ \mathbf{r} \left( A_0 - \sum_{k=2}^{\infty} (k - 1) A_k \right) \mathbf{e}^T &= (\boldsymbol{\pi}_0 + \boldsymbol{\pi}_*) \sum_{k=2}^{\infty} \frac{k(k - 1)}{2} A_k \mathbf{e}^T. \end{aligned}$$

In the following subsection, we numerically compare the considered approaches in terms of speed and reliability.

### 3.5 Numerical results

In this subsection, we compare our approach with the root-finding and matrix-analytic methods for the bulk-service queue. We focus on speed and reliability, i.e., the absence of failures such as a wrong number of roots, or unrealistic result (negative, imaginary or not accurate queue length); see Table 1. We generate 10,000 random parameters  $g = 2, \dots, 30, c = g + 1, \dots, 70$  and load  $\rho \in [0, 0.99]$ . We set  $A_b(z) = (\rho g z / c + 1 - \rho g / c)^c$ , i.e., we consider binomial arrivals. We have used standard *python* methods for computing integrals, finding polynomial roots and solving linear systems of equations.

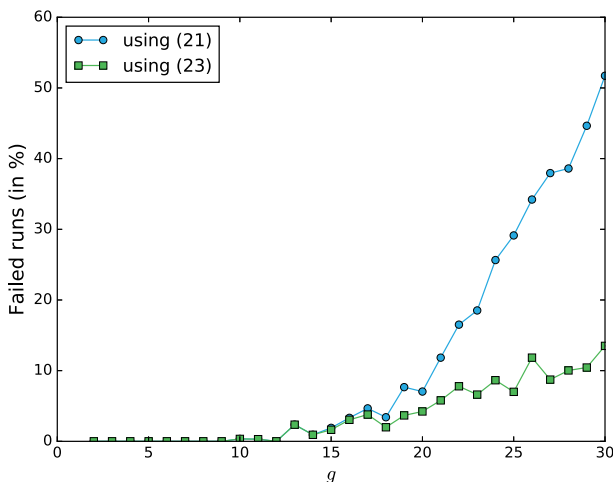
For the root-finding technique, the use of the linear system (21) leads to poor reliability. In almost 12% of the cases, the approach failed to give a realistic result, where in 3% of the cases, the number of roots in the unit circle was wrong, and there was no result at all. The root-finding algorithm together with (23) is two times faster and fails in only 4% of the cases. Note that the failures occur more often for large batch sizes; see Fig. 2. For a batch size of 30, the root-finding method with (21) fails in more than half of the cases. Our approach has no fails for all values of batch sizes and produces realistic results.

The matrix-analytic approach always produces a nonnegative queue length. However, the use of (24) leads to a serious underestimation of the queue length in 10% of the cases. This happens due to slow convergence of the distribution tail for high loads

**Table 1** Comparison of speed and reliability for the expected queue length

Methods	Root-finding using (21)	Root-finding using (23)	Matrix-analytic using (24)	Matrix-analytic using (25)	Our approach
Speed (ms. per run)	7.15	3.37	17.73	17.69	0.99
Absolute difference					
> 0.0001	937	95	4222	0	0
> 0.5	305	0	1018	0	0
Total failures	1179	389	0	0	0
Wrong number of roots	329	329	N.A.	N.A.	N.A.
$\text{Re}(\mathbb{E}L) < -0.0001$	241	60	0	0	0
$\text{Re}(\mathbb{E}L) < -0.5$	112	0	0	0	0
$ \text{Im}(\mathbb{E}L)  > 0.0001$	788	0	0	0	0
$ \text{Im}(\mathbb{E}L)  > 0.5$	258	0	0	0	0

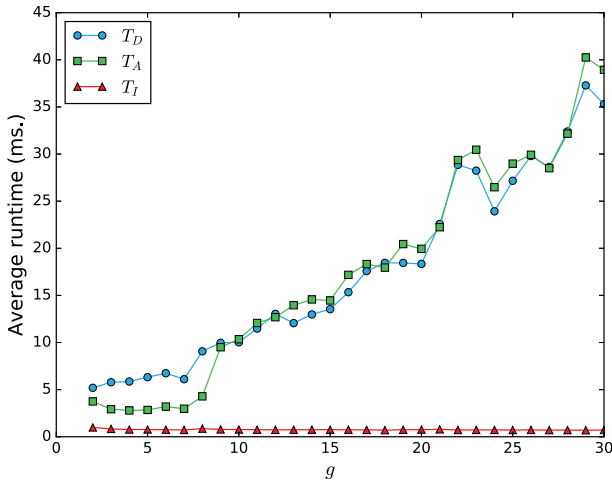
There are 10,000 random combinations of  $g$ ,  $c$  and  $\lambda$ . For the speed reference of our method, only the real part of the integral was calculated. The real and imaginary parts of the expected queue length are denoted by  $\text{Re}(\mathbb{E}L)$  and  $\text{Im}(\mathbb{E}L)$ , respectively. The absolute difference is calculated between our and other approaches



**Fig. 2** The percentage of failed runs for root-finding algorithm depending on the batch size  $g$

( $\rho > 0.9$ ). The calculation of the aggregated distribution (given by  $\pi_0, \pi_*$ ) leads to results that (almost) coincide with the results of our approach, suggesting that both approaches produce accurate results. The main drawback is the slow computational speed, which is mainly due to the computation of the matrix  $G$ , which we find by recursively applying

$$G_{n+1} = \sum_{k=0}^{\infty} A_k G_n^k, \quad n = 0, 1, \dots$$



**Fig. 3** The average runtime as a function of  $g$  for the matrix-analytic approach using (24),  $T_D$ , or (25),  $T_A$ , and for our integral approach (20),  $T_I$

with  $G_0 = 0$ , until absolute values of entries in the matrix  $G_{n+1} - G_n$  are less than  $\varepsilon_M = 0.1^{10}$ . The runtime is longer for bigger sizes of the matrices, i.e., for bigger  $g$ ; see Fig. 3.

We conclude that our contour-integral method outperforms the considered approaches both in speed and reliability. In the following subsection, we summarize the advantages of our method compared to the existing approaches.

### 3.6 Discussion

The main contribution of the paper is our *explicit* method of finding the queue-length distribution. Both the root-finding approach and the matrix-analytic approach require the computation of *implicitly* defined variables, i.e., the roots  $z_1, \dots, z_{g-1}$  or matrix  $G$ . For the root-finding approach, this step is error-prone, resulting in poor reliability of the method, while for the matrix-analytic approach, the main issue is the speed. Note that explicit closed-form results are much easier to manipulate to show a number of structural results such as convexity and sensitivity to input parameters or asymptotic behavior.

In Sect. 3.5, we used a simple example of a bulk-service queue with binomial arrivals. For such arrivals, the pgf  $A_b(z)$  is a polynomial, and  $A_n = 0$  for  $n \geq c/g + 1$ . Thus, we use standard *python* methods to find the roots of the polynomials and do not need to truncate the arrival distribution to find the matrix  $G$ , which might result in an underestimation of the queue length. For our approach, the required pgf  $A_b(z)$  is usually easy to compute even if it is not a polynomial; see Sect. 4.5 below. Therefore, we do not have to choose a truncation bound as in the existing methods.

We note that, for certain arrival distributions, there are formulas to compute the roots; see, for example, [10]. These formulas contain infinite summations, which should be truncated for numerical results. As we saw in Sect. 3.5, a poor precision



of the roots may lead to unrealistic results, which can also happen in the case of a badly chosen truncation bound. For the case of the bulk-service queue, root-free results were obtained in [9]. The average queue length is represented as a (double) infinite summation, which also leads to a truncation problem.

Our results are applicable to a quite general class of queueing systems. This includes several  $M/G/1$ -type queues, such as bulk-service queues and FCTL-type queues described in Sect. 4, but also discrete  $G/G/1$  queues, see Sect. 2.6, for which the matrix-analytic approach is not applicable.

## 4 FCTL-type queues

In this section, we consider several traffic-light models and apply our method to them. First, we explain the basic FCTL model in detail and use our approach to find the average queue length; see Sect. 4.1. Then, we propose several modifications of this model to include the following realistic situations:

1. the lane serves a turning flow, so vehicles slow down before departing;
2. pedestrian and/or bicycle traffic lights have an actuated control;
3. just after the intersection there is a bridge or a railway and a part of the green time may be lost;
4. another lane at the intersection does not have a fixed length of the green time due to an actuated control;
5. times between departures differ due to driver distraction and/or vehicle condition and length.

In Sect. 4.2, we consider the turning flow and prove a decomposition result. Even though situations 2–4 differ a lot, they can be modeled in the same way by assuming that the green and red times consist of a random number of time intervals; see Sect. 4.3. In Sect. 4.4, we suggest a way to model the distraction of the drivers. We conclude this section by presenting numerical examples in Sect. 4.5.

### 4.1 FCTL model

Consider an isolated intersection with fixed-cycle traffic-light control, i.e., the green and red times of each lane are fixed. This system is modeled using a fixed-cycle traffic-light (FCTL) model, which is a basic model in traffic theory and has been extensively studied; see, for example, [6,17].

Due to the fact that the green and red times are fixed, the queues at different lanes are independent and can be considered separately. Thus, we focus on one approaching lane and consider the queue length for this lane. The analysis for the other lanes is the same. Suppose that each delayed vehicle needs the same time  $\tau$  to depart from the lane. Thus, if there is a long queue at the beginning of the green time, we see a departure every  $\tau$  seconds. In what follows, we suppose that time is split in time intervals, each of  $\tau$  seconds. The effective green time consists of  $g \in \mathbb{N}$  time intervals and the effective red time of  $r \in \mathbb{N}$  time intervals. Thus, not more than  $g$  delayed vehicles can depart during the green time. We suppose that each cycle starts with  $g$

green time intervals and then switches to  $r$  red time intervals, where by green and red time intervals we mean time intervals during green and red time, respectively. Together, this gives  $c = g + r$  time intervals in a cycle.

Let  $Y_{n,m}$  be the number of arrivals during the  $n$ th time interval of the  $m$ th cycle with pgf  $Y_{n,m}(z)$ . We make the following two assumptions:

**Assumption 4** (*Independent arrivals assumption*) The numbers of arrivals  $Y_{n,m}$ , for  $m = 1, 2, \dots, n = 0, \dots, c - 1$ , are identically distributed and independent of each other.

Thus, we can denote  $Y_{n,m}$  simply as  $Y$  and  $Y_{n,m}(z)$  as  $Y(z)$ . This assumption is realistic for an intersection that lies far enough from other signal-controlled intersections; for example, if the intersection is isolated. However, if the distance between intersections is short, then the vehicles arrive in platoons and this assumption does not hold. Following [17], we add the so-called FCTL assumption:

**Assumption 5** (*FCTL assumption*) If a vehicle arrives during the green time and finds an empty queue, then it proceeds without delay.

Assumption 5 means that if the queue is cleared before the end of the green time, the vehicles that arrive during the remaining green time will immediately depart. Therefore, the queue length will be 0 until the end of the green time. This assumption is quite realistic for a straight-going flow since vehicles that find no queue will proceed without stopping and, thus, are able to depart at the free-flow speed. However, for turning flows, this assumption is not realistic, especially for right turn (or left turn in the case of left-hand traffic), because the turning vehicles need to slow down. Thus, in the next subsection, we also consider the FCTL model with the one-vehicle assumption (see Assumption 6 below) instead of the FCTL assumption.

Denote by  $X_n(z)$  the pgf of the queue length in the stationary state at the beginning of the  $n$ th time interval during an arbitrary cycle. The pgf exists in the case of a stable system, i.e., when the possible number of departures is greater than the expected number of arrivals:

$$cY'(1) < g.$$

Let  $q_n$  be the stationary probability of having an empty queue at the beginning of the  $n$ th time interval during a cycle, i.e.,  $q_n = X_n(0)$ . Then, under Assumptions 4 and 5, we get

$$\begin{aligned} X_{n+1}(z) &= \frac{X_n(z) - q_n}{z} Y(z) + q_n, & \text{for } n = 0, \dots, g - 1, \\ X_{n+1}(z) &= X_n(z)Y(z), & \text{for } n = g, \dots, c - 2, \\ X_0(z) &= X_{c-1}(z)Y(z). \end{aligned} \quad (26)$$

The first equation in (26) is based on the fact that if there is at least one vehicle in the queue at the beginning of the green time interval, then one vehicle departs and  $Y$  vehicles arrive. In terms of pgfs, this means multiplying by  $Y(z)/z$ . However, if the queue is empty, it remains empty throughout the time interval. Thus,  $q_n$  should not be multiplied by any additional function. During the red time,  $Y$  vehicles arrive each time interval and none depart. Therefore, we only multiply by  $Y(z)$  in the final two equations of (26).

This gives us, after some rearrangement, the pgf of the overflow queue, defined as the queue length at the beginning of the red time:

$$X_g(z) = \frac{\sum_{k=0}^{g-1} q_k z^k (Y(z))^{g-1-k}}{z^g - (Y(z))^c} (z - Y(z)). \tag{27}$$

The expression has the form (1), where  $D(z) = z^g - A(z)$ ,  $A(z) = (Y(z))^c$ ,  $B(z) = Y(z)$ ,  $f(z) = z - Y(z)$  and  $x_k = q_k$ . Note that by Corollary 1,  $f(z) = 0$  inside and on the unit circle only for  $z = 1$ . Thus,  $f(z_k) \neq 0$  for each root  $z_k$  of (2),  $k = 1, \dots, g - 1$ . We need to assume that  $Y(z)$  is an analytic function in the disk  $D_{1+\alpha}$  for some  $\alpha > 0$ . Then, Assumption 1 holds. Since we consider only stable systems, we assume that  $A'(1) = cY'(1) < g$ , which means that  $B'(1) = Y'(1) < 1$ , and, therefore, Assumption 3 holds.

One can prove, see, for example, [6], that the expected overflow queue  $\mathbb{E}X_g = X'_g(1)$  and the expected queue length at an arbitrary point during the cycle are connected in the following way:

$$\begin{aligned} \mathbb{E}L_{\text{FCTL}} &= \frac{r}{c(1 - Y'(1))} X'_g(1) + \frac{r^2 Y'(1)}{2c(1 - Y'(1))} \\ &+ \frac{r(Y''(1) + Y'(1) - (Y'(1))^2)}{2c(1 - Y'(1))^2}. \end{aligned} \tag{28}$$

Now consider  $X'_g(1)$ . From (14), we get the following theorem.

**Theorem 2** For the FCTL model, the overflow queue is given by

$$X'_g(1) = \frac{(Y'(1) - 1)}{2\pi} I + g, \tag{29}$$

where

$$I = \int_{-\pi}^{\pi} \frac{gz^g - zc(Y(z))^{c-1}Y'(z)}{z^g - (Y(z))^c} \frac{z}{z - Y(z)} d\varphi, \tag{30}$$

$z = z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ , and  $\varepsilon$  is chosen as described in Theorem 1.

Plugging (29) in (28) gives, after rearrangement, the following formula:

$$\mathbb{E}L_{\text{FCTL}} = -\frac{r}{2\pi c} I + \frac{gr}{c} + \frac{r(c + g + 1)Y'(1)}{2c(1 - Y'(1))} + \frac{rY''(1)}{2c(1 - Y'(1))^2}. \tag{31}$$

### 4.2 FCTL model for the turning flow

For some cases, it may turn out that the FCTL assumption does not hold. For example, in the case of a right turn at an intersection (or a left turn in the case of left-hand traffic), vehicles that find the queue empty need to decelerate almost to the speed of a delayed vehicle. Thus, rather than the FCTL assumption, it is logical to make the following assumption:

**Assumption 6** (*One-vehicle assumption*) If a number of vehicles arrives during the green time interval and finds the queue empty, then only one of the vehicles proceeds without delay, and the others form a queue.

This assumption means that even if the queue was cleared during the previous green time intervals, the overflow queue may be nonempty. The stability condition in this model coincides with the stability condition of the FCTL model under the FCTL assumption. Equations (26) will change to

$$\begin{aligned} X_{n+1}(z) &= X_n(z) \frac{Y(z)}{z} + q_n \left(1 - \frac{1}{z}\right) Y(0), & \text{for } n = 0, \dots, g - 1, \\ X_{n+1}(z) &= X_n(z) Y(z), & \text{for } n = g, \dots, c - 2, \\ X_0(z) &= X_{c-1}(z) Y(z). \end{aligned} \tag{32}$$

Indeed, if there are no vehicles, then  $\max\{Y - 1, 0\}$  vehicles will queue, with pgf  $(Y(z) - Y(0))/z + Y(0)$ . Therefore, we obtain for  $n < g$  that  $X_{n+1}(z) = (X_n(z) - q_n) \frac{Y(z)}{z} + q_n \left(\frac{Y(z)-Y(0)}{z} + Y(0)\right)$ . This gives us the above result after a small rearrangement. Hence, the pgf of the overflow queue is expressed as

$$X_g(z) = \frac{\sum_{k=0}^{g-1} q_k z^k (Y(z))^{g-1-k}}{z^g - (Y(z))^c} (z - 1) Y(0). \tag{33}$$

As in the case of the FCTL assumption, we can use our method for the FCTL model with the one-vehicle assumption if the system is stable, i.e.,  $g < cY'(1)$ , and the function  $Y(z)$  is analytic in some disk  $D_{1+\alpha}$ .

Note that in the case of Bernoulli arrivals, Assumptions 5 and 6 coincide. In this case,  $Y(z) = (1 - Y(0))z + Y(0)$ , and, thus,  $z - Y(z) = Y(0)(z - 1)$ . Note also that even though the equation for the roots is the same, generally the coefficients  $q_k$  may differ from those under the FCTL assumption. This happens because one of the roots is 1 and plugging it in the numerator gives 0 without resulting in an additional equation to find  $q_k$ . The required  $g$ th equation comes from the fact that  $X_g(1) = 1$ . This normalization equation has different forms for Assumptions 5 and 6. Under Assumption 5, we get that

$$\sum_{k=0}^{g-1} q_k (1 - Y'(1)) = g - cY'(1). \tag{34}$$

However, under Assumption 6, the normalization equation is of the following form:

$$\sum_{k=0}^{g-1} q_k Y(0) = g - cY'(1). \tag{35}$$

We see that these models coincide only if  $1 - Y(0) = Y'(1)$ , i.e., this happens only in the case of Bernoulli arrivals.

From Eqs. (34) and (35) and the fact that the roots are the same in both models, it follows that

$$q_{k,\text{FCTL}}(1 - Y'(1)) = q_{k,1\text{v}}Y(0),$$

where  $q_{k,\text{FCTL}}$  (resp.,  $q_{k,1\text{v}}$ ) is the stationary probability of having an empty queue at the beginning of the  $k$ th time interval during a cycle in the case of Assumption 5 (resp., 6). Moreover, by comparing Eqs. (27), (33) and (26), (32), we get that in the case of a stable system, for each  $k = 0, \dots, c - 1$ ,

$$X_{k,1\text{v}}(z) = X_{k,\text{FCTL}}(z)X_{\text{dif}}(z),$$

where

$$X_{\text{dif}}(z) = \frac{(1 - Y'(1))(z - 1)}{z - Y(z)},$$

$X_{k,\text{FCTL}}(z)$  and  $X_{k,1\text{v}}(z)$  are pgfs at the beginning of the  $k$ th time interval for the FCTL model and the one-vehicle model, respectively. Note that  $X_{\text{dif}}(z)$  is independent of  $g$  and  $c$  with  $X_{\text{dif}}(1) = 1$  and  $X'_{\text{dif}}(1) = \frac{Y''(1)}{2(1-Y'(1))}$ .

Consider the case  $g = c = 1$ . In the case of the FCTL assumption, the queue is always empty once it becomes empty since there is no red time during which vehicles can form a queue. Therefore, in the case of stability, i.e.,  $Y'(1) < 1$ , we get  $X_{0,\text{FCTL}}(z) = 1$ . Thus, we get  $X_{0,1\text{v}} = X_{\text{dif}}(z)$ , i.e.,  $X_{\text{dif}}(z)$  is the pgf of the queue with  $g = c = 1$  and arrivals  $Y(z)$ . Hence, we get the following decomposition result (compare with the decomposition results in polling systems, [5]):

**Theorem 3** *For arbitrary  $g \leq c$ , in the case of a stable system, the queue with the one-vehicle assumption can be considered as a sum of two independent queues: the FCTL queue with the same  $g, c$  and the one-vehicle queue with  $g = c = 1$ .*

Note that  $X_{\text{dif}}(z)$  can be also viewed as a pgf of the queue length of a bottleneck, for example, when part of the road has low speed limit and all the vehicles decelerate before this place.

We can as well change the model to allow more than one vehicle, but not all, to pass the junction if the queue is empty. Then, the changes will be only in the function  $f(z)$ , and we can easily use our method if Assumptions 1 and 3 hold. For the same reason, there would be a similar decomposition result. The expected queue length in this new case will be less than the expected queue length in the case of the one-vehicle assumption but more than the expected queue length in the case of the FCTL assumption.

**Remark 6** The bulk-service model can be seen as an intermediate step between the considered traffic-light models. More precisely, for  $A_b(z) = Y(z)^c$ , it can be shown using a sample path argument that the average queue length satisfies the following inequalities:

$$X'_{g,\text{FCTL}}(1) \leq (X_b^-)'(1) \leq X'_{g,1\text{v}}(1). \tag{36}$$

In traffic terms, the bulk-service model can be seen as a queue with service at the end of the green time that serves  $g$  (or fewer) vehicles and then waits one cycle for new arrivals.

From (36), the difference between the two FCTL models, described by Theorem 3, gives an upper bound on the difference between the FCTL and bulk-service models. In Sect. 4.5.2, we show that most of the time this difference is small.

### 4.3 Random green and red times

Suppose we know that with probability  $\theta_{r,g}$  the red time consists of  $r$  time intervals and the green time after this red time consists of  $g$  time intervals,  $r, g \in \mathbb{N} \cup \{0\}$ ,  $r + g \neq 0$ . Thus, cycles start with a red time, followed by a green time. The lengths of the red time and the green time are independent of the lengths of red and green times before this cycle. We assume that the length of the green time is at most  $N$ , i.e.,  $\sum_{k=0}^{\infty} \sum_{l=N+1}^{\infty} \theta_{k,l} = 0$ . For simplicity, we consider the smallest such  $N$ , so  $\sum_{k=0}^{\infty} \theta_{k,N} > 0$ . As before, in each time interval,  $Y$  vehicles arrive with pgf  $Y(z)$ . We also use the FCTL assumption. One can easily alter the formulas in the case of the one-vehicle assumption.

Let  $X(z)$  be the pgf of the overflow queue and  $X_k^{r,g}(z)$  be the (conditional) pgf of the queue length at the beginning of the  $k$ th green time unit given that there are  $r$  red time units and  $g$  green time units in this cycle. Note that when the choices of  $r$  and  $g$  are known, the pgf of the queue length changes as in (26). Thus,

$$X_0^{r,g}(z) = X(z)Y(z)^r,$$

$$X_{k+1}^{r,g}(z) = X_k^{r,g}(z) \frac{Y(z)}{z} + p_k^{r,g} \left( 1 - \frac{Y(z)}{z} \right),$$

where  $p_k^{r,g}$  is the probability of having an empty queue at the beginning of the  $k$ th green time interval in the case of a cycle with  $r$  red and  $g$  green time intervals. Hence, with probability  $\theta_{r,g}$ , the pgf of the queue length at the beginning of the next red time will be

$$X_g^{r,g}(z) = X(z) \frac{(Y(z))^{g+r}}{z^g} + \left( 1 - \frac{Y(z)}{z} \right) \sum_{k=0}^{g-1} p_k^{r,g} \left( \frac{Y(z)}{z} \right)^{g-1-k}.$$

Therefore, the pgf of the overflow queue satisfies the following equation:

$$X(z) = \sum_{r,g} \theta_{r,g} \left( X(z) \frac{(Y(z))^{g+r}}{z^g} + \left( 1 - \frac{Y(z)}{z} \right) \sum_{k=0}^{g-1} p_k^{r,g} \left( \frac{Y(z)}{z} \right)^{g-1-k} \right).$$

Isolating the function  $X(z)$  gives us

$$X(z) = \frac{\sum_{r,g} \theta_{r,g} \left( 1 - \frac{Y(z)}{z} \right) \sum_{k=0}^{g-1} p_k^{r,g} \left( \frac{Y(z)}{z} \right)^{g-1-k}}{1 - \sum_{r,g} \theta_{r,g} \frac{(Y(z))^{g+r}}{z^g}}.$$

After multiplying both numerator and denominator by  $z^N$  and making some rearrangements, we get

$$X(z) = \frac{\sum_{k=0}^{N-1} q_k(Y(z))^k z^{N-1-k}}{z^N - A(z)}(z - Y(z)),$$

where  $q_k = \sum_{r,g} \theta_{r,g} P_{g-1-k}^{r,g}$  and  $A(z) = \sum_{r,g} \theta_{r,g} (Y(z))^{g+r} z^{N-g}$ . As we see, we can apply our method if  $A(z)$  and  $Y(z)$  are analytic in some disk  $D_{1+\alpha}$ ,  $A'(1) < N$  and  $Y'(1) < 1$ .

Note that if the summation  $\Omega_1(z) = \sum_{r,g} \theta_{r,g} (g+r)(Y(z))^{g+r-1} z^{N-g}$  converges at some point  $a > 1$ , then  $\Omega_1(z)$ ,  $\Omega_1(z)Y'(z)$ ,  $A(z)$  and

$$\Omega_2(z) = \sum_{r,g} \theta_{r,g} (Y(z))^{g+r} (N-g) z^{N-g-1}$$

converge absolutely in  $D_a$ . Then, if the function  $Y(z)$  is analytic in  $D_a$ , the function  $A(z)$  is also analytic in  $D_a$  and  $A'(z) = \Omega_1(z)Y'(z) + \Omega_2(z)$ . For applications,  $A(z)$  is always a finite sum, so it is analytic if  $Y(z)$  is analytic. Note that

$$A'(1) = \sum_{r,g} \theta_{r,g} ((g+r)Y'(1) + (N-g)) = \mathbb{E}(G+R)Y'(1) + N - \mathbb{E}G,$$

where  $G$  and  $R$  are random variables that denote the length of the green and red times during a cycle. Thus, if  $A'(1) < N$ , we get  $\mathbb{E}(G+R)Y'(1) < \mathbb{E}G$ . It immediately follows that  $Y'(1) < 1$ . We see once again that the stability assumption holds in the case of a stable system.

Now, we give three examples of possible applications of this model.

### 4.3.1 Interruption by pedestrians and/or cyclists

Let us focus on one lane and, for simplicity, on one source causing interruptions. Suppose that with probability  $p$  there is an arrival of a cyclist during a cycle. Suppose that we need  $t_c$  green time intervals for switching on and off the cyclists' green time. The green time for pedestrians and cyclists can be provided using the green time  $g$  of this lane or can be added as extra time to the cycle time  $c = g + r$ . In the first case, we have  $\theta_{r,g} = 1 - p$ ,  $\theta_{r+t_c, g-t_c} = p$ , and in the second case, we have  $\theta_{r,g} = 1 - p$ ,  $\theta_{r+t_c, g} = p$ .

### 4.3.2 Interruption by trains and boats

This case is more-or-less the same as the previous one. The only difference is in the length of interruption. We assume that interruption happens during a cycle with probability  $p$ . We also assume that if it happens, then all green time is effectively red. Therefore,  $\theta_{r,g} = 1 - p$ ,  $\theta_{c,0} = p$ .

### 4.3.3 Vehicle-actuated control on another lane

Suppose that another lane has a vehicle-actuated control, and, thus, the length of its green time is not deterministic. In this case, to find an approximation for the queue length on the considered lane, we assume that the lengths of red times during different cycles are independent of each other. Then, if we know the distribution of the green time for the other lane, we get a distribution of the red time for the considered lane. Let  $\theta_{r,g} = p_r$ , where  $p_r$  is the probability of the red time length  $r$ , and  $g$  is fixed. In the case of a fixed cycle, we take  $\theta_{r,c-r} = p_r$ .

**Remark 7** The analysis of the actuated-controlled lane is complicated due to the fact that the arrival and service processes are connected. We leave this problem for future research.

### 4.4 Uncertain departure time

In this subsection, we consider the following extension of the FCTL model. Suppose that, during a green time interval, the driver of the departing vehicle may be distracted with a fixed probability  $p \in [0, 1]$ . Therefore, during the green time, even in the case of a nonempty queue, vehicles do not depart each time interval. Then, each driver has a geometrically distributed number of tries to depart from the queue. This can be modeled using the FCTL model in the following way: Suppose that during the green time there is an extra arrival with probability  $p$ . In this way, we compensate for the uncertainty in the departures. The distribution of the queue length in this model will be the same as the distribution in our new model. So, we consider the case that the number of the arrivals during the green time interval has pgf  $Y(z)(pz + (1 - p))$ . However, during the red time, we have still  $Y(z)$  arrivals. Hence, the pgf of the overflow queue is given by the following formula:

$$X_g(z) = \frac{\sum_{k=0}^{g-1} q_k z^k (Y(z)(pz + 1 - p))^{g-1-k}}{z^g - (Y(z))^c (pz + 1 - p)^g} (z - Y(z)(pz + 1 - p)).$$

Using the same arguments as in Sect. 4.1, one can see that the pgf is of the general form described in Sect. 2. Thus, it is possible to use the same techniques in the case when  $Y(z)$  is analytic in some  $D_{1+\alpha}$ , and the system is stable, i.e.,  $cY'(1) + gp < g$ .

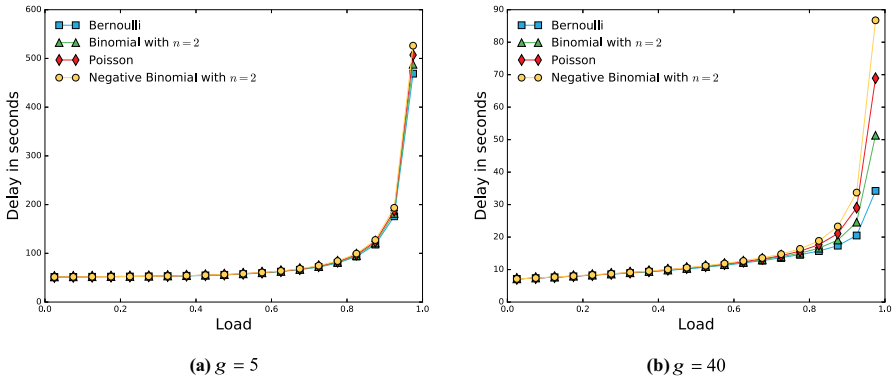
### 4.5 Numerical results

In this subsection, we present the numerical results for FCTL-type models. First, we consider the impact of the variability of the arrival process for the FCTL queue. Then, we discuss the difference between the FCTL assumption and the one-vehicle assumption. Finally, we study the impact of the traffic interruptions caused by cyclists and uncertain departure times. Further numerical results can be found in the technical report [14].



**Table 2** The considered types of the arrival process

Type of arrivals	Pgf $Y(z)$	Variance
Bernoulli	$\lambda z + (1 - \lambda)$	$\lambda - \lambda^2$
Binomial	$\left(\frac{\lambda}{n}z + 1 - \frac{\lambda}{n}\right)^n$	$\lambda - \frac{\lambda^2}{n}$
Poisson	$e^{\lambda(z-1)}$	$\lambda$
Negative binomial	$\left(\frac{n}{n+\lambda-\lambda z}\right)^n$	$\lambda + \frac{\lambda^2}{n}$



**Fig. 4** The expected delay as a function of load for Bernoulli, binomial, Poisson and negative binomial arrivals with  $c = 60, n = 2$  and **a**  $g = 5, \mathbf{b}$   $g = 40$

**4.5.1 Variability of the arrival process**

In this subsection, we show the impact of the variability of the arrival process on the queue length for the FCTL model. In what follows,  $\lambda$  denotes the arrival rate per time interval, i.e.,  $\lambda = Y'(1)$ . In Table 2, we summarize pgfs and variances of the considered arrival processes.

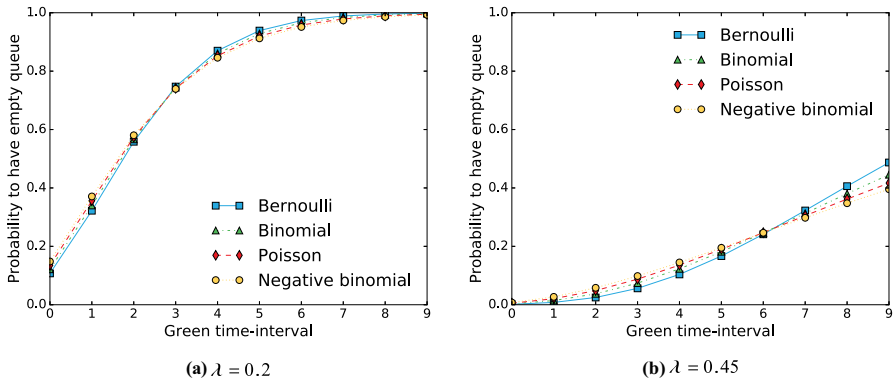
Note that Bernoulli arrivals have the least possible variance for a fixed  $\lambda$ . Indeed, the variance of an arbitrary arrival process with rate  $\lambda$  is equal to  $Y''(1)+Y'(1)-(Y'(1))^2 \geq \lambda - \lambda^2$ . In what follows, the parameter  $n$  of both the binomial and the negative binomial distribution is set to be equal to 2.

First, we focus on the expected delay per vehicle for various types of arrivals. The expected delay is computed using Little’s law  $\mathbb{E}D = \mathbb{E}L/\lambda$ , where  $\mathbb{E}L = \mathbb{E}L_{\text{FCTL}}$  is computed by (31). In Fig. 4, the expected delay is plotted as a function of the load  $\rho = c\lambda/g$  for  $c = 60$  and  $g = 5$  or  $40$ . The expected delay is given in seconds, rather than time intervals, and one time interval is set to be equal to 2 s.

Our first observation is that the higher the variability of the arrival process, the longer the expected delay. However, for each green time  $g$ , the difference is significant only for a load higher than 0.8. As we can see, for high load, the relative difference between expected delays for various types of arrivals is increasing as the green time increases. However, the absolute difference is decreasing. For load  $\rho = 0.9833$  (the highest load in the figures) and  $g = 5, 15, 30, 40$ , the absolute differences are given in Table 3. In

**Table 3** The differences (in seconds) in the expected delay for different types of arrival process,  $c = 60$  and load  $\rho = 0.9833$

Difference	$g = 5$	$g = 15$	$g = 30$	$g = 40$
$\mathbb{E}D_{\text{negbin}} - \mathbb{E}D_{\text{pois}}$	29.1472	28.6778	28.1833	27.7916
$\mathbb{E}D_{\text{pois}} - \mathbb{E}D_{\text{binom}}$	29.1369	28.6156	28.0097	27.5466
$\mathbb{E}D_{\text{binom}} - \mathbb{E}D_{\text{bern}}$	29.1258	28.5392	27.7332	27.0498



**Fig. 5** The probabilities  $q_k$  of a queue being empty,  $k = 0, \dots, g - 1$ , for  $g = 10, c = 20, n = 2$  and **a**  $\lambda = 0.2$ , **b**  $\lambda = 0.45$

the table, we use  $\mathbb{E}D_{\text{bern}}$ ,  $\mathbb{E}D_{\text{binom}}$ ,  $\mathbb{E}D_{\text{pois}}$ ,  $\mathbb{E}D_{\text{negbin}}$  to express the expected delay in the cases of Bernoulli, binomial, Poisson and negative binomial arrivals, respectively.

From the table, we see that the absolute difference is decreasing as the green time increases. In many approximations (see [12,16]), the dependence on the variance is set to be linear for the fixed arrival rate, green times and cycle times. However, we see that this difference increases for higher variance.

Next, we consider the probabilities  $q_k$  of a queue being empty for  $k = 0, \dots, g - 1$ . In Fig. 5, for  $g = 10, c = 20$  and  $\lambda = 0.2, 0.45$ , these probabilities are plotted as functions of  $k$ . For fixed  $g, c$  and  $\lambda$ , the sum of the probabilities is the same for various types of arrival, but the distributions of this sum are different. As we see, for all arrival rates, at the beginning of the green time the probability of a queue being empty is lower for the arrival processes that have lower variability, whereas at the end of green time the situation is reversed. The graphs look very different depending on the arrival rates. For low arrival rates, the graphs are concave and the difference between graphs for different arrival types is small. For high arrival rates, the graphs are convex and the relative difference between arrival types is larger.

**Remark 8** The speed and reliability of our method, which we showed in Sect. 3.5, allows us to consider the green time allocation problem for an isolated traffic-light intersection. This is the problem of choosing the green time length (value of  $g$ ) for different lanes under constraints on the total green time; see [14].

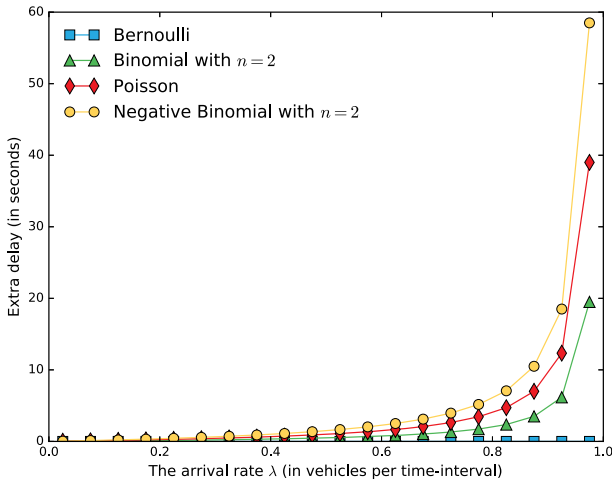


Fig. 6 The expected extra delay due to the one-vehicle assumption as compared to the FCTL assumption

### 4.5.2 Comparing the FCTL and one-vehicle assumptions

Let us first consider  $X'_{\text{dif}}(1)/\lambda$ , i.e., the expected difference in the delay between FCTL and one-vehicle models. We plot this as a function of the arrival rate  $\lambda = Y'(1)$  in Fig. 6.

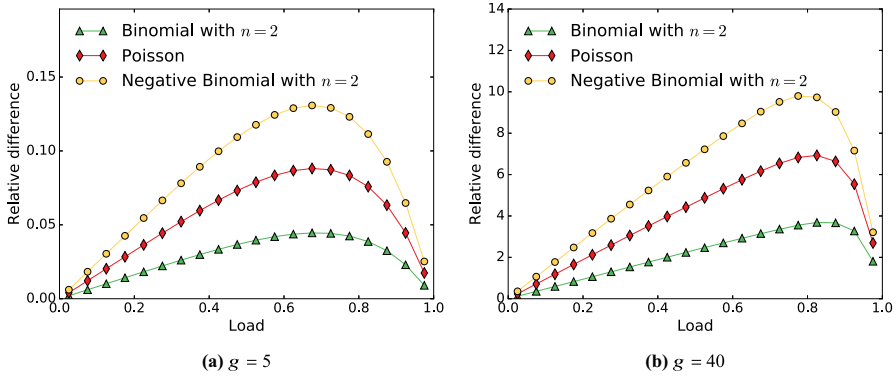
Note that if at an intersection there are at least two conflicting “main” phases, then both of them use less than half of the cycle time. Thus, in the case of a stable system, the arrival rate is lower than 0.5. If we consider, for example, Poisson arrivals, the expected extra delay is  $\frac{1}{\lambda} \cdot \frac{\lambda^2}{2(1-\lambda)} \leq \frac{1}{2(1-\lambda)} \leq 1$  time interval, i.e., not more than 2 s. Therefore, for most of the cases the extra delay will be very short. Similarly, in these cases, the extra queue length is less than one vehicle.

As we have shown, the absolute difference between FCTL and one-vehicle models is quite small. However, it is interesting to consider the relative difference. For the same settings as in Fig. 4, we plot the relative difference in the expected delay in Fig. 7.

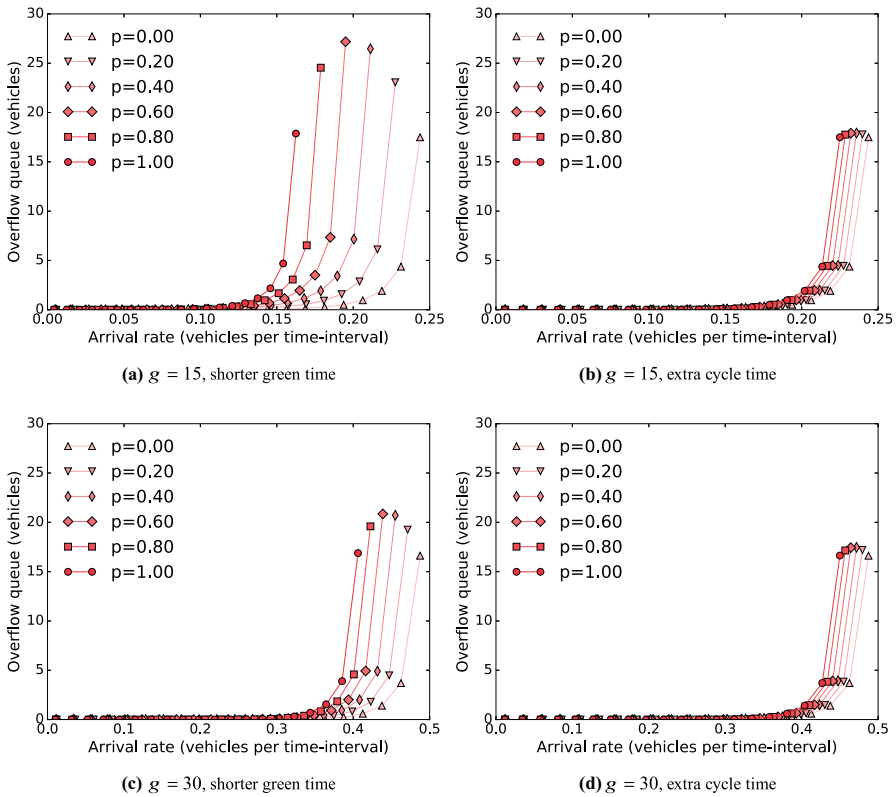
For a shorter green time, the delay observed in the FCTL model is already very long and the expected difference  $X'_{\text{dif}}(1)/\lambda$  is small. Thus, the relative difference is very small. For a longer green time, the delay is shorter and the arrival rate is higher, so the relative difference is higher. As we see, for  $g = 40$ , this relative difference reaches 10% for negative binomial arrivals with  $n = 2$ .

### 4.5.3 Disruption of the traffic

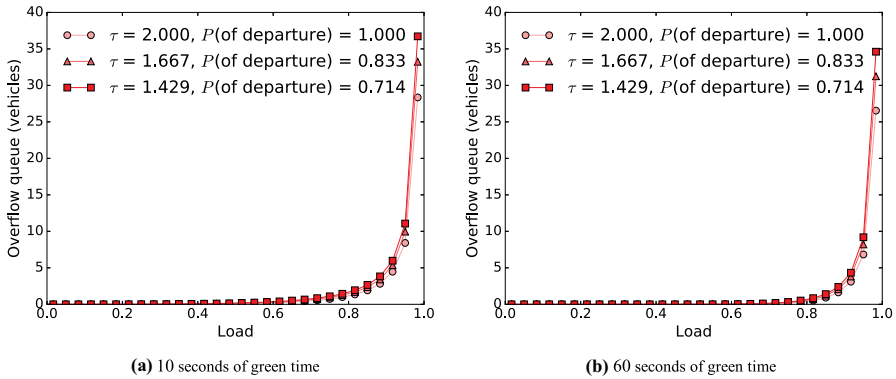
In this subsection, we consider disruption of traffic due to cyclists (or pedestrians). Suppose that cyclists need five time intervals, i.e., 10 s, to cross the road. There are two ways to provide the required green time. We can either shorten the green time of one lane (thus making room for cyclists to cross), or, alternatively, extend the total cycle time and, hence, add extra red time to all lanes. Let  $p$  be the probability of cyclists



**Fig. 7** The relative difference (in %) in the expected delay between the FCTL and one-vehicle models as a function of load for binomial, Poisson and negative binomial arrivals for  $c = 60$ ,  $n = 2$  and **a**  $g = 5$ , **b**  $g = 40$



**Fig. 8** The overflow queue in the case of cyclists as a function of arrival rate. The cycle time is  $c = 60$ . The green time in **a** and **b** is  $g = 15$  and in **c** and **d**  $g = 30$ . In **a** and **c**, the green time for cyclists (5 time intervals) is taken from the green time of the lane. In **b** and **d**, the extra time is added to the cycle



**Fig. 9** The overflow queue for FCTL model with uncertain departure times. The cycle time is 120s, and the green times are **a** 10 and **b** 60 s

arrival during the cycle. In Fig. 8, we plot the overflow queue as a function of the rate for different  $p$  and  $g$  and fixed  $c = 60$ . In each graph we plot only up to load  $\rho = 0.975$ . The arrival process is assumed to be Poisson. As we see, the first way to deal with cyclists is highly disadvantageous for the lane. It significantly decreases the capacity of the lane and increases the overflow queue. The effect is stronger for a shorter green time. The conclusion is that it is better, if possible, to increase the cycle time.

#### 4.5.4 Uncertain departure times

Consider the FCTL model with uncertain departure times that was presented in Sect. 4.4. In Fig. 9, we plot the expected overflow queue as a function of the arrival rate for different probabilities of departure. We suppose that, on average, a vehicle departs in 2s. The length of the time interval is set to be equal to  $\tau = 2p$  seconds, where  $p$  is the probability of departure. The arrival process is assumed to be Poisson. We fixed the cycle time (in seconds) and consider two different green times. Depending on  $p = 1, 5/6, 5/7$ , there are on average five departures in 5, 6 or 7 time intervals.

The uncertainty in departure times does not influence the capacity of the system but only increases the overflow queue and, consequently, the delay. The overflow queue is a bit longer for a green time of 10s (about 2 vehicles for the highest considered load 0.9833). The effect of the uncertainty becomes noticeable for loads over 0.8. Note that the average overflow queue is almost 0 for loads up to 0.7–0.8 (depending on the green time), which means that most of the vehicles are served during one cycle time after the arrival.

### 5 Conclusions

We have presented a new method to calculate the expectation and distribution of the queue length for several discrete-time queueing systems. The solution is given in terms of contour integrals without any implicitly defined variables. We successfully

applied this method to the bulk-service queue and several traffic-light models, providing insights in traffic behavior. For the FCTL model with the one-vehicle assumption, we proved a decomposition result, which allowed us to give a bound on the difference between the FCTL and bulk-service queue-length expectations, which turns out to be small for many relevant applications.

We have shown that our method outperforms the classic root-finding and matrix-analytic methods in terms of speed and reliability. Using our method, the queue-length expectation can be found more than 7 and 17 times faster compared to the root-finding and matrix-analytic methods, respectively. Moreover, we found that the root-finding approach can fail to give a realistic result. This happens more often for the case of many roots, where up to 10% of the considered cases result in no answer and 40% in the negative or nonreal (complex) expected queue length. The speed of the matrix-analytic approach depends on the size of the block matrices, which coincides with the number of roots, and can be up to 54 times slower than our approach. Future work will be on extending our integral method to a wider range of systems.

**Acknowledgements** We would like to thank the anonymous reviewers and the associate editor for their valuable remarks and suggestions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix A: Proof of Lemma 4

Recall that we need to prove that  $zB(w) \neq wB(z)$  for  $z \neq w, z, w \in \bar{D}_1$ , if  $B(z)$  is a pgf with  $B'(1) < 1$ . Consider the Taylor expansion  $\sum_{j=0}^{\infty} b_j z^j$  of  $B(z)$  around zero. Let  $\tilde{B}(z) = \sum_{j=2}^{\infty} b_j z^j$ . Then, it is equivalent to prove that  $z(b_0 + \tilde{B}(w)) \neq w(b_0 + \tilde{B}(z))$  for  $z \neq w, z, w \in \bar{D}_1$ . Now, we can reformulate the lemma in the following way: there is not more than one root of the equation  $z = a(b_0 + \tilde{B}(z))$  in  $\bar{D}_1$  for each  $a \in \mathbb{C}$ . Here,  $a$  represents the value  $w/(b_0 + \tilde{B}(w))$  for some  $w \in \bar{D}_1$ . To make such a reformulation, we need to show that  $b_0 + \tilde{B}(w) \neq 0$  for all  $w \in \bar{D}_1$ .

First, we show that  $b_0 > \tilde{B}(1)$ . Consider the derivative of  $\tilde{B}(z)$  at 1. On the one hand, according to the stability assumption, we get that  $\tilde{B}'(1) = B'(1) - b_1 < 1 - b_1 = \tilde{B}(1) + b_0$ . On the other hand,  $b_k \geq 0$  for  $k \in \mathbb{N}$ , and, thus,

$$\tilde{B}'(1) = \sum_{j=2}^{\infty} j b_j \geq \sum_{j=2}^{\infty} 2 b_j = 2\tilde{B}(1).$$

Finally, we get that  $2\tilde{B}(1) \leq \tilde{B}'(1) < \tilde{B}(1) + b_0$ . Therefore,  $b_0 > \tilde{B}(1)$ .

Now, we need to use the following small lemma:

**Lemma 5** Consider an analytic function  $C(z) = \sum_{j=0}^{\infty} c_j z^j$  in  $D_{r+\delta}$ , where  $r \in \mathbb{R}$ ,  $\delta > 0$ . Suppose that  $c_j \geq 0$  for all  $j \in \mathbb{N} \cup \{0\}$ . Then the absolute value of  $C(z)$  reaches its maximum value in  $r$ , i.e.,  $|C(z)| \leq C(r)$  for each  $z \in \bar{D}_r$ .

**Proof** Since the function  $C(z)$  is analytic in the disk  $D_{r+\delta}$ , the Taylor series converges absolutely. Therefore,  $|C(z)| \leq \sum_{j=0}^{\infty} c_j |z|^j \leq \sum_{j=0}^{\infty} c_j r^j = C(r)$ .  $\square$

As a corollary, we get that  $|\tilde{B}(z)| < b_0$  for each  $z \in \bar{D}_1$ , and, consequently, there are no solutions of the equation  $b_0 + \tilde{B}(z) = 0$  in  $\bar{D}_1$ . Hence, it is sufficient to prove that the equation

$$z = a(b_0 + \tilde{B}(z)) \tag{37}$$

has not more than one root in  $\bar{D}_1$  for each  $a \in \mathbb{C}$ .

If  $a = 0$ , we have the simple equation  $z = 0$  that has not more than one solution. If  $a \neq 0$ , we can uniquely represent it as  $te^{i\varphi}$ , where  $0 \leq \varphi < 2\pi, t > 0$ . We want to prove, for a fixed value of  $\varphi$ , that the number of roots inside the unit disk does not increase when  $t$  increases. To do so, we consider our roots as functions of  $t$  and prove that the absolute value of the root increases as  $t$  increases. Suppose  $z(t)$  is a root of Eq. (37) inside the unit disk. Consider its derivative:

$$\frac{dz}{dt} = \frac{d}{dt}(te^{i\varphi}(b_0 + \tilde{B}(z))) = e^{i\varphi}(b_0 + \tilde{B}(z)) + te^{i\varphi}\tilde{B}'(z)\frac{dz}{dt}.$$

Rearranging and plugging  $z/(t(b_0 + \tilde{B}(z)))$  instead of  $e^{i\varphi}$  give us

$$t \frac{dz}{dt} = \frac{z}{1 - \frac{z\tilde{B}'(z)}{b_0 + \tilde{B}(z)}}.$$

Thus, the derivative of  $|z(t)|^2$  is equal to

$$\frac{d(z\bar{z})}{dt} = \frac{dz}{dt}\bar{z} + \frac{d\bar{z}}{dt}z = 2\operatorname{Re}\left(\frac{dz}{dt}\bar{z}\right) = \frac{2|z|^2\operatorname{Re}\left(1 - \frac{z\tilde{B}'(z)}{b_0 + \tilde{B}(z)}\right)}{t\left|1 - \frac{z\tilde{B}'(z)}{b_0 + \tilde{B}(z)}\right|^2},$$

where  $\bar{z}$  is complex conjugate of  $z$ . To prove that this derivative is positive, we only need to prove the following lemma.

**Lemma 6** *There is some  $\delta > 0$  such that, for each  $z \in D_{1+\delta}$ ,*

$$\operatorname{Re}\left(1 - \frac{z\tilde{B}'(z)}{b_0 + \tilde{B}(z)}\right) > 0. \tag{38}$$

**Proof** First of all, it is sufficient to prove that (38) holds for each  $z \in \bar{D}_1$ . Indeed, if for some point  $z$  inequality (38) holds, then for some neighborhood of  $z$  it also holds. Since  $\bar{D}_1$  is compact, if (38) holds in  $\bar{D}_1$ , it also holds for a small neighborhood  $D_{1+\delta}$ ,  $\delta > 0$ .

Now, note that inequality (38) is equivalent to the following inequality:

$$\operatorname{Re}\left((b_0 + \tilde{B}(z))(b_0 + \tilde{B}(\bar{z})) - z\tilde{B}'(z)(b_0 + \tilde{B}(\bar{z}))\right) > 0.$$

This can be rewritten as

$$b_0^2 > \operatorname{Re} \left( b_0(z\tilde{B}'(z) - \tilde{B}(z) - \tilde{B}(\bar{z})) + (z\tilde{B}'(z) - \tilde{B}(z))\tilde{B}(\bar{z}) \right).$$

Since  $\operatorname{Re}(\tilde{B}(\bar{z})) = \operatorname{Re}(\tilde{B}(z))$ , we finally get that (38) is equivalent to

$$b_0^2 > \operatorname{Re} \left( b_0(z\tilde{B}'(z) - 2\tilde{B}(z)) + (z\tilde{B}'(z) - \tilde{B}(z))\tilde{B}(\bar{z}) \right). \tag{39}$$

Note that  $\operatorname{Re}(z) \leq |z|$  for each  $z \in \mathbb{C}$ , and, thus,

$$\begin{aligned} \operatorname{Re} \left( b_0(z\tilde{B}'(z) - 2\tilde{B}(z)) + (z\tilde{B}'(z) - \tilde{B}(z))\tilde{B}(\bar{z}) \right) \\ \leq b_0|z\tilde{B}'(z) - 2\tilde{B}(z)| + |z\tilde{B}'(z) - \tilde{B}(z)| \cdot |\tilde{B}(\bar{z})|. \end{aligned}$$

Since the functions  $z\tilde{B}'(z) - 2\tilde{B}(z)$ ,  $z\tilde{B}'(z) - \tilde{B}(z)$  and  $\tilde{B}(z)$  are analytic and have nonnegative coefficients in their Taylor expansion at 0, we can use Lemma 5. This gives us that

$$\begin{aligned} \operatorname{Re} \left( b_0(z\tilde{B}'(z) - 2\tilde{B}(z)) + (z\tilde{B}'(z) - \tilde{B}(z))\tilde{B}(\bar{z}) \right) \\ \leq b_0(\tilde{B}'(1) - 2\tilde{B}(1)) + (\tilde{B}'(1) - \tilde{B}(1)) \cdot (\tilde{B}(1)) < b_0(b_0 - \tilde{B}(1)) + b_0\tilde{B}(1) = b_0^2. \end{aligned}$$

Here, we used that  $\tilde{B}'(1) < \tilde{B}(1) + b_0$ . Thus, we proved for each  $z \in \bar{D}_1$  inequality (39), which is equivalent to (38).  $\square$

We proved that  $\frac{d|z\tilde{z}|}{dt} > 0$  for each  $z = z(t) \in D_{1+\delta} \setminus \{0\}$ . Since  $b_0 + \tilde{B}(0) \neq 0$ , we get that  $z(t) \neq 0$ . Therefore, any root of Eq. (37) for  $t > 0$  goes outside  $\bar{D}_{1+\delta}$ . Hence, the number of roots inside  $D_1$  decreases when  $t$  increases. But, for small  $t > 0$ , we can use Rouché’s theorem to show that there is only one root of Eq. (37). Thus, we have proved Lemma 4.  $\square$

## Appendix B: Computational remarks

In this Appendix, we give several computational remarks.

### B.1 Zero-free case

From Assumption 2, it follows that the function  $B(z)$  has at most one zero in  $D_1$ . It is possible that the function  $B(z)$  has no zero in  $D_1$ , for example, if  $B(z)$  is the pgf of a Poisson random variable. We will show here how to simplify the computation in this case. Then, we will give a simplified version of Algorithm 2 and an easy test to check if  $B(z)$  is zero-free in  $D_1$  if  $B(z)$  is a pgf.

Consider  $\tilde{y}_k = z_k/B(z_k)$ . In the same manner as before, we can get formulas for  $\tilde{\eta}_k = \sum_{j=1}^{g-1} \tilde{y}_j^k$  that do not contain any integrals along  $S_\delta$ . More specifically, we



consider

$$\tilde{h}(y) = \sum_{k=0}^{g-1} x_k y^k.$$

Thus,  $\tilde{h}\left(\frac{z}{B(z)}\right) (B(z))^{g-1} = h\left(\frac{B(z)}{z}\right) z^{g-1}$ . We conclude that  $\tilde{y}_k$  are zeros of  $\tilde{h}(y)$ , and, thus,

$$\tilde{h}(y) = x_{g-1} \prod_{k=1}^{g-1} (y - \tilde{y}_k).$$

Using the same argument as in Sect. 2.1, we get

$$\frac{x_{g-1-k}}{x_{g-1}} = (-1)^k \sigma_k (\tilde{y}_1, \dots, \tilde{y}_{g-1}) = (-1)^k \tilde{\sigma}_k. \tag{40}$$

Note that, if  $B(z)$  does not have zeros in  $\bar{D}_1$ , then

$$\tilde{\eta}_k = \frac{1}{2\pi i} \oint_{S_{1+\varepsilon}} \frac{D'(z)}{D(z)} \left(\frac{z}{B(z)}\right)^k dz - 1. \tag{41}$$

Hence, we do not need to compute two integrals as in (7). As before, we need to subtract the residue at 1, which is equal to 1. This leads to the following simplified algorithm to find unknown variables  $x_k$ :

**Algorithm 3** (Computation of  $x_k, k = 0, \dots, g - 1$ , in the case of zero-free  $B(z)$ )

1. Find  $\varepsilon > 0$  such that there are only  $g$  roots of Eq. (2) in the disk  $\bar{D}_{1+\varepsilon}$ , using one of two ways described in Sect. B.3.
2. Compute  $\tilde{\eta}_k$ , using

$$\tilde{\eta}_k = -1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} z \frac{D'(z)}{D(z)} \left(\frac{z}{B(z)}\right)^k d\varphi,$$

where, for convenience,  $z = z(\varphi) = (1 + \varepsilon)e^{i\varphi}$ .

3. Compute  $\tilde{\sigma}_k$  by (8) using  $\tilde{\eta}_k$ .
4. Compute  $x_k, k = 0, \dots, g - 1$ , using (40) together with (5).

The necessary and sufficient condition for a pgf  $B(z)$  to be zero-free in  $D_1$  is given by the following lemma.

**Lemma 7** *The pgf  $B(z)$  is zero-free in  $D_1$  if and only if  $B(-1) > 0$ .*

**Proof** Suppose that  $B(-1) > 0$  and there is a zero  $t$  of  $B(z)$  in  $D_1$ . Since  $B(z)$  has at most one zero and  $B(\bar{t}) = \overline{B(t)}$ , this zero is on the real line. Consider some  $0 < \Delta < \min(B(-1), 1 - B'(1))$  and the function  $C(z) = \frac{B(z) - \Delta}{1 - \Delta}$ . Since  $B(0) \geq 1 - B'(1) > \Delta$ , the function  $C(z)$  is a pgf and  $C'(1) = \frac{B'(1)}{1 - \Delta} < 1$ . Therefore, according to Lemma 4,  $C(z)$  has not more than one zero in the unit disk. However,  $C(t) < 0 < C(-1), C(1)$ . Thus, by the intermediate value theorem, it has at least two zeroes. We get a contradiction.

The rest of the proof follows from the fact that if  $B(-1) \leq 0 < B(1)$ , then by the intermediate value theorem, there is such  $t \in [-1, 1]$  that  $B(t) = 0$ . □

### B.2 Choice of $\delta$

In (7), we use the parameter  $\delta$  such that there are no roots of Eq. (2) in  $\bar{D}_\delta = \{z: |z| \leq \delta\}$ . For a general function  $D(z)$ , we use the following method to find an appropriate  $\delta$ . Consider the integral

$$\frac{1}{2\pi i} \oint_{S_\delta} \frac{D'(z)}{D(z)} dz = \frac{1}{2\pi} \int_{-\pi}^\pi z \frac{D'(z)}{D(z)} d\varphi,$$

where  $z$  in the last integral stands for  $z(\varphi) = \delta e^{i\varphi}$ . By Lemma 1 and the residue theorem, the value of this integral is equal to the number of roots of Eq. (2) inside the disk  $\bar{D}_\delta$ . Therefore, if this integral is equal to zero, then  $\delta$  is correct. Hence, we suggest starting with  $\delta = 1/2$ . Then, check the value of the integral (2). If it is not zero (in this case it should be at least 1), then set  $\delta$  to  $1/4$  and check the integral again. Continuing like this, we will find an appropriate  $\delta$  after a finite number of steps since  $D(z)$  is an analytic function with  $D(0) \neq 0$ .

In the case  $D(z) = z^g - A(z)$ , we prove that any  $\delta < \delta_0 = A(0)/(2 - A(0))$  satisfies our condition. Consider  $A(z) = \sum_{j=0}^\infty a_j z^j$ . Since  $a_j \geq 0$  and  $\sum_{j=0}^\infty a_j = 1$ ,

$$|A(z) - A(0)| = \left| \sum_{j=1}^\infty a_j z^j \right| \leq \sum_{j=1}^\infty a_j |z|^j \leq (1 - A(0))|z|$$

for any  $z$  in the unit disk. Therefore, for  $z \in \bar{D}_\delta$ , where  $\delta < \delta_0$ , we get

$$|A(z)| \geq A(0) - |A(z) - A(0)| \geq A(0) - (1 - A(0))|z| > |z| \geq |z|^g.$$

Thus, there are no roots of Eq. (2) in  $\bar{D}_\delta$  for  $\delta < \delta_0$ .

Note that if  $g > 1$  and  $A(0) \neq 1$ ,  $A(0)/(2 - A(0)) < 1$ , and, thus, for each  $z \in \bar{D}_{\delta_0}$ , we have  $|z| > |z|^g$ . Therefore,  $|A(z)| > |z|^g$  even if  $A(0) - (1 - A(0))|z| = |z|$ , which is equivalent to  $|z| = \delta_0$ . Hence, in almost all cases,  $\delta = \delta_0$  is an appropriate value. Note that  $\delta_0 > 0$  since  $A(0) > 0$ . However,  $\delta_0$  can be so small that computational errors will arise. In this case, we can use the general way to find  $\delta$ .

### B.3 Choice of $\varepsilon$

The choice of  $\varepsilon$  may be done in the same way as the choice of  $\delta$  by computing contour integrals. For  $D(z) = z^g - A(z)$ , one can prove using Rouché’s theorem that there are no roots of Eq. (2) in  $\bar{D}_{1+\varepsilon} \setminus \bar{D}_1$  for  $1 + \varepsilon < z_{-1}$ , where  $z_{-1}$  is the (only) root of Eq. (2) on the open ray  $(1, \infty)$ . Thus, if a lower bound on  $z_{-1}$  is known or is easy to compute numerically, one can use any point on  $(0, z_{-1} - 1)$  as  $\varepsilon$ . However, the smaller  $\varepsilon$ , the greater the computational error for the integral, so we suggest taking

$\varepsilon = \operatorname{argmax}_{z \in [1, z_{-1}]} |z^\varepsilon - A(z)| - 1$  or, in the case when the errors are unlikely to arise, just  $\varepsilon = (z_{-1} - 1)/2$ . In this way, we stay far enough from the roots of Eq. (2).

#### B.4 From a complex integral to real integrals

In this subsection, we give a suggestion how to compute the required complex integrals. We need to compute a complex integral, so we need to compute the imaginary and real parts of the integral (15). This is equivalent to computing the following integrals:

$$\int_{-\pi}^{\pi} \operatorname{Re}(g(z(\varphi))z(\varphi))d\varphi, \quad \int_{-\pi}^{\pi} \operatorname{Im}(g(z(\varphi))z(\varphi))d\varphi.$$

However, in the applications, the performance indicators are real values. Note that in this case the imaginary part vanishes, and we need to compute only one integral.

#### References

- Adan, I.J.B.F., van Leeuwaarden, J.S.H., Winands, E.M.M.: On the application of Rouché's theorem in queueing theory. *Oper. Res. Lett.* **34**(3), 355–360 (2006)
- Al Hanbali, A., de Haan, R., Boucherie, R.J., van Ommeren, J.K.: Time-limited polling systems with batch arrivals and phase-type service times. *Ann. Oper. Res.* **198**(1), 57–82 (2012)
- Bailey, N.T.: On queueing processes with bulk service. *J. R. Stat. Soc. Ser. B (Methodological)* **16**, 80–87 (1954)
- Boon, M.A.A., Janssen, A.J.E.M., van Leeuwaarden, J.S.H., Timmerman, R.W.: Pollaczek contour integrals for the fixed-cycle traffic-light queue. *Queueing Syst.* **91**(1–2), 89–111 (2019)
- Boxma, O.J., Groenendijk, W.P.: Pseudo-conservation laws in cyclic-service systems. *J. Appl. Probab.* **24**(04), 949–964 (1987). <https://doi.org/10.1017/S002190020011681X>
- Darroch, J.N.: On the traffic-light queue. *Ann. Math. Stat.* **35**(1), 380–388 (1964). <https://doi.org/10.1214/aoms/1177703761>
- Frigui, I., Alfa, A.S.: Analysis of a time-limited polling system. *Comput. Commun.* **21**(6), 558–571 (1998)
- He, Q.M.: *Fundamentals of Matrix-Analytic Methods*, vol. 365. Springer, Berlin (2014)
- Janssen, A.J.E.M., van Leeuwaarden, J.S.H.: Analytic computation schemes for the discrete-time bulk service queue. *Queueing Syst.* **50**(2–3), 141–163 (2005)
- Janssen, A.J.E.M., van Leeuwaarden, J.S.H.: Back to the roots of the M/D/s queue and the works of Erlang. *Crommelin and Pollaczek. Stat. Neerl.* **62**(3), 299–313 (2008)
- McNeil, D.R.: A solution to the fixed-cycle traffic light problem for compound Poisson arrivals. *J. Appl. Probab.* **5**(3), 624–635 (1968). <https://doi.org/10.1017/S0021900200114457>
- Miller, A.J.: Settings for fixed-cycle traffic signals. *OR* **14**(4), 373–386 (1963). <https://doi.org/10.2307/3006800>
- Neuts, M.F.: Matrix-analytic methods in queueing theory. *Eur. J. Oper. Res.* **15**(1), 2–12 (1984)
- Oblakova, A., Al Hanbali, A., Boucherie, R.J., Ommeren, J.C.W., Zijm, W.H.M.: Exact expected delay and distribution for the fixed-cycle traffic-light model and similar systems in explicit form. Memorandum Faculty of Mathematical Sciences University of Twente (2056) (2016)
- Riska, A., Smirni, E.: Exact aggregate solutions for M/G/1-type Markov processes. In: Calzarossa, M.C., Tucci, S. (eds.) *ACM SIGMETRICS Performance Evaluation Review*, vol. 30, pp. 86–96. ACM, New York (2002)
- van den Broek, M.S., van Leeuwaarden, J.S.H., Adan, I.J.B.F., Boxma, O.J.: Bounds and approximations for the fixed-cycle traffic-light queue. *Transp. Sci.* **40**(4), 484–496 (2006). <https://doi.org/10.1287/trsc.1050.0146>
- van Leeuwaarden, J.S.H.: Delay analysis for the fixed-cycle traffic-light queue. *Transp. Sci.* **40**(2), 189–199 (2006). <https://doi.org/10.1287/trsc.1050.0125>

18. van Ommeren, J.C.W.: The discrete-time single-server queue. *Queueing Syst.* **8**(1), 279–294 (1991)
19. Vinberg, E.B.: *A Course in Algebra*, vol. 56. American Mathematical Society, Providence (2003). <https://doi.org/10.1090/gsm/056>
20. Webster, F.V.: Traffic signal settings. Technical Report (1958)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.