

A Taylor series expansions approach to queues with train arrivals

Koen De Turck Dieter Fiems Sabine Wittevrongel Herwig Bruneel
SMACS research group, Department TELIN, Ghent University
St-Pietersnieuwstraat 41, 9000 Gent, Belgium
{kdeturck,df,sw,hb}@telin.UGent.be

ABSTRACT

We assess the performance of a discrete-time queueing system with train arrivals. Arrivals at the queue stem from a number of active sessions, each generating a packet in a slot with fixed probability q . Since an exact analysis is not feasible for $q \neq 1$, we rely on Taylor-series expansions around $q = 0$ of the joint probability generating functions of the number of active sessions and the queue content. These expansions are then either combined with the known generating function for $q = 1$ if the system is stable for $q = 1$, or with heavy-traffic results if this is not the case. In both cases, we obtain expressions for the moments of queue content and packet delay and assess the accuracy of our approximations by means of simulation.

1. INTRODUCTION

Input traffic at various nodes in packet switched telecommunication networks typically exhibits various levels of correlation. It is well known that input correlation significantly affects queueing performance and hence there is a continuing interest in analytically tractable queueing models which can accurately capture arrival correlation. There is a particular interest in Markovian arrival models, including models with a finite state space such as the discrete-time batch-Markovian arrival model [1, 2], or with a structured infinite state space such as the discrete autoregressive arrival models [3, 4], branching arrival models [5] and train arrival models [6, 7, 8, 16].

Train arrival models, which are also referred to as session-based arrivals or queueing models with $M/G/\infty$ -input, are a particular type of two-level models: arriving packets are organized in larger entities denoted by sessions or trains. The exhibited time correlation then follows from the intricate interplay between the arrival process of new sessions, the duration of these sessions, and the packet arrival process within a single session. While train-arrival models come natural — the concept of sessions is omnipresent at higher layers of the network protocol stack — applicability of train-arrival

models in literature is somewhat limited by the implicit assumption that the load associated to a single session should equal or exceed the service speed. Albeit implicit, this assumption is key to the tractability of the queueing analysis: it implies that the arrival process is in a fixed (Markovian) state when the queue is empty which facilitates the queueing analysis. Lifting this assumption poses some mathematical difficulties and is the subject of this paper.

To the best of our knowledge, the queueing problem at hand cannot be solved by exact analytical techniques. Therefore, heavy-traffic techniques have been applied to models with a related arrival model [9, 10]. Heavy-traffic results offer a good deal of insight in the performance model and reasonable numerical agreement of systems operating near their maximal capacity. In this paper, we focus on light-traffic approximations. In particular, we develop a Taylor-series expansion in the probability with which an active session produces packets. Some previous work on Taylor-series expansions for queueing models has been reported. Overviews are given in [15, 17]. A variety of techniques have been applied to this problem as well as to related topics such as perturbation theory and light-traffic approximations. Sample-path based techniques [18], Palm theory [19], ladder-height distributions [20] have been reported in the literature. In this paper, we take the functional equation of the joint generating function as a starting point, and we find expansions of the generating function (thus, of the entire distribution), instead of just in the expected value.

The remainder of this paper is organised as follows. In the next section, the queueing model under study is described in detail and the necessary notation is introduced. The analysis of this model is then presented in section 3 while our approximations are numerically evaluated in section 4. Finally, we justify our Taylor series expansions in section 5, and draw conclusions in section 6.

2. QUEUEING MODEL

We consider a discrete-time queueing system. Time is divided into fixed-length intervals, called slots. During the consecutive slots, packets arrive at the queueing system, are stored in an infinite capacity buffer and are transmitted in order of arrival. The transmission times of the consecutive packets are fixed and equal to the slot length. Moreover, transmission of packets is synchronised with respect to slot boundaries which implies that packets cannot be transmitted during their arrival slot. We investigate a single-server

system: if packets are present in the buffer at the beginning of the slot, only a single packet leaves the system at the end of the slot.

The packets that arrive in the queueing system are generated by a number of active sessions; while being active, a session generates a single packet in a slot with a fixed probability q or no packet with probability $1 - q$, independently from slot to slot. The number of new sessions that start in slot k is denoted by N_k . The consecutive N_k constitute a sequence of independent and identically distributed random variables with common probability generating function $N(z)$. For further use, let $\lambda_i = \mathbb{E}[N_0^i]$ denote the i th moment of the number of new sessions in a slot, also let $\lambda = \lambda_1$. Sessions typically span multiple slots. The length (in slots) of a session is an independent geometrically distributed random variable. Let p denote the probability that an active session continues during the next slot. The probability mass function of the session length then equals $s(n) = p^{n-1}(1-p)$ for $n = 1, 2, \dots$

We now relate the number of active sessions at consecutive slots, as well as the number of active sessions in a slot and the number of arrivals that they generate. Let $B_k^{(q)}(t)$ ($\hat{B}_k^{(p)}(t)$) be a sequence of counting functions of independent and identically distributed Bernoulli processes, q (p) being the success probability. For a fixed t , $B_k^{(q)}(t)$ is a binomially distributed random variable with parameters t (the number of trials) and q (the success probability). The number of active sessions in slot k is denoted by X_k . In view of the description of the arrival model above, the number of active sessions in slot $k+1$ can be expressed in terms of the number of active sessions in slot k as follows,

$$X_{k+1} = B_k^{(p)}(X_k) + N_{k+1}. \quad (1)$$

Finally, the number of packet arrivals A_k in a slot can be expressed in terms of the number of active sessions during slot k as follows,

$$A_k = \hat{B}_k^{(q)}(X_k). \quad (2)$$

These equations describe the arrival process and complement the evolution of the queue size, described in the following section.

3. QUEUEING ANALYSIS

Having settled the notation and the details of the model at hand, we now proceed to the Taylor-series approximations. In section 3.1, we first establish the system equations of the queueing model and the functional equation for the probability generating function of the (Markovian) state of this system in equilibrium. By means of a Taylor-series expansion of this functional equation in the probability q around $q = 0$, light-traffic approximations are extracted for the probability generating function, mean and variance of the queue content (and packet delay) in Section 3.2. For $q = 1$, the queueing system is either stable or not. If the system is not stable for $q = 1$, there exists a value $0 < q_m < 1$ such that the queueing system is stable for $q < q_m$. Therefore, the heavy-traffic approximation for $q \rightarrow q_m$ is considered in Section 3.3. If the system is stable for $q = 1$, we recall the known analytical results for this case in section 3.4. In either case, we obtain an approximant for the different moments of queue

content and packet delay by combining these results with our light-traffic approximation.

3.1 Functional equation

Let U_k denote the queue content at the beginning of slot k . If there are packets in the queue at the beginning of slot k , a single packet departs at the end of this slot, whereas no packet departs if the queue is empty at the beginning of slot k . Moreover, all packet arrivals during slot k are in the queue at the beginning of slot $k+1$. Hence, we have the following system equation,

$$U_{k+1} = (U_k - 1)^+ + A_k, \quad (3)$$

where the notation $(\cdot)^+$ is the usual shorthand notation for $\max(0, \cdot)$.

It is easy to verify that the queue content U_k and the number of active sessions $X_k - 1$ completely describe the state (in the Markovian sense) of the queueing system at the beginning of slot k . Therefore, let

$$P_k(x, z) = \mathbb{E} \left[x^{X_{k-1}} z^{U_k} \right] \quad (4)$$

denote the joint probability generating function of the number of active sessions during slot $k-1$ and the queue content at the beginning of slot k . We now express the corresponding generating function at the beginning of slot $k+1$ in terms of this generating function,

$$\begin{aligned} P_{k+1}(x, z) &= \mathbb{E}[x^{X_k} z^{U_{k+1}}] = \mathbb{E}[x^{X_k} z^{(U_k - 1)^+ + A_k}] \\ &= \frac{1}{z} \left[P_k(1 - p + px(1 - q + qz), z) \right. \\ &\quad \left. + (z - 1)P_k(1 - p + px(1 - q + qz), 0) \right] \\ &\quad \times N(x(1 - q + qz)). \end{aligned} \quad (5)$$

Here, the last equation follows from the observation that A_k is binomially distributed with success probability q , and with the number trials equal to X_k , where X_k is the sum of N_k and a binomially distributed term with success probability p and number of trials equal to X_{k-1} .

As the queue under study belongs essentially to the broad class of $G/G/1$ -systems with a stationary input process, it is straightforward to show (e.g. by means of a Loynes approach) that the queueing system under investigation reaches equilibrium if,

$$\rho = \frac{\lambda q}{1 - p} < 1. \quad (6)$$

Under this assumption, there exists a stationary ergodic process (U_k^*, X_k^*) to which the process (U_k, X_{k-1}) converges almost surely, starting from any initial state (U_1, X_0) . Let $P(x, z)$ denote the joint probability generating function of the number of active sessions X_0^* and the queue content U_1^* in equilibrium. In view of (5), this probability generating function then satisfies the following functional equation,

$$\begin{aligned} P(x, z) &= \frac{1}{z} \left[P(\bar{p} + px(\bar{q} + qz), z) \right. \\ &\quad \left. + (z - 1)P(\bar{p} + px(\bar{q} + qz), 0) \right] N(x(\bar{q} + qz)). \end{aligned} \quad (7)$$

Here we introduced $\bar{p} = 1 - p$ and $\bar{q} = 1 - q$ to simplify notation. Despite its apparent simplicity, this functional equation cannot be solved in closed form for general q to the best of our knowledge. Therefore, we focus on Taylor-series expansions in the next section.

3.2 Taylor series around $q = 0$

In the remainder, it is assumed that the generating function $P(x, z)$ is analytic in x, z and q in a neighbourhood of $x = z = q = 0$. It is shown in section 5 that this assumption is indeed justified. For ease of notation, let $M_{i,j}(x)$ denote the coefficient in the term $[q^i z^j]$ of the expansion of $P(x, z)$ in q and z . In other words, by the assumed analyticity, the probability generating function $P(x, z)$ can be expressed as follows,

$$P(x, z) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} M_{i,j}(x) z^j q^i. \quad (8)$$

Proving that $M_{i,j}(x) = 0$ for $j > i$. We now show by induction that $M_{i,j}(x) = 0$ for $j > i$. That is, the i th term in the expansion of $P(x, z)$ in q is a polynomial of order i in z . By plugging in the expansion (8) in the functional equation (7), and by comparing terms in $[q^0 z^k]$ on either side of the equation, we find,

$$\begin{aligned} M_{0,0}(x) &= N(x)(M_{0,1}(\bar{p} + px) + M_{0,0}(\bar{p} + px)), \\ M_{0,k-1}(x) &= N(x)M_{0,k}(\bar{p} + px), \quad \text{for } k > 1. \end{aligned}$$

By construction, $M_{0,1}(x)$ is a partial probability generating function: all coefficients in its series expansion are non-negative. Moreover, from the first equality above, we find $M_{0,1}(1) = 0$. This immediately implies $M_{0,1}(x) = 0$ for $|x| < 1$. The second equality above then shows that $M_{0,k}(x) = 0$ as well for all $k > 1$. This is the base case of our proof by induction (i.e. for $i = 0$).

For the inductive step, assume that $M_{i,j}(x) = 0$ for $j > i$ and $i < k$, i.e. that the terms in q^0 to q^{k-1} are polynomials in z of order 0 to $k-1$. We show that then also $M_{k,j}(x) = 0$. For ease of notation, let $[X]_{[q^k z^l]}$ denote the coefficient of $q^k z^l$ in the Taylor series expansion of X . From the functional equation we find for $m \geq 1$,

$$\begin{aligned} [P(x, z)]_{[q^k z^{k+m}]} &= \\ & \left[P(\bar{p} + px(\bar{q} + qz), z) N(x(\bar{q} + qz)) \right]_{[q^k z^{k+m+1}]} \\ & + \left[P(\bar{p} + px(\bar{q} + qz), 0) N(x(\bar{q} + qz)) \right]_{[q^k z^{k+m}]} \\ & - \left[P(\bar{p} + px(\bar{q} + qz), 0) N(x(\bar{q} + qz)) \right]_{[q^k z^{k+m+1}]} \cdot \quad (9) \end{aligned}$$

The two last terms of the former equation are zero. This follows from the observation that the coefficients of z^n in $P(\bar{p} + px(\bar{q} + qz), 0) N(x(\bar{q} + qz))$ are at least of the order n in q (for any $n \geq 0$). By this observation and by plugging

in the Taylor series expansion (8), we get,

$$\begin{aligned} M_{k,k+\ell}(x) &= \\ & = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[M_{i,j}(\bar{p} + px(\bar{q} + qz)) N(x(\bar{q} + qz)) \right]_{[q^{k-i} z^{k+\ell+1-j}]} \\ & = \sum_{i=0}^{k-1} \sum_{j=0}^i \left[M_{i,j}(\bar{p} + px(\bar{q} + qz)) N(x(\bar{q} + qz)) \right]_{[q^{k-i} z^{k+\ell+1-j}]} \\ & \quad + \sum_{j=0}^{k+\ell+1} \left[M_{k,j}(\bar{p} + px(\bar{q} + qz)) N(x(\bar{q} + qz)) \right]_{[q^0 z^{k+\ell+1-j}]} \\ & = \sum_{j=0}^{k+\ell+1} \left[M_{k,j}(\bar{p} + px(\bar{q} + qz)) N(x(\bar{q} + qz)) \right]_{[q^0 z^{k+\ell+1-j}]} \cdot \quad (10) \end{aligned}$$

The second equality uses the fact that the $M_{i,j}(x) = 0$ for $i < k$ and $j > i$, while the first summation again disappears in the third equation by the observation that the coefficients of z^n are at least of the order n in q . Finally, we find,

$$M_{k,k+\ell}(x) = M_{k,k+\ell+1}(\bar{p} + px) N(x), \quad \text{for } \ell \geq 1. \quad (11)$$

By construction, $\sum_{\ell} M_{k,k+\ell}(x)$ converges for $x = 1$. This is only possible if $M_{k,k+\ell}(1) = 0$ for $\ell \geq 1$ since equation (11) implies $M_{k,k+\ell}(1) = M_{k,k+\ell+1}(1)$ for $\ell \geq 1$. Repeated application of equation (11) further shows,

$$M_{k,k+1}(x) = M_{k,k+\ell}(f_{\ell}(x)) \prod_{m=1}^{\ell-1} N(f_m(x)) \quad (12)$$

with $f_1(x) = x$ and $f_n(x) = 1 - p + pf_{n-1}(x)$, $n > 1$. Taking the limit for $n \rightarrow \infty$, the product on the right-hand side of (12) converges for all $|x| < 1$. Moreover, $\lim_{\ell \rightarrow \infty} M_{k,k+\ell}(f_{\ell}(x)) = \lim_{\ell \rightarrow \infty} M_{k,k+\ell}(1) = 0$ which in turn implies $M_{k,k+1}(x) = 0$ for all $|x| < 1$. By equation (11), we finally find that $M_{k,k+\ell}(x) = 0$ for $|x| < 1$ and $\ell \geq 1$. Hence, given that $M_{i,j}(x) = 0$ for $i < k$ and $j > i$, we have shown that $M_{k,j}(x) = 0$ for $j > k$. We conclude that the probability generating function $P(x, z)$ can be expressed as follows,

$$P(x, z) = \sum_{i=0}^{\infty} \sum_{j=0}^i M_{i,j}(x) z^j q^i. \quad (13)$$

Functional equations for $M_{i,j}(x)$. In view of this expression, the functional equation (7) can be rewritten as follows,

$$\begin{aligned} \sum_{i=0}^{\infty} \sum_{j=0}^i M_{i,j}(x) z^j q^i &= \frac{1}{z} \left[\sum_{i=0}^{\infty} \sum_{j=0}^i M_{i,j}(\bar{p} + px(\bar{q} + qz)) z^j q^i \right. \\ & \left. + (z-1) \sum_{i=0}^{\infty} M_{i,0}(\bar{p} + px(\bar{q} + qz)) q^i \right] N(x(\bar{q} + qz)). \quad (14) \end{aligned}$$

Collecting the terms in $q^i z^j$ ($i = 0, 1, \dots$ and $j = 0, \dots, i$) on both sides of the former equation, we then obtain a system of functional equations. This system can be used to find the series expansion in q of the moments of the queue content as will be shown in the following subsection. For ease of

notation, let $\widehat{M}_{i,j}(x) = M_{i,j}(1-p+px)$. The systems of functional equations corresponding to $i=0$, $i=1$ and $i=2$ are displayed below. For $i=0$, we have,

$$M_{0,0}(x) = \widehat{M}_{0,0}(x)N(x).$$

For $i=1$, we have,

$$\begin{cases} M_{1,0}(x) &= \widehat{M}_{1,1}(x)N(x) + \widehat{M}_{1,0}(x)N(x) \\ &\quad - \widehat{M}_{0,0}(x)xN'(x) - pxN(x)\widehat{M}'_{0,0}(x) \\ M_{1,1}(x) &= \widehat{M}_{0,0}(x)xN'(x) + px\widehat{M}'_{0,0}(x)N(x) \end{cases}$$

For $i=2$, we have,

$$\begin{cases} M_{2,0}(x) &= \widehat{M}_{2,1}(x)N(x) + \widehat{M}_{2,0}(x)N(x) \\ &\quad - \widehat{M}_{1,1}(x)xN'(x) - px\widehat{M}'_{1,1}(x)N(x) \\ &\quad - \widehat{M}_{1,0}(x)xN'(x) - px\widehat{M}'_{1,0}(x)N(x) \\ &\quad + 2(px)^2\widehat{M}''_{0,0}(x)N(x) + px^2\widehat{M}'_{0,0}(x)N'(x) \\ &\quad + 2x^2\widehat{M}_{0,0}(x)N''(x) \\ M_{2,1}(x) &= \widehat{M}_{2,2}(x)zN(x) + \widehat{M}_{1,1}(x)xN'(x) \\ &\quad + px\widehat{M}'_{1,1}(x)N(x) + \widehat{M}_{1,0}(x)xN'(x) \\ &\quad + px\widehat{M}'_{1,0}(x)N(x) - 4(px)^2\widehat{M}''_{0,0}(x)N(x) \\ &\quad - 2px^2\widehat{M}'_{0,0}(x)N'(x) - 4x^2\widehat{M}_{0,0}(x)N''(x) \\ M_{2,2}(x) &= 2(px)^2\widehat{M}''_{0,0}(x)N(x) + px^2\widehat{M}'_{0,0}(x)N'(x) \\ &\quad + 2x^2\widehat{M}_{0,0}(x)N''(x) \end{cases}$$

As already illustrated by the systems above, the systems of equations grow quickly in size and manual calculation becomes tedious. Higher-order expressions are omitted but can be easily calculated with Mathematica or Maple. The numerical evaluation of the series expansion in section 4 includes higher-order terms. For now, we only mention that these functional equations have the following form,

$$\mathbf{M}_i(x) = \mathbf{M}_i(1-p+px)\mathcal{A}_i(x) + \mathbf{N}_i(x), \quad (15)$$

with,

$$\mathbf{N}_i(x) = \sum_{k=0}^{i-1} \sum_{l=0}^i \mathbf{M}_k^{(l)}(1-p+px)\mathcal{B}_{k,l}(x). \quad (16)$$

Here \mathbf{M}_i is a row vector with entries $M_{i,j}$ ($j=0, \dots, i$) and \mathcal{A}_i and $\mathcal{B}_{j,k}$ are known matrices. Furthermore, $\mathbf{M}_k^{(l)}$ is the l th derivative of \mathbf{M}_k . Notice that $\mathbf{N}_i(x)$ is a vector which can be expressed in terms of known functions and in terms of \mathbf{M}_k , $k < i$. Hence, \mathbf{N}_i is known if one solves the equations (15) for consecutive i .

The functional equation (15) can be solved by repeated application of this equation. We have,

$$\mathbf{M}_i(x) = \mathbf{M}_i(1) \prod_{\ell=1}^{\infty} \mathcal{A}_i(f_{\ell}(x)) + \sum_{j=1}^{\infty} \mathbf{N}_i(f_j(x)) \prod_{\ell=1}^{j-1} \mathcal{A}_i(f_{\ell}(x)). \quad (17)$$

The unknown vectors $\mathbf{M}_i(1)$ can be determined as follows. Plugging $x=1$ into the matrix representation (15) of the functional equations yields,

$$\mathbf{M}_i(1) = \mathbf{M}_i(1)\mathcal{A}_i(1) + \mathbf{N}_i(1). \quad (18)$$

For all i , the matrix $\mathcal{A}_i(1)$ is singular such that the former expression does not allow to solve for $\mathbf{M}_i(1)$. However the

normalisation condition yields an extra equation for each i . In view of equation (13), the normalisation condition $P(1,1) = 1$ yields

$$P(1,1) = \sum_{i=0}^{\infty} \sum_{j=0}^i M_{i,j}(1)q^i, \quad (19)$$

such that

$$M_{0,0}(1) = 1, \quad \sum_{j=0}^i M_{i,j}(1) = 0, \quad (20)$$

for $i \geq 1$.

Moments. We now focus on the Taylor series expansion in q of the n th moment of the queue content. By the representation (13), this expansion can be expressed as follows,

$$\mathbb{E}[U_{\text{LT}}^n] = \sum_{i=1}^{\infty} \theta_{i,n} q^i, \quad \theta_{i,n} = \sum_{j=0}^i j^n M_{i,j}(1). \quad (21)$$

In other words, for the calculation of the n th order expansion, the unknown values $M_{i,j}(1)$ need to be determined for $i=0, \dots, n$ and $j=0, \dots, i$. For the calculation of the k th coefficient, equation (16) shows that we need to determine the values of the derivatives up to order k of $\mathbf{M}_i(x)$ for $x=1$, $i < k$. These can be found by differentiating either equation (17) or — more conveniently — equation (15) and by evaluating in $x=1$. In particular, the 3rd order Taylor series expansion of the first two moments of the queue content are displayed below,

$$\begin{aligned} \mathbb{E}[U_{\text{LT}}] &= \frac{\lambda_1}{1-p}q + \left(\frac{\lambda_2}{2(1-p)^2} + \frac{2p\lambda_1^2 - (1-p)\lambda_1}{2(1+p)(1-p)^2} \right) q^2 \\ &\quad + \left(\frac{p}{2(1-p)(p^2+p+1)} \lambda_3 \right. \\ &\quad + \frac{(5p^3+3p^2+1)\lambda_1 - p(1+2p)(1-p)^2 \lambda_2}{2(1-p)^2(p^2+p+1)(1+p)} \lambda_2 \\ &\quad + \frac{4p^4 - p^3 + p^2 - 1}{2(1+p)(p^2+p+1)(1-p)^2} \lambda_1^2 \\ &\quad \left. + \frac{2p(1+2p^3)\lambda_1^2 - 2p^3(1-p)^2 \lambda_1}{2(1+p)(p^2+p+1)(1-p)^3} \lambda_1 \right) q^3 + O(q^4), \quad (22) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[U_{\text{LT}}^2] &= \frac{\lambda_1}{1-p}q + \left(\frac{3\lambda_2}{2(1-p)^2} + \frac{6p\lambda_1^2 - 3(1-p)\lambda_1}{2(1+p)(1-p)^2} \right) q^2 \\ &\quad + \left(\frac{(15p^3+13p^2+2p+3)\lambda_1 - 6p^4 + 6p \lambda_2}{2(1-p)^2(1+p)(1+p+p^2)} \lambda_2 \right. \\ &\quad - \frac{(9p+2)}{3(1-p)(p^2+p+1)} \lambda_3 \\ &\quad + \frac{2p(6p^3+2p^2+3)\lambda_1 - 12p^5 + 15p^4 - 2p^3 + p^2 + p - 3}{2(1+p)(p^2+p+1)(1-p)^3} \lambda_1^2 \\ &\quad \left. - \frac{9p^3 - 2p - 2}{3(p^2+p+1)(1-p^2)} \lambda_1 \right) q^3 + O(q^4). \quad (23) \end{aligned}$$

The accuracy of these expressions is evaluated in section 4.

Packet delay. We consider a discrete-time queueing system with single-slot packet transmission times. Hence, the prob-

ability generating functions of queue content $P(1, z)$ and delay $D(z)$ relate as follows [11]:

$$D(z) = \frac{P(1, z) - P(1, 0)}{1 - P(1, 0)}, \quad (24)$$

where $1 - P(1, 0)$ equals the load ρ of the queueing system:

$$1 - P(1, 0) = \frac{\lambda q}{1 - p}.$$

Therefore, the Taylor-series expansion in q of the n moment of the delay is expressed as follows,

$$E[D_{LT}^n] = \sum_{i=0}^{\infty} \hat{\theta}_{i,n} q^i. \quad (25)$$

where the coefficients $\hat{\theta}_{i,n}$ are

$$\hat{\theta}_{i,n} = \theta_{i+1,n} \frac{1-p}{\lambda} \quad (26)$$

3.3 Heavy-traffic analysis

In this section, we assume that $\lambda \geq 1 - p$. In this case, the queueing system is not stable for $q = 1$ and the moments of the queue content diverge to ∞ for $q \rightarrow q_m = \frac{1-p}{\lambda}$. Hence, in this case, we aim for heavy-traffic limit results, as pioneered by Kingman in the seminal paper [21].

The performance measures are given in terms of the load and the asymptotic variance. The load $\rho = \frac{\lambda q}{1-p}$ is given as before; the asymptotic variance is defined as follows. Consider the random variable S_κ denoting the total number of arrivals during a period of κ slots, say slot 1 to slot κ . The asymptotic variance is then given by,

$$V = \lim_{\kappa \rightarrow \infty} \frac{1}{\kappa} \text{Var}[S_\kappa]. \quad (27)$$

Let $Y_k^{(\kappa)}$ denote the number of arrivals in slots 1 to κ that belong to a session that started in slot k and recall that N_k denotes the number of new sessions started in slot k . Moreover, let $B_{m,n}$ denote a doubly indexed sequence of Bernoulli distributed random variables with success probability q_m and let G_m be the duration of the m th session that starts in slot k . Then $Y_k^{(\kappa)}$ distributes as,

$$Y_k^{(\kappa)} \stackrel{d}{=} \sum_{m=1}^{N_k} \sum_{n=1}^{\widehat{G}_m^{(\kappa-k+1)}} B_{m,n}, \quad \widehat{G}_m^{(\ell)} = \max(G_m, \ell), \quad (28)$$

and hence,

$$\text{Var}[Y_k^{(\kappa)}] = \lambda \text{Var}[\widehat{G}_m] q_m^2 + \lambda E[\widehat{G}_m] q_m (1 - q_m) + (\lambda_2 - \lambda^2) E[\widehat{G}_m]^2 q_m^2. \quad (29)$$

Taking the limit for $\kappa \rightarrow \infty$, we have,

$$V = \lim_{\kappa \rightarrow \infty} \text{Var}[Y_k^{(\kappa)}] = \lambda \frac{p}{(1-p)^2} q_m^2 + \lambda \frac{1}{1-p} q_m (1 - q_m) + (\lambda_2 - \lambda^2) \frac{1}{(1-p)^2} q_m^2, \quad (30)$$

which further simplifies to,

$$V = \frac{2p-1}{\lambda} + \frac{\lambda_2}{\lambda^2}. \quad (31)$$

The first two moments of the buffer content under heavy traffic can then be found to be equal to (using eg. [21]):

$$E[U_{HT}] = \frac{V}{2(1-\rho)} = (1-p) \frac{(2p-1)\lambda + \lambda_2}{2\lambda^2(1-p-q\lambda)},$$

$$E[U_{HT}^2] = \frac{V^2}{2(1-\rho)^2} = 2 \left((1-p) \frac{(2p-1)\lambda + \lambda_2}{2\lambda^2(1-p-q\lambda)} \right)^2. \quad (32)$$

In order to combine the Taylor expansion in the neighbourhood of $q = 0$ (light traffic) and the heavy-traffic results we use an idea proposed by [22]: as light-traffic approximations of the n th order provide the exact expressions for the first k derivatives around $\rho = 0$, and heavy-traffic approximations provide the exact asymptote around $\rho = 1$, an expression of the form

$$E[U_{LT+HT}^n] = \theta_{1,n} q + \theta_{2,n} q^2 + \dots + \theta_{k,n} q^k + \left(\frac{q\lambda}{1-p} \right)^{k+1} E[U_{HT}^n] \quad (33)$$

retains all the information contained in the two approximations. Note that due to the aforementioned result of [11], we see that in heavy traffic and for one-slot service times, the distributions of buffer content and packet delay coincide. Hence a similar combination of light-traffic and heavy-traffic results can be obtained for the moments of the packet delay.

3.4 The case $q = 1$

As already mentioned, assuming $\lambda/(1-p) < 1$, one can explicitly solve the functional equation for $q = 1$. Equation (7) then simplifies to,

$$P(x, z) = \frac{N(xz)}{z} \left(P(1-p+pxz, z) + (z-1)P(1-p+pxz, 0) \right). \quad (34)$$

For $q = 1$, active sessions imply arrivals such that there are no active sessions if the queue is empty. This means that the unknown function $P(x, 0)$ is a constant: $P(x, 0) = P(0, 0) = 1 - \lambda/(1-p)$. By this observation, all moments of the queue content can be found; see [6] for details. For example, the mean queue content equals,

$$E[U_1] = \frac{\lambda p(1-p) + \lambda^2(1-5p) - \lambda_2(1-p)}{(1-p)^2(2\lambda-1+p)}. \quad (35)$$

Expressions for higher-order moments are rather spacious and are therefore omitted. As for the heavy-traffic results presented above, results for $q = 1$ can be combined with light-traffic results into an approximant for all $q \in [0, 1]$. In particular, Padé approximants [23] that additionally satisfy the values of the performance measures at hand for $q = 1$ can be used to this end. A number of numerical examples are displayed in the following section.

4. NUMERICAL EVALUATION

In this section, we provide a number of numerical examples and verify our approximants by means of simulation. In order to limit the number of parameters, we make the additional assumption that the number of new sessions that start in a slot is a Poisson distributed random variable with mean λ . Hence, the arrival process is characterised by the triple (λ, p, q) .

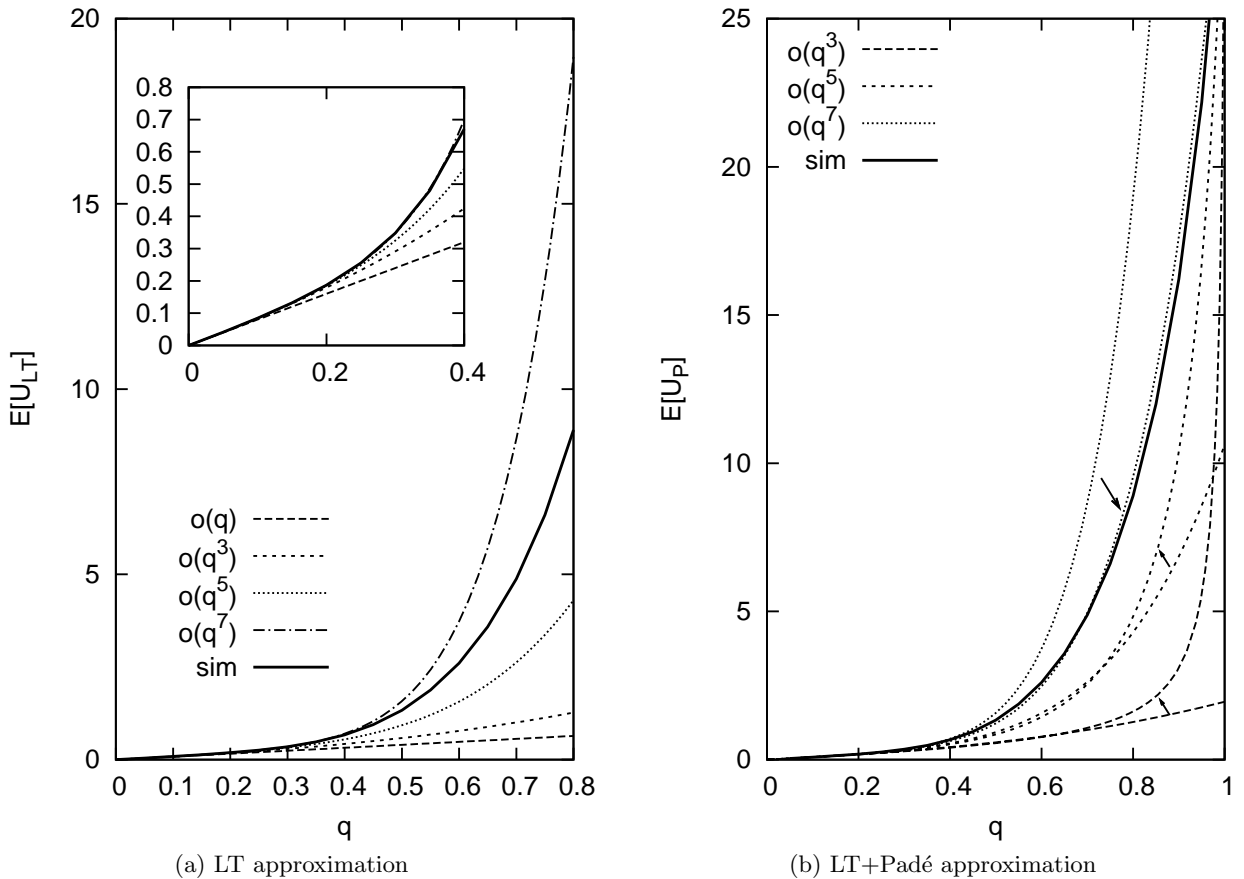


Figure 1: Taylor-series expansion for the mean queue content up to different orders as depicted versus q .

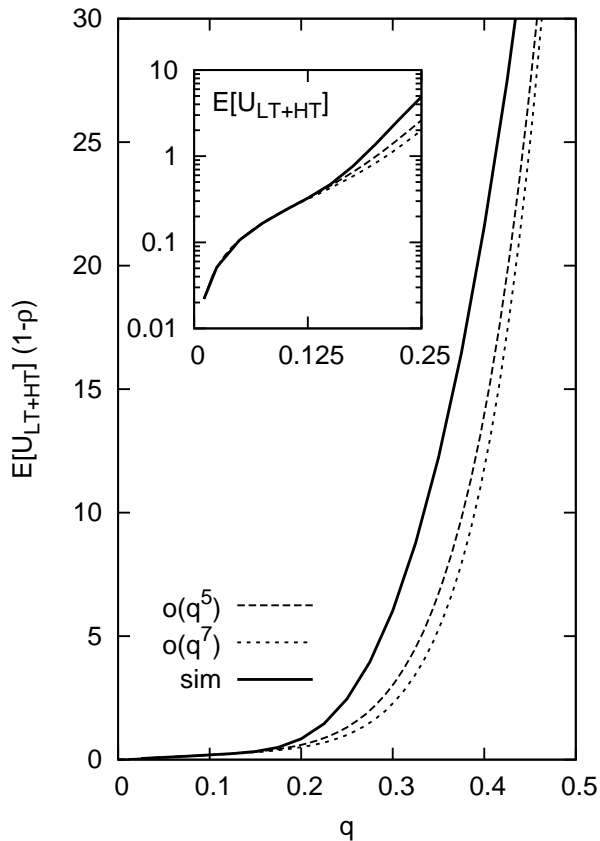
Figure 1 depicts the mean queue content versus q . We assume $p = 0.9$ and $\lambda = 0.08$ such that the mean session duration is 10 slots and such that the load is 80% if sessions always produce packets ($q = 1$). Figure 1(a) depicts the light-traffic approximation curves as obtained in section 3.2 as well as simulation results for this performance measure. To illustrate the accuracy for small q , we enlarged the curves for $0 \leq q \leq 0.4$ in the subplot. Higher-order approximants are more accurate for small q . However, for larger q , this is not necessarily the case. Since light-traffic results are inaccurate for larger q , we combine them with the exact expression for $q = 1$ by means of a Padé approximant (see section 3.4). The best results were obtained by Padé approximants with a first order denominator. The results are depicted in figure 1(b). For all orders of LT approximants, we depict both the original approximant and the Padé approximant (an arrow points from the original approximant to the Padé approximant). This approach improves the accuracy of the approximants for all orders of the expansions. In particular, the 7th order expansion yields a very accurate approximant.

While the queue was stable for $q = 1$ in the preceding example, this is not the case in Figure 2. Here, we assume $p = 0.99$ and $\lambda = 0.02$ such that the mean session duration is 100 slots and such that the load is 100% if sessions produce packets with probability $q_m = 0.5$. Figures 2(a) and 2(b) de-

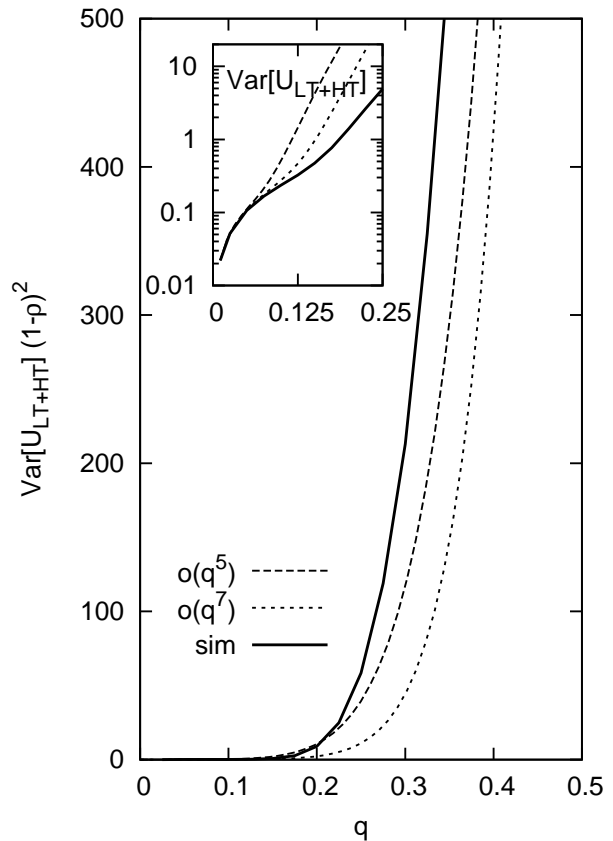
pic the mean and variance of the queue content versus the probability q . Curves of the various approximants — we here consider the combination of light-traffic and heavy-traffic results, see section 3.3 — are depicted as well as simulation results. To facilitate the display of the heavy-traffic results, we scale mean and variance with $(1 - \rho)$ and $(1 - \rho)^2$, respectively. Moreover, we again zoom in (without scaling) on the values for small q to assess the accuracy of the approximation in this region. From Figure 2(a), we observe that our approximants for the mean queue content are accurate for loads up to $\rho = 40\%$ ($q = 0.2$). For higher-order moments, the approximants are less accurate though, see Figure 2(b).

5. AN ANALYTICITY PROOF

In this section we show that it is justified to perform the Taylor-series expansions. The analyticity of a stationary solution is not automatic and is in general not as straightforward to establish as one may be inclined to think [17]. The basic problem is that the analyticity of finite-horizon queueing characteristics (which is easier to establish) does not automatically extend to stationary characteristics. To establish analyticity of the latter, there are broadly speaking two approaches. The first, developed by Malyshev, Menshikov and Fayolle [12], is an extension to the Foster-Lyapunov approach to the ergodicity of Markov chains, and makes use of some Banach spaces. It is restricted to Markovian systems, see also [13]. The other approach was developed by



(a) mean queue content



(b) variance of the queue content

Figure 2: LT+HT approximant for the mean and variance of the queue content up to different orders as depicted versus q .

Błaszczyszyn in the more general framework of stationary point processes, and makes use of Palm theory. In this paper, we have opted for the former approach.

For the purpose of this section, we view the stochastic model of this paper as a family of Markov chains $\{L_q\}$ (parametrised by the parameter q). The goal of this section is to establish the following theorem:

THEOREM 1. *The family $\{L_q\}$ of Markov chains admits a unique stationary distribution $\pi_q(\cdot)$ in a neighbourhood of $q = 0$. Moreover $\pi_q(\cdot)$ depends analytically on q .*

In the proof of this theorem, we make use of a general theorem reported by Fayolle et al. in [12], more precisely Theorem 7.3.2. It is based on the uniform convergence of operators associated with the family of Markov chains $\{L_q\}$ (parametrised by the parameter q) on a countable two-dimensional state space $A = \mathbb{N} \times \mathbb{N}$. We denote a generic element of A as (s, u) , where s denotes the number of active sessions and u denotes the buffer content.

For convenience, we include the formulation of Theorem 7.3.2 in [12] here. It makes use of some Banach-space machinery, and hence we must state some definitions. Let \mathfrak{M} be the Ba-

nach space of real measures on A with the variation norm, defined as follows:

$$\|\mu\| = \sup\{|\mu(E)| : E \subset A\}. \quad (36)$$

Further, let \mathfrak{B} be the Banach algebra of linear bounded operators $\mathfrak{M} \rightarrow \mathfrak{M}$. A Markov chain L_ϵ induces the operator $P_\epsilon \in \mathfrak{B}$ in the usual sense. A set $M \subset \mathfrak{M}$ is called a set of uniform convergence for an operator ($P \in \mathfrak{B}$) if $PM \subset M$ and there exists a function $\phi(n)$ exist such that

$$\sum_{n=1}^{\infty} \phi(n) =: \phi < \infty, \quad (37)$$

and a $y \in M$ exist such that

$$\|P^n x - y\| < \phi(n), \quad (38)$$

for all n and $x \in M$.

THEOREM 2. *Let P_q depend analytically on q as a function taking its values in the Banach algebra of operators \mathfrak{B} , and assume that the following conditions are met.*

- (i) *For the operator P_0 there exist two sets M_0 and M_1 of uniform convergence such that $M_1 \subset M_0$ and such that $\inf_{x \in M_0} \|x\| > 0$.*

(ii) There is a $q_0 > 0$ such that $P_q x \in M_2$, for all $|q| < q_0$ and any $x \in M_1$.

(iii) There is a $q_1 > 0$ such that

$$x_1 + \frac{(P_q - P_0)x_2}{\|P_q - P_0\|} \in M_2,$$

for $|q| < q_1$ and any $x_1, x_2 \in M_1$.

Then there is a $q_2 > 0$ such that, for $|q| < q_2$ and $x \in M_1$, the limit

$$\lim_{n \rightarrow \infty} P_q^n x = r(q)$$

exists and depends analytically on q .

We now prove Theorem 1.

PROOF. Remark that the Markov chain decomposable for $q = 0$. Indeed, there exist $\hat{p}_{**}, \tilde{p}_{**}$ such that

$$P_0 x(s', u') = \sum_{s, u} x(s, u) \hat{p}_{uu'} \tilde{p}_{ss'}.$$

Every state $(s, u), u > 0$ is transient. The Markov chain restricted to the smaller state space $(0, s)$, however, is irreducible and aperiodic. The subchain with transition probabilities $\tilde{p}_{ss'}$ describes the evolution of the number of active sessions. Denote its stationary distribution by $\sigma(s)$, and let $\sigma^*(s) = \sum_{k=s}^{\infty} \sigma(k)$. Further, the transient subchain $\hat{p}_{uu'}$ describes the buffer content evolution if $q = 0$: $\hat{p}_{uu'} = 1$ if $u' = (u - 1)^+$ and zero otherwise.

Picking the right sets M_1 and M_2 is the most difficult part of the proof. Choose $M_1 = \{\sigma\}$. This means that we only prove convergence if the initial distribution equals σ , but as we can establish the existence of a unique stationary distribution by means of a Loynes' argument (under the condition that $\lambda q < 1 - p$), the set is big enough for our purposes. Further, we choose $M_2 = \{x : \forall u, s : |x(u, s)| \leq \alpha \sigma(s) \sigma^*(u), \sum x(\cdot, \cdot) = 1\}$. Note that conditions (ii) and (iii) of Theorem 2 are trivially fulfilled for this choice of M_1 and M_2 , as is the fact that M_1 is a set of uniform convergence. Hence we must only verify that M_2 is a set of uniform convergence for the operator P_0 . This consists of two steps:

- We show that $PM_2 \subset M_2$. Indeed, let $x \in M_2$ then we have for $y = P_0 x$:

$$\begin{aligned} |y(j)| &= \left| \sum_i p_{ij} x(i) \right| \leq \sum_{i_1, i_2} x(i_1, i_2) \tilde{p}_{i_1 j_1} \hat{p}_{i_2 j_2}, \\ &\leq \alpha \sigma(i_1) \sigma^*((i_2 - 1)^+) \\ &\leq \alpha \sigma(i_1) \sigma^*(i_2) \end{aligned} \quad (39)$$

- For $y(s, u) = \pi(s)1(u = 0)$, we show that $\|P_0^n x - y\| < \phi(n)$, where $\sum_n \phi(n) =: \phi < \infty$. Indeed, pick

$\phi(n) = \alpha \sum_{i=n}^{\infty} \theta_{q_0}(i)$, and we see that

$$\begin{aligned} &\|P^n x - y\| \\ &= \sup_{j_1, j_2} \left| \sum_{i_1, i_2} x(i_1, i_2) \tilde{p}_{i_1 j_1}^n \hat{p}_{i_2 j_2}^n - \pi(j_1)1(j_2 = 0) \right| \\ &\leq \sup_{j_2} \left| \sum_{i_2} \hat{p}_{i_2 j_2}^n x_2(i_2) - 1(j_2 = 0) \right| \\ &= \left| \sum_{i=n}^{\infty} x_2(i) \right| \\ &\leq \phi(n) \end{aligned} \quad (40)$$

As $\sum_{i=0}^{\infty} i \sigma^*(i) < \infty$, we have $\phi < \infty$, as required.

This concludes the proof. \square

Note that from the analyticity of the stationary distribution, the analyticity of $P(x, z)$ in a neighbourhood of $x = z = q = 0$ is evident.

6. CONCLUSIONS

In this paper, we investigate a discrete-time queueing system with session-based arrivals with small packet arrival rate q within the sessions. As this system defies analytical solution, we evaluate Taylor-series expansions of the joint probability generating function of the system state at slot boundaries. We explicitly prove the analyticity and polynomial form of the steady-state generating function, so that our derivations are mathematically justified.

Numerical examples show that for not too heavy load ($\rho < 0.2$), even relatively low order series expansions have good accordance with simulation results. When combined with heavy-traffic or solutions for the case $q = 1$, we get good correspondence in an extended interval and reasonable correspondence over the entire load interval.

Acknowledgement

The second author is a postdoctoral fellow with the Research Foundation, Flanders (F.W.O.-Vlaanderen), Belgium.

7. REFERENCES

- [1] C. Blondia and O. Casals. Statistical multiplexing of VBR sources: A matrix-analytic approach. *Performance Evaluation*, 16(1-3):5-20, 1992.
- [2] C. Herrmann. The complete analysis of the discrete time finite DBMAP/G/1/N queue. *Performance Evaluation*, 43(2-3):95-121, February 2001.
- [3] G. U. Hwang and K. Sohraby. On the exact analysis of a discrete-time queueing system with autoregressive inputs. *Queueing Systems*, 43(1-2):29-41, January 2003.
- [4] F. Kamoun. The discrete-time queue with autoregressive inputs revisited. *Queueing Systems*, 54(3):185-192, November 2006.
- [5] D. Fiems, J. Walraevens, and H. Bruneel. Queues with Galton-Watson-type arrivals. In *Proceedings of the Belarussian Winter Workshop on Queueing Theory*, Minsk, Belarussia, 2009.
- [6] H. Bruneel. Packet delay and queue length for statistical multiplexers with low-speed access lines. *Computer Networks and ISDN Systems*, 25(12):1267-1277, 1993.

- [7] S. Wittevrongel. Discrete-time buffers with variable-length train arrivals. *Electronics Letters*, 34(18):1719–1721, 1998.
- [8] L. Hoflack, S. De Vuyst, S. Wittevrongel, and H. Bruneel. Analytic traffic model of web server. *Electronics Letters*, 44(1):61–62, 2008.
- [9] K. Tsoukatos and A. Makowski. Heavy traffic limits associated with $M/G/\infty$ input processes. *Queueing Systems*, 34(1/4):101 – 130, 2000.
- [10] M. Mandjes. *Large deviations for Gaussian queues: modelling communication networks*. John Wiley and Sons, 2007.
- [11] B. Vinck and H. Bruneel. Delay analysis for single server queues. *Electronics Letters*, 32(9):802–803, 1996.
- [12] G. Fayolle, V.A. Malyshev and M.V. Menshikov. *Topics in the constructive theory of countable Markov chains*. Cambridge University Press, 1995.
- [13] E. Altman, K. E. Avrachenkov, and R. Nunez-Queija. Perturbation analysis for denumerable Markov chains with application to queueing models. *Advances In Applied Probability*, 36(3):839–853, 2004.
- [14] V. A. Malyshev and M. V. Menshikov. Ergodicity, continuity and analyticity of countable Markov chains (in Russian). *Trudy Moscow Matematicheskogo Obschestva*, 39, 1979.
- [15] I. Kovalenko. Rare events in queueing systems – a survey. *Queueing Systems*, 16:1–49, 1994.
- [16] L. Hoflack, S. De Vuyst, S. Wittevrongel and H. Bruneel. Discrete-time buffer systems with session-based arrival streams. *Performance Evaluation* (2010), doi:10.1016/j.peva.2009.12.007.
- [17] B. Błaszczyszyn, T. Rolski, and V. Schmidt. *Light-traffic approximation in queues and related stochastic models*, pages 379–406. CRC Press, Probability and Stochastics Series, 1995.
- [18] M. I. Reiman and B. Simon. Open queueing systems in light traffic. *Mathematics of Operations Research*, 14(1):26–59, 1989.
- [19] B. Błaszczyszyn. Factorial moment expansion for stochastic systems. *Stochastic Processes and their Applications*, 56(2):321 – 335, 1995.
- [20] S. Asmussen. Light traffic equivalence in single-server queues. *The Annals of Applied Probability*, 2(3):555–574, 1992.
- [21] J. F. C. Kingman. On queues in heavy traffic. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(2):383–392, 1962.
- [22] S. Burton. A simple relationship between light and heavy traffic limits. *Oper. Res.*, 40(S2):342–345, 1992.
- [23] G. A. Baker Jr., P. Graves-Morris, P. A. Carruthers, Padé approximants, Part I: Basic Theory, *Encyclopedia of mathematics and its applications*, Vol. 13., 1981.