Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

A. Brandt[1] and M. Brandt

Asymptotic Results and a Markovian Approximation for the $M(n)/M(n)/s+GI$ System

[1] Humboldt-Universität zu Berlin

ZIB-Report 00-12 (April 2000)

Asymptotic Results and a Markovian Approximation for the $M(n)/M(n)/s + GI$ System

Andreas Brandt
Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin, Germany

Manfred Brandt
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Takustr. 7, D-14195 Berlin, Germany

Abstract

In this paper for the $M(n)/M(n)/s + GI$ system, i.e. for a s-server queueing system where the calls in the queue may leave the system due to impatience, we present new asymptotic results for the intensities of calls leaving the system due to impatience and a Markovian system approximation where these results are applied. Furthermore, we present a new proof for the formulae of the conditional density of the virtual waiting time distributions, recently given by Movaghar for the less general $M(n)/M/s+GI$ system. Also we obtain new explicit expressions for refined virtual waiting time characteristics as a byproduct.

Mathematics Subject Classification (MSC 1991): 60K25, 68M20, 60G10.

Keywords: $M(n)/M(n)/s+GI$ system; limited waiting times; finite buffer; virtual waiting times; blocking-, abandon probability; impatience rates; asymptotic results; $M(n)/M(n)/s$ system with waiting place dependent impatient rates; Markovian approximation.

1 Introduction

The $M(n)/M(n)/s + GI$ system is a s-server queueing system with a (potential) unlimited waiting room with FCFS queueing discipline where the calls waiting in the queue for service are impatient, cf. Figure 1. The arrival and service processes are allowed to be state dependent with respect to the number n of calls in the system, but the cumulative service rate is assumed to be constant for $n > s$. We assume that the sequence of the arrival rates λ_n is bounded and that $\lambda_n > 0$ for $n \geq 0$ or that there exists a positive integer k such that $\lambda_n > 0$ for $0 \leq n < s + k$ and $\lambda_n \equiv 0$ for $n \geq s + k$. Concerning the cumulative rate μ_n of finishing service we assume $\mu_0 = 0$ and $\mu_n \equiv \mu_* > 0$ for $n > s$. Each call arriving at the system has a maximal waiting time I. If the offered waiting time W^o (i.e. the time which a call would have to wait for accessing a server if it were sufficiently patient) exceeds I, then the call departs from the system after having waited time I.

The maximal waiting times are assumed to be i.i.d. with a general distribution $C(u) := P(I \leq u)$, $u \in \mathbb{R}_+$, which may be defective, i.e., $P(I = \infty) > 0$ is not excluded. In the notation $M(n)/M(n)/s + GI$ the first $M(n)$ denotes the arrival process, the second $M(n)$ the service process depending on $\min(n, s+1)$ only, i.e., on the number of busy servers and additionally whether there are calls waiting for service. The symbol GI stands for the i.i.d. maximal waiting times. Note, that if $\lambda_n > 0$ for $0 \leq n < s + k$ and $\lambda_n \equiv 0$ for $n \geq s + k$, then we have the case of a limited waiting room with k waiting places ($M(n)/M(n)/s/k + GI$ system). If $\mu_n = \min(n, s)\mu$ for $n \geq 0$ ($\mu_* := s\mu$), then we have a $M(n)/M/s + GI$ system and if additionally $\lambda_n \equiv \lambda > 0$ then a $M/M/s + GI$ system.

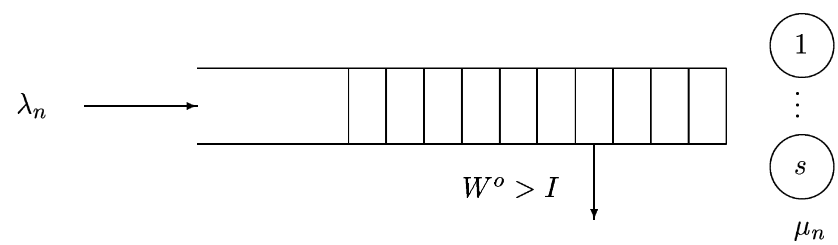

Figure 1: *The $M(n)/M(n)/s+GI$ system with impatient calls and state dependent arrival and service rates, where n denotes the number of calls in the system, W^o the offered waiting time and I the maximal waiting time.*

The general $M(n)/M(n)/s + GI$ system is treated in [BB1], where in particular the density of the detailed state process and the occupancy distribution are derived and an application for the use of general state dependent departure rates is given, cf. also [BB2]. Independently, Movaghar [Mov] derived the occupancy distribution and the conditional offered virtual waiting time distribution besides other quantities for the $M(n)/M/s + GI$ system. For further results on special cases and results for queues with impatience we refer to [BH], [Ju1], [Mov], [BB1] and the references therein.

The main results and organization of the paper are as follows: In Section 2 various performance measures for the $M(n)/M(n)/s + GI$ system are recalled from [BB1] since they are used later in the paper. Section 3 deals with the offered virtual waiting time distribution for the $M(n)/M(n)/s + GI$ system. Recently, for the $M(n)/M/s + GI$ system Movaghar [Mov] derived an explicit formula for the density of the offered virtual waiting time if ℓ calls are in the queue by using tricky probabilistic arguments. We give a new rigorous proof of this result for the more general $M(n)/M(n)/s + GI$ system by using explicit results from [BB1]. The proof given here is completely different to those in [Mov] for the special case of a $M(n)/M/s+GI$ system. The advantage of our alternative proof lies in the fact that it is not quite clear whether the arguments in [Mov] can be extended to state dependent service rates and that some of the arguments are a little bit vague or not outlined rigorously. Although our proof is more analytical we obtain as a byproduct also new refined virtual waiting time characteristics, cf. Eq. (3.10). In Section 4 monotonicity and asymptotic results are derived for the departure intensities of calls due to impatience. Using these results in Section 5 we construct a simple Markovian approximation for the $M(n)/M(n)/s+GI$ system. The idea is to replace the individual maximal waiting times of the calls by waiting place dependent impatience rates, i.e., with each waiting place, which are numbered by $i=1,2,\ldots$, there is associated an impatience rate $\beta_i\,(\geq 0)$. This system – which is a simple Markov model – is denoted by $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$, where $M(\beta_i)_{i=1}^{\infty}$ stands for the waiting place dependent impatient rates and $M(n)/M(n)$ has the same meaning as above. The relevant performance measures for this system can easily be obtained since the process of the number of calls in the system is a simple birth-death process. The impatience rates β_i can be chosen such that the occupancy distribution of the $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system is fitted to those of the $M(n)/M(n)/s+GI$ system. If $\lambda_{i+s-1}>0$ then β_i is uniquely determined. In general, the impatience rates β_i can be chosen such that they only depend on i, μ_* and on the distribution $C(u)$ of the maximal waiting times. The fitting of the occupancy distribution implies the fitting of other performance measures. From the asymptotic results given in Section 4 it follows that the β_i converge to the intensity corresponding to an exponential fitting if the m-th moment of the distribution $C(u)$ is finite for any $m>2$. However, numerical examples given in Section 5 show that the β_i may significantly differ from their limit for smaller i. Also the relation to the fitting technique used in [BB2] is discussed. The proposed approximation is useful if in a given queueing network one wants to approximate a $\cdot/M(n)/s+GI$ node by a simple birth-death node. The use of the approximation lies in reducing the numerical complexity for computing performance measures in queueing networks.

2 State process and performance measures for the $M(n)/M(n)/s+GI$ system

In this section we recall some notation and results for the $M(n)/M(n)/s+GI$ system from [BB1]. Throughout this section we assume that the queueing system is stable, where the stability condition will be given later. If n calls are in the system then $\ell := (n-s)_+$ calls are waiting in the queue for service. (The notation $\ell := (n-s)_+$ will also be used in the following.) We number the waiting calls according to their positions in the queue. By the FCFS discipline the first call in the queue will be potentially the next for service. Let

$N(t)$	–	number of calls in the system at time t;
$L(t) := (N(t)-s)_+$	–	number of waiting calls at time t;
$(X_1(t),\dots,X_{L(t)}(t))$	–	vector of the residual maximal waiting times of waiting calls ordered according to their positions in the queue at time t;
$(I_1(t),\dots,I_{L(t)}(t))$	–	vector of the original maximal waiting times of the waiting calls ordered according to their positions in the queue at time t;
$p(n) := P(N(t)=n)$	–	stationary distribution of the number of calls in the system;

$$P(n;\, x_1,\dots,x_\ell;\, u_1,\dots,u_\ell) := P(N(t)=n;\ X_1(t)\le x_1,\dots,X_\ell(t)\le x_\ell;\ I_1(t)\le u_1,\dots,I_\ell(t)\le u_\ell)$$

– stationary distribution on $\{N(t)=n\}$.

For fixed $n>s$ the support of $P(n;x_1,\dots,x_\ell;u_1,\dots,u_\ell)$ is contained in

$$\Omega_\ell := \{(x_1,\dots,x_\ell;u_1,\dots,u_\ell)\in\mathbb{R}_+^{2\ell}\ :\ u_1-x_1\ge\dots\ge u_\ell-x_\ell\ge 0\}.$$

Density of the state process. Assume that the following assumptions are satisfied:

(A1) $C(u)$ is non-defective, i.e. $\lim_{u\to\infty} C(u)=1$,

(A2) $C(u)$ has a continuous density $c(u)$.

Then the density

$$p(n;x_1,\dots,x_\ell;u_1,\dots,u_\ell) := \frac{\partial^{2\ell}}{\partial x_1\dots\partial x_\ell\partial u_1\dots\partial u_\ell}P(n;x_1,\dots,x_\ell;u_1,\dots,u_\ell)$$

of the state process $(N(t),X_1(t),\dots,X_{L(t)}(t),I_1(t),\dots,I_{L(t)}(t))$ with right-continuous sample paths is continuous on Ω. In [BB1] these densities are obtained by solving a system of integral equations explicitly, cf. [BB1] Eqs. (2.10) and (2.17):

$$p(n) = g\Big(\prod_{i=0}^{n-1}\lambda_i\Big)\Big(\prod_{i=n+1}^{s}\mu_i\Big), \qquad n=0,1,\dots,s;$$

$$p(n;x_1,\dots,x_\ell;u_1,\dots,u_\ell) = \mathbb{1}\{(x_1,\dots,x_\ell;u_1,\dots,u_\ell)\in\Omega_\ell\}\, g\Big(\prod_{i=0}^{n-1}\lambda_i\Big)\Big(\prod_{i=1}^{\ell}c(u_i)\Big)e^{-\mu_*(u_1-x_1)}, \qquad n=s+1,s+2,\dots \tag{2.1}$$

The normalization factor g is given by

$$g^{-1} = \sum_{j=0}^{s-1} \Big(\prod_{i=0}^{j-1} \lambda_i\Big)\Big(\prod_{i=j+1}^{s} \mu_i\Big) + \sum_{j=0}^{\infty} \Big(\prod_{i=0}^{s+j-1} \lambda_i\Big) F_j, \tag{2.2}$$

where

$$F_j := \frac{1}{j!} \int_0^\infty F(\xi)^j e^{-\xi} \mathrm{d}\xi, \quad j = 0, 1, 2, \dots, \tag{2.3}$$

or alternatively

$$F_0 = 1; \quad F_j = \frac{1}{(j-1)!} \int_0^\infty F(\xi)^{j-1} F'(\xi) e^{-\xi} \mathrm{d}\xi, \quad j = 1, 2, \dots; \tag{2.4}$$

$$F(\xi) := \int_0^{\xi/\mu_*} (1 - C(\eta)) \mathrm{d}\eta, \quad \xi \in \mathbb{R}_+. \tag{2.5}$$

From (2.1) for the (marginal) densities of the vector $(N(t), X_1(t), \dots, X_{L(t)}(t))$ we obtain immediately

$$\begin{aligned} p_x(n; x_1, \dots, x_\ell) &:= \frac{\partial^\ell}{\partial x_1 \dots \partial x_\ell} P(N(t) = n; X_1(t) \le x_1, \dots, X_\ell(t) \le x_\ell) \\ &= g_n \int_{\xi_1 \ge \dots \ge \xi_\ell \ge 0} e^{-\mu_* \xi_1} \Big(\prod_{i=1}^{\ell} c(\xi_i + x_i)\Big) \mathrm{d}\xi_1 \dots \mathrm{d}\xi_\ell, \\ &\qquad n = s+1, s+2, \dots, \end{aligned} \tag{2.6}$$

where

$$g_n := g \prod_{i=0}^{n-1} \lambda_i \tag{2.7}$$

and g is given by (2.2). The support of this density is contained in $\mathbb{R}_+^\ell$.

Stability condition and occupancy distribution. For a general maximal waiting time distribution $C(u)$ in [BB1] Theorem 3.1 it is shown that the system is *stable* iff the right-hand side of (2.2) is finite, i.e., the stability condition reads

$$J := \sum_{j=0}^{\infty} \Big(\prod_{i=s}^{s+j-1} \lambda_i\Big) F_j < \infty, \tag{2.8}$$

and in case of a stable system the occupancy distribution is given by

$$p(n) = \begin{cases} g_n \prod_{i=n+1}^{s} \mu_i, & n = 0, 1, \ldots, s, \\ g_n F_{n-s}\,, & n = s+1, s+2, \ldots \end{cases} \tag{2.9}$$

Remark 2.1 1. For the $M(n)/M/s + GI$ system the corresponding stability condition (2.8) and occupancy distribution (2.9) (as well as other quantities) are given in [Mov], where a completely different approach is used than in [BB1]. However, it seems that the density (2.1) can not be obtained by the approach given in [Mov].

2. In case of a limited waiting room of capacity k, i.e. $\lambda_n \equiv 0$ for $n \geq s+k$, the stability condition (2.8) is always fulfilled.

3. For the $M/M(n)/s + GI$ system, i.e. if $\lambda_n \equiv \lambda > 0$ for $n = 0, 1, 2, \ldots$, from (2.8), (2.3), (2.5) we obtain

$$J = \int_0^\infty \exp\left(\lambda F(\xi) - \xi\right) \mathrm{d}\xi = \int_0^\infty \exp\left(\mu_* \int_0^{\xi/\mu_*} \left(\frac{\lambda}{\mu_*}(1 - C(\eta)) - 1\right) \mathrm{d}\eta\right) \mathrm{d}\xi,$$

and thus the system is stable, i.e. $J < \infty$, iff

$$\frac{\lambda}{\mu_*} \lim_{u \to \infty} (1 - C(u)) < 1. \tag{2.10}$$

For the $M/M/s + GI$ system the stability condition (2.10) was given in [BH]. In the general case we obtain the above expression for J as an upper bound if $\lambda_n \leq \lambda$ for $n = s, s+1, \ldots$ Thus in the general case (2.10) may be used as a sufficient stability condition. In particular, the $M(n)/M(n)/s + GI$ system is stable if $C(u)$ is non-defective.

Departure rates due to impatience. Assuming (A1), (A2) the intensity α_ℓ of calls leaving the system due to impatience conditioned upon $\ell := n - s > 0$ calls are in the queue is given by

$$\alpha_\ell = \frac{1}{p(s+\ell)} \sum_{i=1}^{\ell} \int_{\mathbb{R}_+^{2\ell-1}} p(s+\ell; x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_\ell; u_1, \ldots, u_\ell) \, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_\ell \mathrm{d}u_1 \ldots \mathrm{d}u_\ell$$

provided $p(s+\ell) > 0$. Using (2.9), (2.7), (2.3)-(2.5) and the last equation on p. 10 in [BB1] for $x \downarrow 0$ we obtain

$$\alpha_\ell = \frac{F_{\ell-1}}{F_\ell} - \mu_*, \quad \ell = 1, 2, \ldots, k, \tag{2.11}$$

where $k := \sup\{j \in \mathbb{Z}_+ : \lambda_{j+s-1} > 0\}$. In case of a limited waiting room ($k < \infty$) let α_ℓ be defined by (2.11) for $\ell > k$. Since the α_ℓ depend only on ℓ, $C(u)$ and μ_*, in case of a $M(n)/M/s + GI$ system one has the same α_ℓ with $\mu_* = s\mu$ if $1/\mu$ is the mean call service time. For this latter case (2.11) is given in [Mov], too. For particular $C(u)$ the previous formulae can be specialized appropriately, cf. the preprint version of [BB1] and [Mov].

Performance measures. The *cumulative mean call arrival intensity* Λ is given by

$$\Lambda = \sum_{n=0}^{\infty} \lambda_n p(n). \tag{2.12}$$

Since the λ_n are bounded, Λ is finite. The probability p_I that a call will leave the system due to impatience later (*abandon probability*) and the probability p_W that a typical arriving call has to wait for service are given by

$$p_I = \frac{1}{\Lambda} \sum_{n=s+1}^{\infty} \alpha_{n-s} p(n), \quad p_W = 1 - \frac{1}{\Lambda} \sum_{n=0}^{s-1} \lambda_n p(n). \tag{2.13}$$

The mean waiting time EW in the queue of a typical arriving call is given by Little's formula

$$EW = \frac{1}{\Lambda} \sum_{n=s+1}^{\infty} (n-s) p(n). \tag{2.14}$$

In case of a $M/M(n)/s/k + GI$ system, the probability p_B that an arriving call will be blocked (*blocking probability*) is given by

$$p_B = p(s+k). \tag{2.15}$$

3 Virtual waiting time for the $M(n)/M(n)/s + GI$ system

For a given stable $M(n)/M(n)/s + GI$ system in steady state let V be the *offered virtual waiting time* of a call, i.e. the time which an arriving virtual call at $t = 0$ would have to wait for accessing a server if it were sufficiently patient, and V_n be the offered virtual waiting time given that n calls are in the system, i.e. $P(V_n \le u) = P(V \le u \,|\, N(0) = n)$. We make the convention that the virtual call arriving at $t = 0$ acts like a real call, i.e., it is incorporated into the dynamics of the system after $t = 0$ but it is not counted in $N(t)$. In particular, if $N(0-0) \ge s$ then immediately after $t = 0$ the service rate is μ_*. The dynamics of the model and the assumptions imply $V_n = 0$ for $n = 0, \ldots, s-1$ and $V_n > 0$ for $n = s, s+1, \ldots$ as well as that the conditional densities $v(u|n) = \frac{\mathrm{d}}{\mathrm{d}u} P(V_n \le u)$, $u \in \mathbb{R}_+$, $n = s, s+1, \ldots$, exist.

Recently, for the $M(n)/M/s + GI$ system Movaghar [Mov] formula (3.17) proved – in our notations – the following result

$$v(u|n) = \mu_* e^{-\mu_* u} \frac{F(\mu_* u)^{n-s}}{(n-s)!\, F_{n-s}}, \quad u > 0, \quad n = s, s{+}1, \ldots, \tag{3.1}$$

where $\mu_* := s\mu$, $1/\mu$ denotes the mean service time, and $F(\xi)$, F_j are given by (2.3)-(2.5). The proof bases on intricate probabilistic arguments. The densities $v(n;u) = \frac{\mathrm{d}}{\mathrm{d}u} P(N(0) = n; V \leq u) = p(n)v(u|n)$, $n = s, s{+}1, \ldots$, of the offered virtual waiting time V on $\{N(0) = n\}$, which are more convenient in the following considerations, are obtained from (2.9), (2.4) and (3.1) as

$$v(n;u) = g_n \mu_* e^{-\mu_* u} \frac{F(\mu_* u)^{n-s}}{(n-s)!}, \quad u > 0, \quad n = s, s{+}1, \ldots, \tag{3.2}$$

and the distribution of V is given by

$$P(V = 0) = \sum_{n=0}^{s-1} p(n); \quad \frac{\mathrm{d}}{\mathrm{d}u} P(V \leq u) = \sum_{n=s}^{\infty} v(n;u), \quad u > 0. \tag{3.3}$$

In this section by using complete different arguments than in [Mov] we prove that (3.2) remains valid for the more general $M(n)/M(n)/s + GI$ system:

Theorem 3.1 For a general stable $M(n)/M(n)/s + GI$ system the densities $v(n;u)$, $n = s, s{+}1, \ldots$, are given by (3.2).

Proof. The proof is divided into several steps. We consider the stable $M(n)/M(n)/s + GI$ system in steady state with its stationary state process $(N(t), X_1(t), \ldots, X_{L(t)}(t), I_1(t), \ldots, I_{L(t)}(t))$, $t \in \mathbb{R}$, where at $t = 0$ there arrives a virtual call of unlimited maximal waiting time, i.e., at $t = 0{+}0$ there are $N(0){+}1$ calls in the system relevant for the dynamics.

1. If $N(0) = s$ then the time until a server becomes free and the virtual call goes into service is exponentially distributed with parameter μ_* in view of $\mu_n \equiv \mu_*$ for $n > s$. This implies

$$v(s;u) = p(s)\mu_* e^{-\mu_* u}, \quad u \in \mathbb{R}_+, \tag{3.4}$$

which is just (3.2) for $n = s$ because of (2.9). (Note, the calls arriving after $t = 0$ do not have any impact on the service process until the begin of the service for the virtual call.) In the succeeding steps we will prove (3.2) for $n > s$, i.e. if $\ell = n - s > 0$ calls are in the queue.

2. In this step we look into the details of the service and impatience process. For a given vector $k = (k_1, \ldots, k_h) \in \{0,1\}^h$ let $|k| := k_1 + \ldots + k_h$ and, if $|k| > 0$, let $1 \leq i_1 < i_2 < \ldots < i_{|k|} \leq h$ the indices of those components of k where $k_i = 1$, i.e., $k_i = 1$ for $i \in \{i_1, \ldots, i_{|k|}\}$ and $k_i = 0$ for $i \in \{1, \ldots, h\} \backslash \{i_1, \ldots, i_{|k|}\}$. By convention let $i_0 := 0$, $i_{|k|+1} := h+1$ in the following.

2.1. According to the definition of the state process we index the arriving virtual call at $t = 0$ by $\ell + 1$ and assign to it the residual maximal waiting time $X_{\ell+1}(0) := \infty$ according to its unlimited maximal waiting time. Then, as long as there are calls in the queue, the process of served calls and hence of calls going into service is a Poisson process with intensity μ_*. Also, since the service process during the interval $(0, V]$ will not be affected by newly arriving calls, we can stop the arrival process after $t = 0$ for determining $v(s+\ell; u)$.

2.2. On $\{N(0) = s + \ell\}$ let $X_i := X_i(0)$, $i = 1, \ldots, \ell + 1$, and $K_i = 1\,(0)$, $i = 1, \ldots, \ell$, if the i-th call in the queue at $t = 0$ later goes into service (gets lost due to impatience). Notice, the virtual call at position $\ell + 1$ in the queue at $t = 0$ always goes later into service, since its maximal waiting time is unlimited. The state space of the vector $K = (K_1, \ldots, K_\ell)$ is $\{0,1\}^\ell$. Consider for a given $k \in \{0,1\}^\ell$ the event $\{K = k\}$. Then the $|k|$ calls at the positions $i_1, \ldots, i_{|k|}$ in the queue at $t = 0$ will be served later and the remaining $\ell - |k|$ calls will get lost due to impatience. Denote by $W_0 < \ldots < W_{|k|}$ the consecutive time instants after $t = 0$ where calls go into service, i.e., since we are on $\{K = k\}$, at W_j, $0 \leq j < |k|$, the call with index i_{j+1} goes into service and at $W_{|k|}$ finally the virtual call, i.e., $V = W_{|k|}$. As mentioned above, the W_j are part of a Poisson process of intensity μ_*. Denote further by $Y_1, \ldots, Y_{\ell-|k|}$ the vector of the residual maximal waiting times at $t = 0$ of all calls leaving the queue due to impatience later. By the system dynamics it holds

$$(Y_1, \ldots, Y_{\ell-|k|}) = (X_1, \ldots, X_{i_1-1}, X_{i_1+1}, \ldots, X_{i_{|k|}-1}, X_{i_{|k|}+1}, \ldots, X_\ell) \tag{3.5}$$

and

$$\{K = k\} = \bigcap_{j=0}^{|k|} \left(\left(\bigcap_{r=i_j+1}^{i_{j+1}-1} \{X_r \leq W_j\} \right) \bigcap \{X_{i_{j+1}} > W_j\} \right), \tag{3.6}$$

where on the r.h.s. the inequality $X_{i_{j+1}} > W_j$ is always fulfilled for $j = |k|$ in view of $X_{i_{|k|+1}} = X_{\ell+1} = \infty > W_{|k|}$.

3. In this step, which is the main part of the proof, we assume that (A1) and (A2) hold, thus the densities $p_x(n; x_1, \ldots, x_\ell)$ of the residual maximal waiting times are given by (2.6), (2.7).

3.1. For a given $k \in \{0,1\}^\ell$ the densities

$$p_{w,y}(k; w_0, \ldots, w_{|k|}; y_1, \ldots, y_{\ell-|k|}) := \frac{\partial^{\ell+1}}{\partial w_0 \ldots \partial w_{|k|} \partial y_1 \ldots \partial y_{\ell-|k|}}$$

$$P(K = k; W_0 \leq w_0, \ldots, W_{|k|} \leq w_{|k|}; Y_1 \leq y_1, \ldots, Y_{\ell-|k|} \leq y_{\ell-|k|})$$

have the support, cf. (3.6),

$$\Omega_k := \{(w_0,\dots,w_{|k|};y_1,\dots,y_{\ell-|k|}) : w_0 \le \dots \le w_{|k|}, \prod_{j=0}^{|k|}\prod_{r=i_j+1}^{i_{j+1}-1} \mathbb{1}\{x_r \le w_j\} = 1\}, \tag{3.7}$$

where the transformation of variables

$$(y_1,\dots,y_{\ell-|k|}) := (x_1,\dots,x_{i_1-1},x_{i_1+1},\dots,x_{i_{|k|}-1},x_{i_{|k|}+1},\dots,x_\ell)$$

is used, cf. (3.5). From (3.6) and since $W_0,\dots,W_{|k|}$ are the first $|k|+1$ points of a Poisson process of intensity μ_*, for $(w_0,\dots,w_{|k|};y_1,\dots,y_{\ell-|k|})\in\Omega_k$ by the law of total probability we obtain

$$p_{w,y}(k;w_0,\dots,w_{|k|};y_1,\dots,y_{\ell-|k|}) = \left(\prod_{j=0}^{|k|}\mu_* e^{-\mu_*(w_j-w_{j-1})}\right)$$
$$\int\limits_{\mathbb{R}_+^{|k|}} \left(\prod_{j=0}^{|k|-1}\mathbb{1}\{x_{i_{j+1}} > w_j\}\right) p_x(s+\ell;x_1,\dots,x_\ell)\mathrm{d}x_{i_1}\dots\mathrm{d}x_{i_{|k|}}, \tag{3.8}$$

where the convention $w_{-1} := 0$ and the same transformation of variables is used. Applying the explicit expression (2.6) for $p_x(s+\ell;x_1,\dots,x_\ell)$ and taking into account (3.7), from (3.8) by using Fubini's theorem and integrating with respect to $\mathrm{d}y_1\dots\mathrm{d}y_{\ell-|k|}$ $(= \mathrm{d}x_1\dots\mathrm{d}x_{i_1-1}\mathrm{d}x_{i_1+1}\dots\mathrm{d}x_{i_{|k|}-1}\mathrm{d}x_{i_{|k|}+1}\dots\mathrm{d}x_\ell)$ and $\mathrm{d}x_{i_1}\dots\mathrm{d}x_{i_{|k|}}$ it follows

$$p_w(k;w_0,\dots,w_{|k|}) := \frac{\partial^{|k|+1}}{\partial w_0\dots\partial w_{|k|}}P(K=k;W_0\le w_0,\dots,W_{|k|}\le w_{|k|})$$
$$= g_{s+\ell}\,\mu_*^{|k|+1}e^{-\mu_* w_{|k|}} \int\limits_{\xi_1\ge\dots\ge\xi_\ell\ge 0} e^{-\mu_*\xi_1}$$
$$\left(\prod_{j=0}^{|k|}\left(\prod_{r=i_j+1}^{i_{j+1}-1}\left(\bar{C}(\xi_r)-\bar{C}(\xi_r+w_j)\right)\right)\bar{C}(\xi_{i_{j+1}}+w_j)\right)\mathrm{d}\xi_1\dots\mathrm{d}\xi_\ell, \tag{3.9}$$

where $\bar{C}(u) := 1-C(u)$ is the complementary distribution function and where we use the notation $\xi_{h+1} := -\infty$ and $\bar{C}(-\infty) := 1$ as also in the following. The support of $p_w(k;w_0,\dots,w_{|k|})$ is $\Omega_{k,w} = \{(w_0,\dots,w_{|k|}) : 0\le w_0\le\dots\le w_{|k|}\}$.
3.2. On $\{K=k\}$ the offered virtual waiting time V at $t=0$ is given by $W_{|k|}$. By integrating the density (3.9) with respect to $\mathrm{d}w_0\dots\mathrm{d}w_{|k|-1}$ we obtain the density of V on $\{K=k\}$

$$p(k;u) := \frac{\mathrm{d}}{\mathrm{d}u}P(K=k;V\le u) = g_{s+\ell}\,\mu_* e^{-\mu_* u}I(k;u), \tag{3.10}$$

where

$$I(k;u) := \mu_*^{|k|} \int\limits_{0\le w_0\le\ldots\le w_{|k|}} \int\limits_{\xi_1\ge\ldots\ge\xi_\ell\ge 0} e^{-\mu_*\xi_1} \Big(\prod_{j=0}^{|k|}\Big(\prod_{r=i_j+1}^{i_{j+1}-1}\big(\bar C(\xi_r)-\bar C(\xi_r+w_j)\big)\Big)\bar C(\xi_{i_{j+1}}+w_j)\Big) \mathrm{d}\xi_1\ldots\mathrm{d}\xi_\ell \mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1} \tag{3.11}$$

and in (3.11) the convention $w_{|k|} := u$ is used.

3.3. Now, define for $m\in\mathbb{Z}_+$, $h\in\mathbb{Z}_+\setminus\{0\}$, $k\in\{0,1\}^h$ and $u\in\mathbb{R}_+$

$$I_m(k;u) := \mu_*^{|k|} \int\limits_{0\le w_0\le\ldots\le w_{|k|}} \int\limits_{\xi_1\ge\ldots\ge\xi_h\ge 0} e^{-\mu_*\xi_1}\frac{1}{m!}\Big(\int\limits_{\xi_1}^{\xi_1+w_0}\bar C(\eta)\mathrm{d}\eta\Big)^m \Big(\prod_{j=0}^{|k|}\Big(\prod_{r=i_j+1}^{i_{j+1}-1}\big(\bar C(\xi_r)-\bar C(\xi_r+w_j)\big)\Big)\bar C(\xi_{i_{j+1}}+w_j)\Big) \mathrm{d}\xi_1\ldots\mathrm{d}\xi_h \mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1}, \tag{3.12}$$

where the convention $w_{|k|} := u$ is used again, and for $m\in\mathbb{Z}_+$ and $u\in\mathbb{R}_+$

$$I_m(u) := \frac{1}{m!}\Big(\int\limits_0^u \bar C(\eta)\mathrm{d}\eta\Big)^m. \tag{3.13}$$

In case of $|k| = 0$ the integral with respect to $\mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1}$ does not occur in (3.12), and the r.h.s. of (3.12) reduces to

$$I_m(0,\ldots,0;u) = \int\limits_{\xi_1\ge\ldots\ge\xi_h\ge 0} e^{-\mu_*\xi_1}\frac{1}{m!}\Big(\int\limits_{\xi_1}^{\xi_1+u}\bar C(\eta)\mathrm{d}\eta\Big)^m \Big(\prod_{r=1}^{h}\big(\bar C(\xi_r)-\bar C(\xi_r+u)\big)\Big)\mathrm{d}\xi_1\ldots\mathrm{d}\xi_h. \tag{3.14}$$

Note, for $m = 0$ and $h = \ell$ the r.h.s. of (3.12) corresponds to those of (3.11), i.e.

$$I(k;u) = I_0(k;u), \quad k\in\{0,1\}^\ell, \quad u\in\mathbb{R}_+. \tag{3.15}$$

3.4. The $I_m(k;u)$ and $I_m(u)$ satisfy a recursion which is crucial and will be proved next.

Lemma 3.2 For $m\in\mathbb{Z}_+$, $h\in\mathbb{Z}_+\setminus\{0\}$, $k\in\{0,1\}^h$, $u\in\mathbb{R}_+$ it holds

$$I_{m+1}(k;u)=I_m(0,k;u)+I_m(1,k;u), \tag{3.16}$$

and for $m\in\mathbb{Z}_+$, $h\in\mathbb{Z}_+\setminus\{0\}$, $u\in\mathbb{R}_+$ it holds

$$I_{m+h}(u)=\sum_{k\in\{0,1\}^h} I_m(k;u). \tag{3.17}$$

Proof. Let $m\in\mathbb{Z}_+$ and $k\in\{0,1\}^h$. Let $k^{(\nu)}:=(\nu,k)\in\{0,1\}^{h+1}$, $\nu\in\{0,1\}$. Remember, $i_1,\dots,i_{|k|}$ denote the indices of k in ascending order where the components take the value 1. Analogously denote by $i_1^{(\nu)},\dots,i_{|k^{(\nu)}|}^{(\nu)}$ the corresponding indices for $k^{(\nu)}$. From $k^{(\nu)}=(\nu,k)$ it follows that $|k^{(0)}|=|k|$ and $i_j^{(0)}=i_j+1$, $j=1,\dots,|k^{(0)}|+1$, as well as $|k^{(1)}|=|k|+1$ and $i_j^{(1)}=i_{j-1}+1$, $j=1,\dots,|k^{(1)}|+1$. Using the convention $w_{|k^{(0)}|}:=u$ from (3.12) for $I_m(k^{(0)};u)$ we obtain

$$\begin{aligned}I_m(k^{(0)};u)=\mu_*^{|k^{(0)}|}&\int\limits_{0\le w_0\le\dots\le w_{|k^{(0)}|}}\ \int\limits_{\xi_1\ge\dots\ge\xi_{h+1}\ge 0} e^{-\mu_*\xi_1}\frac{1}{m!}\Bigg(\int\limits_{\xi_1}^{\xi_1+w_0}\bar C(\eta)\mathrm{d}\eta\Bigg)^m\\ &\Bigg(\prod_{j=0}^{|k^{(0)}|}\Bigg(\prod_{r=i_j^{(0)}+1}^{i_{j+1}^{(0)}-1}\Big(\bar C(\xi_r)-\bar C(\xi_r+w_j)\Big)\Bigg)\bar C(\xi_{i_{j+1}^{(0)}}+w_j)\Bigg)\\ &\mathrm{d}\xi_1\dots\mathrm{d}\xi_{h+1}\mathrm{d}w_0\dots\mathrm{d}w_{|k^{(0)}|-1}.\end{aligned} \tag{3.18}$$

Taking into account $|k^{(0)}|=|k|$ and $i_j^{(0)}=i_j+1$, $j=1,\dots,|k^{(0)}|+1$, and substituting ξ_i by ξ_{i-1} yields

$$\begin{aligned}I_m(k^{(0)};u)=\mu_*^{|k|}&\int\limits_{0\le w_0\le\dots\le w_{|k|}}\ \int\limits_{\xi_0\ge\dots\ge\xi_h\ge 0} e^{-\mu_*\xi_0}\frac{1}{m!}\Bigg(\int\limits_{\xi_0}^{\xi_0+w_0}\bar C(\eta)\mathrm{d}\eta\Bigg)^m\\ &\Big(\bar C(\xi_0)-\bar C(\xi_0+w_0)\Big)\Bigg(\prod_{j=0}^{|k|}\Bigg(\prod_{r=i_j+1}^{i_{j+1}-1}\Big(\bar C(\xi_r)-\bar C(\xi_r+w_j)\Big)\Bigg)\bar C(\xi_{i_{j+1}}+w_j)\Bigg)\\ &\mathrm{d}\xi_0\dots\mathrm{d}\xi_h\mathrm{d}w_0\dots\mathrm{d}w_{|k|-1}.\end{aligned} \tag{3.19}$$

By using Fubini's theorem, $i_1^{(1)}=1$, the convention $w_{|k^{(1)}|}:=u$ and by integrating with respect to w_0 from (3.12) for $I_m(k^{(1)};u)$ we find

$$I_m(k^{(1)};u) = \mu_*^{|k^{(1)}|} \int\limits_{0\le w_1\le\ldots\le w_{|k^{(1)}|}} \int\limits_{\xi_1\ge\ldots\ge\xi_{h+1}\ge 0} e^{-\mu_*\xi_1} \frac{1}{(m+1)!}\Big(\int\limits_{\xi_1}^{\xi_1+w_1} \bar{C}(\eta)\mathrm{d}\eta\Big)^{m+1}$$
$$\Big(\prod_{j=1}^{|k^{(1)}|}\Big(\prod_{r=i_j^{(1)}+1}^{i_{j+1}^{(1)}-1}\big(\bar{C}(\xi_r)-\bar{C}(\xi_r+w_j)\big)\Big)\bar{C}(\xi_{i_{j+1}^{(1)}}+w_j)\Big)$$
$$\mathrm{d}\xi_1\ldots\mathrm{d}\xi_{h+1}\mathrm{d}w_1\ldots\mathrm{d}w_{|k^{(1)}|-1}. \tag{3.20}$$

Taking into account $|k^{(1)}| = |k|+1$ and $i_j^{(1)} = i_{j-1}+1$, $j = 1,\ldots,|k^{(1)}|+1$, as well as substituting w_i by w_{i-1} (in particular $w_{|k^{(1)}|} = u$ by $w_{|k|} = u$) and ξ_i by ξ_{i-1} yields

$$I_m(k^{(1)};u) = \mu_*^{|k|} \int\limits_{0\le w_0\le\ldots\le w_{|k|}} \int\limits_{\xi_0\ge\ldots\ge\xi_h\ge 0} \mu_* e^{-\mu_*\xi_0} \frac{1}{(m+1)!}\Big(\int\limits_{\xi_0}^{\xi_0+w_0} \bar{C}(\eta)\mathrm{d}\eta\Big)^{m+1}$$
$$\Big(\prod_{j=0}^{|k|}\Big(\prod_{r=i_j+1}^{i_{j+1}-1}\big(\bar{C}(\xi_r)-\bar{C}(\xi_r+w_j)\big)\Big)\bar{C}(\xi_{i_{j+1}}+w_j)\Big)$$
$$\mathrm{d}\xi_0\ldots\mathrm{d}\xi_h\mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1}. \tag{3.21}$$

From (3.19) and (3.21) it follows

$$I_m(k^{(0)};u) + I_m(k^{(1)};u) = \mu_*^{|k|} \int\limits_{0\le w_0\le\ldots\le w_{|k|}} \int\limits_{\xi_0\ge\ldots\ge\xi_h\ge 0}$$
$$\frac{\mathrm{d}}{\mathrm{d}\xi_0}\Big(-e^{-\mu_*\xi_0}\frac{1}{(m+1)!}\Big(\int\limits_{\xi_0}^{\xi_0+w_0} \bar{C}(\eta)\mathrm{d}\eta\Big)^{m+1}\Big)$$
$$\Big(\prod_{j=0}^{|k|}\Big(\prod_{r=i_j+1}^{i_{j+1}-1}\big(\bar{C}(\xi_r)-\bar{C}(\xi_r+w_j)\big)\Big)\bar{C}(\xi_{i_{j+1}}+w_j)\Big)$$
$$\mathrm{d}\xi_0\ldots\mathrm{d}\xi_h\mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1}. \tag{3.22}$$

In view of (3.12), integration with respect to ξ_0 provides (3.16). Note, in case of $|k| = 0$ the integral with respect to $\mathrm{d}w_0\ldots\mathrm{d}w_{|k|-1}$ does not occur in (3.19), (3.21) and (3.22).

Further, from (3.14) we obtain

$$I_m(0;u) = \int\limits_{\xi_1\ge 0} e^{-\mu_*\xi_1}\frac{1}{m!}\Bigg(\int\limits_{\xi_1}^{\xi_1+u}\bar{C}(\eta)\mathrm{d}\eta\Bigg)^m\Big(\bar{C}(\xi_1)-\bar{C}(\xi_1+u)\Big)\mathrm{d}\xi_1 \tag{3.23}$$

and from (3.12) using Fubini's theorem

$$\begin{aligned} I_m(1;u) &= \mu_* \int\limits_{0\le w_0\le u}\int\limits_{\xi_1\ge 0} e^{-\mu_*\xi_1}\frac{1}{m!}\Bigg(\int\limits_{\xi_1}^{\xi_1+w_0}\bar{C}(\eta)\mathrm{d}\eta\Bigg)^m\bar{C}(\xi_1+w_0)\mathrm{d}\xi_1\mathrm{d}w_0 \\ &= \int\limits_{\xi_1\ge 0}\mu_* e^{-\mu_*\xi_1}\frac{1}{(m+1)!}\Bigg(\int\limits_{\xi_1}^{\xi_1+u}\bar{C}(\eta)\mathrm{d}\eta\Bigg)^{m+1}\mathrm{d}\xi_1. \end{aligned} \tag{3.24}$$

In view of (3.13), the representation of $I_m(0;u)+I_m(1;u)$ given by (3.23) and (3.24) yields (3.17) for the case of $h=1$ by integration.
For $m\in\mathbb{Z}_+$, $h\in\mathbb{Z}_+\setminus\{0\}$ and $u\in\mathbb{R}_+$ by means of (3.16) we conclude

$$\sum_{k\in\{0,1\}^{h+1}} I_m(k;u) = \sum_{k\in\{0,1\}^h}\Big(I_m(0,k;u)+I_m(1,k;u)\Big) = \sum_{k\in\{0,1\}^h} I_{m+1}(k;u). \tag{3.25}$$

Applying (3.25) and (3.17) for $h=1$ provides (3.17) for arbitrary $h\in\mathbb{Z}_+\setminus\{0\}$.

□

3.5. Now we are in the position to prove the explicit formula (3.2) for the density $v(s+\ell;u)$. From (3.10), (3.15) we obtain

$$v(s+\ell;u) = \sum_{k\in\{0,1\}^\ell} p(k;u) = g_{s+\ell}\,\mu_* e^{-\mu_* u}\sum_{k\in\{0,1\}^\ell} I_0(k;u). \tag{3.26}$$

Thus from Lemma 3.2 Eq. (3.17), (3.13) and (2.5) we conclude

$$v(s+\ell;u) = g_{s+\ell}\,\mu_* e^{-\mu_* u}\frac{1}{\ell!}\Bigg(\int\limits_0^u\bar{C}(\eta)\mathrm{d}\eta\Bigg)^\ell = g_{s+\ell}\,\mu_* e^{-\mu_* u}\,\frac{F(\mu_* u)^\ell}{\ell!}, \tag{3.27}$$

which is just (3.2).

4. The case of a general distribution $C(u)$ of the maximal waiting times is obtained by considering $C(u)$ as the limit in distribution of a sequence of non-defective distributions $C_\nu(u)$ with continuous density. From (3.2) applied to $C_\nu(u)$ by arguments of continuity we obtain (3.2) for general distributions $C(u)$.

□

4 Asymptotic results for the departure rates due to impatience

For justifying the Markovian system approximation and its interpretations given in the next section we prove a couple of new results for the impatience rates α_ℓ, cf. (2.11).

Lemma 4.1 It holds
(i) $\alpha_\ell > 0$, $\ell = 1, 2, \ldots$,
(ii) $(\alpha_\ell + \mu_*)/\ell > 1/EI$, $\ell = 1, 2, \ldots$,
(iii) α_ℓ increases strictly monotonically in ℓ,
(iv) $(\alpha_\ell + \mu_*)/\ell$ decreases strictly monotonically in ℓ.

Proof. (i) From (2.5) we conclude $F'(\xi) = (1 - C(\xi/\mu_*))/\mu_* \le 1/\mu_*$ and $\lim_{\xi\to\infty} F'(\xi) < 1/\mu_*$. Thus the assertion follows from (2.11) and (2.3), (2.4).
(ii) Since $F(\xi)$ is an increasing function, from (2.11), (2.3), (2.5) we obtain

$$\alpha_\ell + \mu_* > \frac{\ell}{\lim\limits_{\xi\to\infty} F(\xi)} = \frac{\ell}{\int\limits_0^\infty (1 - C(\eta))\mathrm{d}\eta} = \frac{\ell}{EI}.$$

(iii) Using Fubini's theorem from (2.11), (2.3), (2.4) for $\ell = 2, 3, \ldots$ we find

$$\frac{\alpha_\ell + \mu_*}{\alpha_{\ell-1} + \mu_*} - 1 = \frac{1}{2} \frac{\int\limits_{\mathbb{R}_+^2} F(\xi)^{\ell-2} F(\eta)^{\ell-2} (F(\eta) - F(\xi))(F'(\xi) - F'(\eta)) e^{-(\xi+\eta)} \mathrm{d}\xi \mathrm{d}\eta}{\int\limits_0^\infty F(\xi)^{\ell-1} F'(\xi) e^{-\xi} \mathrm{d}\xi \int\limits_0^\infty F(\xi)^{\ell-2} e^{-\xi} \mathrm{d}\xi}.$$

Since $F(\xi)$ is non-negative and monotonically increasing and $F'(\xi)$ decreases monotonically, it follows that the integrand in the numerator is non-negative over $\mathbb{R}_+^2$. Since the integrand does not vanish everywhere, we have

$$\frac{\alpha_\ell + \mu_*}{\alpha_{\ell-1} + \mu_*} - 1 > 0, \quad \ell = 2, 3, \ldots,$$

and (iii) is proved.
(iv) Using Fubini's theorem from (2.11), (2.3) for $\ell = 2, 3, \ldots$ we obtain

$$\frac{\alpha_\ell + \mu_*}{\ell} - \frac{\alpha_{\ell-1} + \mu_*}{\ell - 1} = -\frac{1}{2} \frac{\int\limits_{\mathbb{R}_+^2} F(\xi)^{\ell-2} F(\eta)^{\ell-2} (F(\xi) - F(\eta))^2 e^{-(\xi+\eta)} \mathrm{d}\xi \mathrm{d}\eta}{\int\limits_0^\infty F(\xi)^{\ell} e^{-\xi} \mathrm{d}\xi \int\limits_0^\infty F(\xi)^{\ell-1} e^{-\xi} \mathrm{d}\xi}. \quad (4.1)$$

Thus the sequence $(\alpha_\ell + \mu_*)/\ell$, $\ell = 1, 2, \ldots$, decreases strictly monotonically.

□

Theorem 4.2 (i) It holds
$$\lim_{\ell\to\infty}\frac{\alpha_\ell}{\ell}=\frac{1}{EI}. \tag{4.2}$$
(ii) If there exists $m\in(1,\infty)$ such that the $m-$th moment of the distribution $C(u)$ of the maximal waiting times is finite, then moreover it holds
$$\frac{\alpha_\ell}{\ell}=\frac{1}{EI}+\mathcal{O}(\ell^{(1-m)/m}). \tag{4.3}$$
(iii) It holds
$$\limsup_{\ell\to\infty}(\alpha_\ell-\alpha_{\ell-1})=\frac{1}{EI}. \tag{4.4}$$
(iv) If there exists $m\in(2,\infty)$ such that the $m-$th moment of the distribution $C(u)$ is finite, then moreover it holds
$$\alpha_\ell-\alpha_{\ell-1}=\frac{1}{EI}+\mathcal{O}(\ell^{(2-m)/m}), \tag{4.5}$$
and especially we have
$$\lim_{\ell\to\infty}(\alpha_\ell-\alpha_{\ell-1})=\frac{1}{EI}. \tag{4.6}$$

Proof. (i) From Lemma 4.1 (ii) and (iv) for $\ell=1,2,\ldots$ it follows
$$\frac{1}{EI}<\frac{\alpha_\ell+\mu_*}{\ell}\le\left(\prod_{j=1}^{\ell}\frac{\alpha_j+\mu_*}{j}\right)^{1/\ell}.$$
Thus from (2.11), (2.3), (2.4) we obtain
$$\frac{1}{EI}<\frac{\alpha_\ell+\mu_*}{\ell}\le\left(\int_0^\infty F(\xi)^\ell e^{-\xi}\mathrm{d}\xi\right)^{-1/\ell}. \tag{4.7}$$
As $F(\xi)$ is non-negative and increases monotonically, for $x\in\mathbb{R}_+$ and $\ell=1,2,\ldots$ it holds
$$\int_0^\infty F(\xi)^\ell e^{-\xi}\mathrm{d}\xi\ge\int_x^\infty F(x)^\ell e^{-\xi}\mathrm{d}\xi=F(x)^\ell e^{-x}, \tag{4.8}$$
yielding
$$\frac{1}{EI}<\frac{\alpha_\ell+\mu_*}{\ell}\le\frac{e^{x/\ell}}{F(x)}.$$
Since $(\alpha_\ell+\mu_*)/\ell$ decreases monotonically in ℓ thus we conclude
$$\frac{1}{EI}\le\lim_{\ell\to\infty}\frac{\alpha_\ell}{\ell}\le\frac{1}{F(x)},$$

and because of $\lim\limits_{x\to\infty} F(x) = EI$ we obtain (4.2).
(ii) From (4.8), (4.7) for $x \in \mathbb{R}_+$ and $\ell = 1, 2, \dots$ it follows

$$F(x) \le EI \min(e^{\frac{x-x_\ell}{\ell}}, 1), \qquad x_\ell := \ln\left(\frac{(EI)^\ell}{\int\limits_0^\infty F(\xi)^\ell e^{-\xi} \mathrm{d}\xi} \right) > 0.$$

Therefore by twofold integration by parts in view of (2.5) we obtain

$$\frac{\mu_*^{m-1}}{m(m-1)} \int\limits_0^\infty u^m \mathrm{d}C(u) = \int\limits_0^\infty \xi^{m-2}(EI - F(\xi))\mathrm{d}\xi$$

$$\ge EI \int\limits_{x_\ell/3}^{2x_\ell/3} \xi^{m-2}(1 - e^{\frac{\xi - x_\ell}{\ell}})\mathrm{d}\xi \ge \frac{1}{2} EI \left(\frac{x_\ell}{3}\right)^{m-1} \left(1 - e^{-\frac{x_\ell}{3\ell}}\right)$$

$$\ge \frac{1}{2} EI\, \ell^{m-1} \left(1 - e^{-\frac{x_\ell}{3\ell}}\right)^m,$$

consequently it holds $1 - e^{-\frac{x_\ell}{3\ell}} \le \mathcal{O}(\ell^{(1-m)/m})$. Thus there exists $C_m \in (0, \infty)$ such that $x_\ell \le C_m \ell^{1/m}$, i.e., it holds

$$\int\limits_0^\infty F(\xi)^\ell e^{-\xi} \mathrm{d}\xi \ge (EI)^\ell \exp(-C_m \ell^{1/m}), \quad \ell = 1, 2, \dots \tag{4.9}$$

From (4.7) and (4.9) we find

$$\frac{1}{EI} < \frac{\alpha_\ell + \mu_*}{\ell} \le \frac{1}{EI} \exp(C_m \ell^{(1-m)/m}), \quad \ell = 1, 2, \dots,$$

yielding (4.3).
(iii) Let $\alpha_0 := 0$. In view of (4.2) we obtain

$$\limsup_{\ell\to\infty} (\alpha_\ell - \alpha_{\ell-1}) \ge \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{j=1}^{\ell} (\alpha_j - \alpha_{j-1}) = \lim_{\ell\to\infty} \frac{\alpha_\ell}{\ell} = \frac{1}{EI}.$$

Since the sequence $(\alpha_\ell + \mu_*)/\ell$ decreases monotonically, for $\ell = 2, 3, \dots$ we have

$$\alpha_\ell - \alpha_{\ell-1} = \frac{\alpha_\ell + \mu_*}{\ell} + (\ell - 1)\left(\frac{\alpha_\ell + \mu_*}{\ell} - \frac{\alpha_{\ell-1} + \mu_*}{\ell - 1} \right) \le \frac{\alpha_\ell + \mu_*}{\ell}. \tag{4.10}$$

Thus from (4.2) moreover it follows

$$\limsup_{\ell\to\infty} (\alpha_\ell - \alpha_{\ell-1}) \le \lim_{\ell\to\infty} \frac{\alpha_\ell + \mu_*}{\ell} = \frac{1}{EI}.$$

(iv) Let $M_\ell := \{\xi \in \mathbb{R}_+ : F(\xi) \le EI \exp(-C_m \ell^{(1-m)/m})\}$ for $\ell = 1, 2, \ldots$ As $f(x) := x^\ell (EI - x)^2$ increases monotonically for $x \in [0, (\ell/(\ell+2))EI]$, because of (4.9) for $\ell \ge (2/C_m)^m$ it follows

$$\frac{\int_0^\infty F(\xi)^\ell (EI - F(\xi))^2 e^{-\xi} \mathrm{d}\xi}{\int_0^\infty F(\xi)^\ell e^{-\xi} \mathrm{d}\xi}$$

$$\le \frac{\int_{M_\ell} (EI)^{\ell+2} \exp(-C_m \ell^{1/m})(1 - \exp(-C_m \ell^{(1-m)/m}))^2 e^{-\xi} \mathrm{d}\xi}{\int_0^\infty F(\xi)^\ell e^{-\xi} \mathrm{d}\xi}$$

$$+ \frac{\int_{\mathbb{R}_+ \setminus M_\ell} F(\xi)^\ell (EI)^2 (1 - \exp(-C_m \ell^{(1-m)/m}))^2 e^{-\xi} \mathrm{d}\xi}{\int_0^\infty F(\xi)^\ell e^{-\xi} \mathrm{d}\xi}$$

$$\le 2(EI)^2 (1 - \exp(-C_m \ell^{(1-m)/m}))^2 \le 2C_m^2 (EI)^2 \ell^{(2-2m)/m},$$

and we conclude that

$$(\ell - 1) \frac{\int_0^\infty F(\xi)^{\ell-2} (EI - F(\xi))^2 e^{-\xi} \mathrm{d}\xi}{\int_0^\infty F(\xi)^{\ell-2} e^{-\xi} \mathrm{d}\xi} = \mathcal{O}(\ell^{(2-m)/m}). \tag{4.11}$$

From (4.10), (4.1), (2.3) and (2.11) for $\ell = 2, 3, \ldots$ we obtain

$$\alpha_\ell - \alpha_{\ell-1} = \frac{\alpha_\ell + \mu_*}{\ell} \left(1 - \left(\frac{\alpha_{\ell-1} + \mu_*}{\ell - 1} \right)^2 \right.$$

$$\left. \frac{\ell - 1}{2} \frac{\int_{\mathbb{R}_+^2} F(\xi)^{\ell-2} F(\eta)^{\ell-2} (F(\xi) - F(\eta))^2 e^{-(\xi+\eta)} \mathrm{d}\xi \mathrm{d}\eta}{\left(\int_0^\infty F(\xi)^{\ell-2} e^{-\xi} \mathrm{d}\xi \right)^2} \right).$$

Because of $(F(\xi) - F(\eta))^2 \le (EI - F(\xi))^2 + (EI - F(\eta))^2$ we find

$$\frac{\alpha_\ell + \mu_*}{\ell} \left(1 - \left(\frac{\alpha_{\ell-1} + \mu_*}{\ell - 1} \right)^2 (\ell - 1) \frac{\int_0^\infty F(\xi)^{\ell-2} (EI - F(\xi))^2 e^{-\xi} \mathrm{d}\xi}{\int_0^\infty F(\xi)^{\ell-2} e^{-\xi} \mathrm{d}\xi} \right)$$

$$\le \alpha_\ell - \alpha_{\ell-1} \le \frac{\alpha_\ell + \mu_*}{\ell}.$$

Thus (4.3) and (4.11) yield (4.5).

□

5 Approximation of the $M(n)/M(n)/s+GI$ system by a $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system

5.1 The $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system

Consider the $M(n)/M(n)/s+GI$ system under the same assumptions as in Section 1 but with the following modified impatience mechanism, cf. Figure 2: With each waiting place, which are numbered by $i = 1, 2, \ldots$, there is associated an impatience rate $\beta_i\,(\geq 0)$, $i = 1, 2, \ldots$ A call waiting on place i leaves the queue and the system due to impatience with rate β_i. The calls behind it move up in the queue according to the FCFS discipline. We denote this system by $M(n)/M(n)/s + M(\beta_i)_{i=1}^{\infty}$, where $M(\beta_i)_{i=1}^{\infty}$ stands for the waiting place dependent impatience mechanism. The cumulative impatience rate if there are ℓ calls in the queue is

$$\alpha_\ell := \sum_{i=1}^{\ell} \beta_i, \quad \ell = 1, 2, \ldots \tag{5.1}$$

If $\beta_i \equiv \beta$, $i = 1, 2, \ldots$, then clearly the system corresponds to those of a $M(n)/M(n)/s + M$ system. If $\lambda_n > 0$ for $0 \leq n < s + k$ and $\lambda_n \equiv 0$ for $n \geq s + k$, then we have the case of a limited waiting room with k waiting places ($M(n)/M(n)/s/k+M(\beta_i)_{i=1}^{\infty}$ system), and the system dynamics are independent of the definition of the β_i for $i > k$.

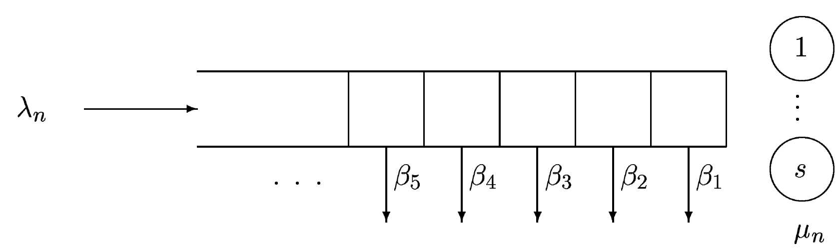

Figure 2: *The $M(n)/M(n)/s + M(\beta_i)_{i=1}^{\infty}$ system with waiting place dependent impatience rates β_i.*

Remark 5.1 Since a call finds at its arrival a random number of calls in the queue the impatience time of a call depends on the queueing process, i.e., in contrast to the $M(n)/M(n)/s+GI$ model there is a dynamic impatience of the calls.

The number $N(t)$ of calls in the $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system is a birth-death process with birth rates λ_n and death rates

$$\Phi_n = \begin{cases} \mu_n, & n=0,1,\ldots,s, \\ \mu_* + \alpha_{n-s}, & n=s+1,s+2,\ldots \end{cases} \tag{5.2}$$

The stationary distribution $p(n) := P(N(t)=n)$, $n \in \mathbb{Z}_+$, is given by

$$p(n) = \begin{cases} g_n \prod_{i=n+1}^{s} \mu_i\,, & n=0,1,\ldots,s, \\ g_n \left(\prod_{i=s+1}^{n} \Phi_i \right)^{-1}, & n=s+1,s+2,\ldots, \end{cases} \tag{5.3}$$

where g_n is given by (2.7) and g by

$$g^{-1} = \sum_{j=0}^{s-1} \left(\prod_{i=0}^{j-1} \lambda_i \right) \left(\prod_{i=j+1}^{s} \mu_i \right) + \sum_{j=0}^{\infty} \left(\prod_{i=0}^{s+j-1} \lambda_i \right) \left(\prod_{i=s+1}^{s+j} \Phi_i \right)^{-1}. \tag{5.4}$$

Note, that (5.3) and (5.4) have the same structure as (2.9) and (2.2), respectively. Since the λ_n are bounded, the system is stable iff

$$\sum_{j=0}^{\infty} \left(\prod_{i=s}^{s+j-1} \lambda_i \right) \left(\prod_{i=s+1}^{s+j} \Phi_i \right)^{-1} < \infty. \tag{5.5}$$

In the following we assume that the system is stable and in steady state. Then the cumulative mean call arrival intensity Λ, the abandon probability p_I, the probability p_W that a typical arriving call has to wait for service and the mean waiting time EW in the queue of a typical arriving call are given by (2.12)-(2.14) again. In case of a $M/M(n)/s/k+M(\beta_i)_{i=1}^{\infty}$ system the blocking probability p_B is given by (2.15) again. One can derive explicit formulae for the distribution of the workload in the system, the virtual waiting time and the waiting times of calls being served and/or getting lost due to impatience. They are mixtures of convolutions of exponential distributions.

Remark 5.2 For a $M(n)/M(n)/s$ FCFS queueing system with arbitrary cumulative Markovian impatience rates $\alpha_\ell\,(\geq 0)$ if there are $\ell\,(\geq 1)$ calls waiting in the queue, the quantities $p(n)$, Λ, p_I, p_W, EW and p_B are given by the same formulae as for the $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system. In such a model the impatience cannot be connected with waiting place dependent impatience rates, in general.

However, if the sequence $(\alpha_\ell)_{\ell=1}^\infty$ increases monotonically than the process $N(t)$ of the number of calls corresponds to those of a $M(n)/M(n)/s + M(\beta_i)_{i=1}^\infty$ system with waiting place dependent impatience rates $\beta_i := \alpha_i - \mathbb{1}\{i>1\}\alpha_{i-1}$ (≥ 0), $i \geq 1$, i.e., a transformation into a $M(n)/M(n)/s + M(\beta_i)_{i=1}^\infty$ system is possible.

5.2 Approximation by fitting the occupancy distribution

Since we deal in the following with the $M(n)/M(n)/s + GI$ system and the $M(n)/M(n)/s + M(\beta_i)_{i=1}^\infty$ system simultaneously, we endow the performance quantities of the $M(n)/M(n)/s + GI$ system with the index C, i.e. $p_C(n)$, $p_{I,C}$, $\alpha_{\ell,C}$, and those of the $M(n)/M(n)/s + M(\beta_i)_{i=1}^\infty$ system with β, i.e. $p_\beta(n)$, $p_{I,\beta}$, $\alpha_{\ell,\beta}$.

Let us consider a stable $M(n)/M(n)/s+GI$ system in steady state. A reasonable Markovian approximation for the given $M(n)/M(n)/s + GI$ system would be a $M(n)/M(n)/s+M(\beta_i)_{i=1}^\infty$ system with the same arrival- and service process and such impatience rates β_i, $i = 1, 2, \ldots$, that the occupancy distribution is fitted:

$$p_\beta(n) = p_C(n), \quad n \in \mathbb{Z}_+. \tag{5.6}$$

Because of (2.2), (2.7), (2.9) and (5.3), (5.4) the fitting (5.6) is equivalent to

$$\left(\prod_{i=s+1}^{s+j} \Phi_i \right)^{-1} = F_j, \quad j = 1, 2, \ldots, k,$$

where again $k := \sup\{j \in \mathbb{Z}_+ : \lambda_{j+s-1} > 0\}$ as in Section 2. In view of (5.2), (2.4) these identities are equivalent to

$$\alpha_{\ell,\beta} = \frac{F_{\ell-1}}{F_\ell} - \mu_*, \quad \ell = 1, 2, \ldots, k. \tag{5.7}$$

In case of a limited waiting room ($k < \infty$) let $\alpha_{\ell,\beta}$ be defined by (5.7) for $\ell > k$, too. Comparing (5.7) and (2.11) we see that the cumulative impatience rates are fitted, too, i.e. $\alpha_{\ell,\beta} = \alpha_{\ell,C}$, $\ell = 1, 2, \ldots$ From Lemma 4.1 we know that the $\alpha_{\ell,C}$ are strictly increasing and positive. Now, taking into account (5.1) it follows that the rates

$$\beta_i := \alpha_{i,\beta} - \mathbb{1}\{i>1\}\alpha_{i-1,\beta} = \alpha_{i,C} - \mathbb{1}\{i>1\}\alpha_{i-1,C}, \quad i = 1, 2, \ldots, \tag{5.8}$$

are positive and that the corresponding $M(n)/M(n)/s + M(\beta_i)_{i=1}^\infty$ system meets the occupancy distribution and the cumulative impatience rates of the given $M(n)/M(n)/s + GI$ system. Altogether, for a given $M(n)/M(n)/s + GI$ system (5.7) and (5.8) provide an algorithm for constructing a $M(n)/M(n)/s+M(\beta_i)_{i=1}^\infty$ system with the same occupancy distribution and cumulative impatience rates as well as the same arrival- and service process. The system dynamics of any $M(n)/M(n)/s + M(\beta_i)_{i=1}^{\infty}$ system satisfying the fitting (5.6) is uniquely determined as the β_i are uniquely determined for $i = 1, 2, \dots, k$.

From (2.11)-(2.15) and (5.6), (5.7) it follows that the fitting of the occupancy distribution implies a fitting of the cumulative mean call arrival intensity, of the abandon probability, of the probability that a typical arriving call has to wait for service, of the mean waiting time and of the blocking probability in case of a $M/M(n)/s/k + GI$ system, i.e.

$$\Lambda_\beta = \Lambda_C, \ \ p_{I,\beta} = p_{I,C}, \ \ p_{W,\beta} = p_{W,C}, \ \ EW_\beta = EW_C, \ \ p_{B,\beta} = p_{B,C}.$$

Although several quantities will be fitted, this clearly is not true for the different waiting time distributions. The considerations given above show that a fitting of the cumulative impatience rates $\alpha_{\ell,\beta} = \alpha_{\ell,C}$, $\ell = 1, 2, \dots$, yields the same fitting. The latter approach has been used in [BB2] in the framework of a two-queue priority system and its application to a performance analysis of an inbound call center with an integrated voice-mail-server. However, the monotonicity and asymptotic results for the impatience rates of the $M(n)/M(n)/s + GI$ system given in Section 4 provide more insight into the Markovian approximation. In particular, from Theorem 4.2 it follows that the β_i defined by (5.7), (5.8) converge to the intensity $1/EI$ corresponding to an exponential fitting if there exists $m \in (2, \infty)$ such that the m-th moment of the distribution $C(u)$ of the maximal waiting times I is finite, and moreover, that

$$\beta_i = 1/EI + \mathcal{O}(i^{(2-m)/m}). \qquad (5.9)$$

In a straight forward manner the results of Section 4 can also be applied in the framework of [BB2].

5.3 The fitted waiting place dependent impatience rates in case of constant impatience times

As an example we determine the fitted waiting place dependent impatience rates in case of approximating the $M(n)/M(n)/s + D$ system, i.e. in case of constant impatience times $I \equiv \tau$. From (2.5) it follows $F(\xi) = \min(\xi/\mu_*, \tau)$, and thus (2.3) provides

$$F_\ell = e^{-\mu_* \tau} \tau^\ell \sum_{j=0}^{\infty} \frac{1}{(\ell + j)!} (\mu_* \tau)^j, \quad \ell = 0, 1, \dots$$

Hence the fitting (5.7) yields

$$\alpha_{\ell,\beta} = \frac{\ell}{\tau} \left(\sum_{j=0}^{\infty} \frac{\ell!}{(\ell + j)!} (\mu_* \tau)^j \right)^{-1}, \quad \ell = 1, 2, \dots,$$

and from (5.8) thus it follows $\beta_1 = \alpha_{1,\beta} = \mu_*/(e^{\mu_*\tau} - 1)$ and in general the numerically stable representation

$$\beta_i = \frac{1}{\tau} \frac{\sum_{j=0}^{\infty} \frac{i!\,(j+1)}{(i+j)!} (\mu_*\tau)^j}{\left(\sum_{j=0}^{\infty} \frac{i!}{(i+j)!} (\mu_*\tau)^j\right)\left(\sum_{j=0}^{\infty} \frac{(i-1)!}{(i+j-1)!} (\mu_*\tau)^j\right)}, \quad i = 1, 2, \ldots \tag{5.10}$$

Table 1: *Fitted waiting place dependent impatience rates β_i in case of constant impatience times $I \equiv \tau$ and $\mu_* = 1$.*

i	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 8$	$\tau = 16$
1	1.541494	0.581977	0.156518	0.018657	0.000336	0.000000
2	1.820500	0.810234	0.299161	0.061991	0.002356	0.000002
3	1.907538	0.898406	0.381472	0.111668	0.008193	0.000013
4	1.944382	0.938409	0.425807	0.152532	0.019009	0.000062
5	1.963096	0.959213	0.450591	0.181516	0.033695	0.000231
6	1.973811	0.971204	0.465288	0.201024	0.049676	0.000677
7	1.980489	0.978672	0.474519	0.214077	0.064618	0.001648
8	1.984918	0.983610	0.480615	0.222943	0.077286	0.003421
9	1.988002	0.987032	0.484816	0.229104	0.087416	0.006203
10	1.990233	0.989496	0.487817	0.233493	0.095270	0.010019
20	1.997516	0.997407	0.497168	0.246595	0.119919	0.050135
30	1.998892	0.998858	0.498786	0.248620	0.123182	0.059059
40	1.999376	0.999361	0.499330	0.249262	0.124092	0.061045
50	1.999600	0.999593	0.499577	0.249542	0.124460	0.061719
60	1.999722	0.999718	0.499709	0.249689	0.124643	0.062017
70	1.999796	0.999793	0.499787	0.249775	0.124747	0.062173
80	1.999844	0.999842	0.499838	0.249830	0.124811	0.062265
90	1.999877	0.999875	0.499873	0.249867	0.124854	0.062323
100	1.999900	0.999899	0.499897	0.249893	0.124884	0.062362

The representation (5.10) provides that in case of constant impatience times it holds

$$\beta_i < 1/EI, \quad i = 1, 2, \ldots, \tag{5.11}$$

and we obtain the asymptotic estimate

$$\beta_i = 1/EI - \mu_* i^{-2} + \mathcal{O}(i^{-3}), \tag{5.12}$$

cf. (5.9).

In Table 1 there are given the first waiting place dependent impatience rates β_i of the $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system approximating the $M(n)/M(n)/s+D$ system. By choosing $1/\mu_*$ as unit of time without loss of generality we may assume $\mu_* = 1$ in Table 1. The results of Table 1 show that the approximation of a $M(n)/M(n)/s+GI$ system by a $M(n)/M(n)/s+M(\beta_i)_{i=1}^{\infty}$ system, as proposed in Section 5.2, may lead to impatience rates β_i which for smaller i, in particular in case of $1/EI \ll \mu_*$, significantly differ from $1/EI$. Note, that in case of exponentially distributed impatience times ($GI = M$) the approximation is trivially $\beta_i = 1/EI$, $i = 1, 2, \ldots$, i.e., the approximation provides just the original $M(n)/M(n)/s+M$ system.

References

[BBH] Baccelli, F., Boyer, P., Hebuterne, G., Single server queues with impatient customers. Adv. Appl. Prob. 16 (1984) 887–905.

[BH] Baccelli, F., Hebuterne, G., On queues with impatient customers. in: F.J. Kylstra (Ed.), *Performance '81.* North-Holland Publishing Company, 1981, 159–179.

[Ba1] Barrer, D.Y., Queueing with impatient customers and indifferent clerks. Operations Research 5 (1957) 644–649.

[Ba2] Barrer, D.Y., Queueing with impatient customers and ordered service. Operations Research 5 (1957) 650–656.

[BB1] Brandt, A., Brandt, M., On the $M(n)/M(n)/s$ queue with impatient calls. Performance Evaluation 35 (1999) 1–18. (Preprint SC 97-19, Konrad-Zuse-Zentrum für Informationstechnik Berlin, available by http://www.zib.de/PaperWeb/abstracts/SC-97-19).

[BB2] Brandt, A., Brandt, M., On a two-queue priority system with impatience and its application to a call center. Methodology and Computing in Applied Probability 1 (1999) 191–210. (Preprint SC 98-21, Konrad-Zuse-Zentrum für Informationstechnik Berlin, available by http://www.zib.de/PaperWeb/abstracts/SC-98-21).

[Br1] Brodi, S.M., On an integro-differential equation for systems with τ-waiting (in Ukrainian). Dop. AN URSR No. 6 (1959) 571–573.

[Br2] Brodi, S.M., On a service problem (in Russian). Tr. V. Vsesojuznogo soveščanija po teorii verojatnostej i matematičeskoj statistike, Izd-vo AN Arm. SSR, Erevan (1960) 143–147.

[Dal] Daley, D.J., General customer impatience in the queue $GI/G/1$. J. Appl. Prob. 2 (1965) 186–205.

[FKAS] Franken, P., König, D., Arndt, U., Schmidt, V., *Queues and Point Processes.* Akademie-Verlag, Berlin; Wiley, Chichester, 1982.

[GK] Gnedenko, B.W., Kowalenko, I.N., *Einführung in die Bedienungstheorie.* Akademie-Verlag, Berlin, 1974 (first edition in Russian, Nauka, 1966).

[GKo] Gnedenko, B.W., König, D., *Handbuch der Bedienungstheorie Vol. II.* Akademie-Verlag, Berlin, 1984.

[Hau] Haugen, R.B., Queueing systems with several input streams and time out. Telektronikk nr. 2, (1978) 100–106.

[HS] Haugen, R.B., Skogan, E., Queueing systems with stochastic time out. IEEE Trans. Commun. vol. COM-28 (1980) 1984–1989.

[Ju1] Jurkevič, O.M., On the investigation of many-server queueing systems with bounded waiting time (in Russian). Izv. Akad. Nauk SSSR Techničeskaja kibernetika No. 5 (1970) 50–58.

[Ju2] Jurkevič, O.M., On many-server systems with stochastic bounds for the waiting time (in Russian). Izv. Akad. Nauk SSSR Techničeskaja kibernetika No. 4 (1971) 39–46.

[Ju3] Jurkevič, O.M., On many-server systems with general service time distribution of "impatient customers" (in Russian). Izv. Akad. Nauk SSSR Techničeskaja kibernetika No. 6 (1973) 80–89.

[Lav] Lavenberg, S.S. (Ed.), *Computer Performance Modelling Handbook.* Academic Press, New York, 1983.

[MS] Mandelbaum, A., Shimkin, N., A model for rational abandonments from invisible queues, Unpublished paper, Technion, Haifa, 1998.

[Mov] Movaghar, A., On queueing with customer impatience until the beginning of service. Queueing Systems 29 (1998) 337–350.

[Sin] Singh, M., Steady state behaviour of serial queueing processes with impatient customers. Math. Operationsforsch. Ser. Statist. 15 (2) (1984) 289–298.

[Teg] Teghem, J., Use of discrete transforms for the study of a GI/M/s queue with impatient customer phenomena. Zeitschrift für Operations Research 23 (1979) 95–106.

[Wal] Wallstrøm, B., A queueing system with time outs and random departures. Proc. ITC 8 Melbourne, Australia, 1976, paper 231.