Department of Applied Mathematics and Physics, Kyoto University
Technical Report #2000-004 (August 8, 2000)

Subexponential asymptotics of the waiting time distribution in the stationary single-server queue

Tetsuya Takine

Department of Applied Mathematics and Physics
Graduate School of Informatics, Kyoto University
Kyoto 606-8501, Japan
E-mail: takine@amp.i.kyoto-u.ac.jp

ABSTRACT

This paper considers the subexponential asymptotics of the tail distributions of virtual waiting times, actual waiting times and sojourn times in stationary work-conserving single-server queues. To this end, a single-server queue with a Markovian arrival stream (MAS) is first examined, where arrivals are modulated by the underlying Markov chain with finite states and service time distributions may depend on the states of the underlying Markov chain immediately before and after arrivals. Under the assumption that the equilibrium distribution of the overall (i.e., customer-average) service time distribution is subexponential, a subexponential asymptotic formula is shown for the virtual waiting time distribution. Further when customers are served on a FIFO basis, the actual waiting time and the sojourn time distributions are shown to have the same asymptotics as the virtual waiting time distribution. As you will see, a common asymptotic constant of these three distributions is given only in terms of the utilization factor. Therefore the same asymptotic formulas hold in the FIFO G/G/1 queue as well, because MAS is dense in stationary marked point processes.

1 INTRODUCTION

This paper studies subexponential asymptotics of the virtual waiting time, actual waiting time and sojourn time distributions in the stationary work-conserving FIFO single-server queue. In the past, this problem was attacked by considering the ladder-height distribu-

tion representing the actual waiting time process and its Weiner-Hopf factorization. The subexponential asymptotics of the actual waiting time in the GI/GI/1 queue was obtained by Pakes [12] under the assumption that the equilibrium distribution of service times is subexponential. Owing to the properties of subexponential distributions, this result leads to the same asymptotics for the sojourn time distribution [15]. Also, using the invariance relationship between virtual and actual waiting times [14], one can readily show that the virtual waiting time distribution has the same asymptotics, too. Asmussen et al. generalized Pakes' result to the MMPP/G/1 queue with state-dependent services [2], and further Jelenković and Lazar obtained the subexponential asymptotics in the Markov-modulated G/G/1 queue [7]. These two papers have a common feature, i.e., they study the time-reversed process of ladder-heights which is closely related to the actual waiting time, and due to this, both results in [2] and [7] were obtained under the assumption that the service time distribution is heavy-tailed, together with the subexponentiality of its equilibrium distribution. Even though the assumption imposed in [2] and [7] is more restrictive than in [12], Jelenković and Lazar pointed out that the actual waiting time in the Markov modulated G/G/1 queue has the same asymptotics as in the GI/GI/1 queue [7].

The purpose of this paper is to show that under the assumption of the subexponential equilibrium distribution of service times, the waiting time distribution in the G/G/1 queue indeed has the same subexponential asymptotics as in the GI/GI/1 queue. To this end, we first studies the subexponential asymptotics of the virtual waiting time distribution in the stationary single-server queue with MAS (Markovian arrival stream) input [16, 20], where customer arrivals are modulated by the underlying Markov chain with finite states and service time distributions may depend on the states of the underlying Markov chain immediately before and after arrivals. Our starting point is a closed-form formula for the virtual waiting time distribution, which was recently found in [20]. As you will see, under the assumption that the equilibrium distribution of customer-average service time distribution is subexponential, all the virtual waiting time, actual waiting time and sojourn time distributions have exactly the same asymptotics as in the GI/GI/1 queue, where the asymptotic constant is given only in terms of the utilization factor.

The dependency between arrivals and services in MAS naturally arises in queues with the superposition of independent MAPs (Markovian arrival processes) [10] and phase-type renewal arrival streams [11], each of which has i.i.d. service times [16, 17, 19]. Therefore the results for the queue with MAS input are of independent interest. However, it is more important that MAS is dense in stationary marked point processes (MPPs) [1]. Thus under the assumption that the equilibrium distribution of service times is subexponential,

the waiting time distribution in the work-conserving single-server queue with MPP arrivals (i.e., the G/G/1 queue) has the same subexponential asymptotics as in the single-server queue with MAS input. Note here that the assumption of the subexponential equilibrium distribution of service times *does not* imply that the service time distribution itself is heavy-tailed (see Remark 6.2 in [15]). Thus our result is more general than those in [2, 7] in this regard, too.

The rest of the paper is divided into four sections. Section 2 summarizes the definition and some properties of the subexponential distribution. Section 3 describes the single-server queue with MAS input and some known results. Section 4 provides subexponential asymptotic formulas. The proof of the main theorem is given in section 5.

2 Subexponential distribution

For any distribution function $F(x)$ with finite mean f, let $\overline{F}(x)$ denote the complementary distribution function of $F(x)$, i.e., $\overline{F}(x) = 1 - F(x)$, and $F_e(x)$ denote the equilibrium distribution of $F(x)$:

$$F_e(x) = f^{-1} \int_0^x \overline{F}(y)dy.$$

A distribution function $F(x)$ (and the corresponding random variable F) is called heavy-tailed if $\overline{F}(x) > 0$ for all $x \geq 0$ and

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1, \qquad \text{for all } y \geq 0. \tag{1}$$

It is known that the convergence in (1) is uniform on compact y sets [13]. Let $F^{(n)}(x)$ denote the n-fold convolution of $F(x)$ with itself, i.e.,

$$F^{(n)}(x) = \int_0^x F(x-y)dF^{(n-1)}(y), \qquad n = 2, 3, \ldots,$$

with $F^{(1)}(x) = F(x)$, and $\overline{F^{(n)}}(x)$ denote $1 - F^{(n)}(x)$. Then the heavy-tailed $F(x)$ (and the corresponding random variable F) is called subexponential if

$$\lim_{x\to\infty} \frac{\overline{F^{(2)}}(x)}{\overline{F}(x)} = 2.$$

Readers are referred to [6, 7, 15] for details. In what follows, we provide some lemmas that we need in the sequel.

Lemma 1 (Corollary 3.3 in [15]) *For any heavy-tailed equilibrium distribution function $F_e(x)$,*

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{F_e}(x)} = 0.$$

Lemma 2 (Theorem 2.7 [4]) *Suppose $F(x)$ is subexponential and $G(x)$ is any distribution function such that*

$$\lim_{x\to\infty} \frac{\overline{G}(x)}{\overline{F}(x)} = c < \infty.$$

If $c > 0$, $G(x)$ is subexponential.

Lemma 3 (Lemma 10 in [7]) *Let F denote a subexponential random variable and let G_i's $(i = 1, \ldots, k)$ be independent random variables. Suppose*

$$\lim_{x\to\infty} \frac{\Pr(G_i > x)}{\Pr(F > x)} = g_i \in [0, \infty), \qquad i = 1, \ldots, k.$$

(a). The following equality holds.

$$\lim_{x\to\infty} \frac{\Pr(G_1 + \cdots + G_k > x)}{\Pr(F > x)} = \sum_{i=1}^{k} g_i. \tag{2}$$

(b) Let g^* denote $\max(g_1, \ldots, g_k)$. Then for each $\epsilon > 0$, there exists some constant $K = K(\epsilon, g^*) < \infty$ such that

$$\frac{\Pr(G_1 + \cdots + G_k > x)}{\Pr(F > x)} \le K(1+\epsilon)^k, \qquad \textit{for all } x \ge 0,$$

where K is independent of k.

3 Single-server queue with MAS input

We now consider a stationary work-conserving single-server queue with the following characteristics. Customer arrivals follow a stationary Markovian arrival process (MAP) with representation $(\boldsymbol{C}, \boldsymbol{D})$, where $\boldsymbol{C}$ and $\boldsymbol{D}$ are $M \times M$ matrices [10]. *MAP* is a class of semi-Markovian arrival processes, and the infinitesimal generator of the underlying continuous-time Markov chain which governs customer arrivals is given by $\boldsymbol{C} + \boldsymbol{D}$, where $\boldsymbol{C}$ has negative diagonal elements and nonnegative off-diagonal elements, and all elements in $\boldsymbol{D}$ are non-negative. Note that

$$(\boldsymbol{C} + \boldsymbol{D})\boldsymbol{e} = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{e}$ denotes an $M \times 1$ vector whose elements are all equal to one. We assume that the underlying Markov chain represented by the infinitesimal generator $(\boldsymbol{C}+\boldsymbol{D})$ is irreducible. We denote the set of states of the underlying Markov chain by $\mathcal{M} = \{1, \ldots, M\}$. When a state transition driven by $\boldsymbol{D}$ occurs, a customer arrives to the system. We assume that $D_{i,j} > 0$ at least for some $i, j \in \mathcal{M}$, so that arrivals are certain. Readers are referred to [10] for details.

We assume that service times of customers who arrive to the system with a transition from state $i \in \mathcal{M}$ to state $j \in \mathcal{M}$ (whose rate is given by the (i,j)th element $D_{i,j} > 0$ of $\boldsymbol{D}$) are i.i.d. (independent and identically distributed) according to a distribution function $H_{i,j}(x)$. Thus, service times of customers may depend on the states immediately before and after arrivals, and MAP with state-dependent services described above is called Markovian arrival stream (MAS) [1].

Without loss of generality, we assume that $H_{i,j}(0) = 0$ for $i,j \in \mathcal{M}$ such that $D_{i,j} > 0$. For convenience, we further assume that for $i,j \in \mathcal{M}$ such that $D_{i,j} = 0$, $H_{i,j}(x) = 1$ for all $x \geq 0$. Let $\boldsymbol{D}(x)$ denote an $M \times M$ matrix whose (i,j)th element $D_{i,j}(x)$ is given by $D_{i,j}H_{i,j}(x)$. Note that MAS can be represented by $(\boldsymbol{C}, \boldsymbol{D}(x))$. We denote the stationary probability vector of the underlying Markov chain by $\boldsymbol{\pi}$ which satisfies

$$\boldsymbol{\pi}(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{0}, \qquad \boldsymbol{\pi}\boldsymbol{e} = 1. \tag{4}$$

We denote the jth element of $\boldsymbol{\pi}$ by π_j. Let ρ denote the utilization factor, i.e.,

$$\rho = \boldsymbol{\pi} \int_0^\infty x d\boldsymbol{D}(x)\boldsymbol{e}, \tag{5}$$

which is assumed to be less than one, so that the queue is stable [9, 16].

We define V and S as generic random variables representing the virtual waiting time (i.e., the amount of stationary work in system) and the state of the stationary underlying Markov chain, respectively. With those, let $\boldsymbol{v}(x)$ denote a $1 \times M$ vector whose jth element represent the stationary joint probability $\Pr(V \leq x, S = j)$. The vector probability distribution $\boldsymbol{v}(x)$ was studied in [16, 18], where the infinitesimal generator $\boldsymbol{Q}$ of a certain irreducible Markov chain with M states plays an important role. Note that $\boldsymbol{Q}$ satisfies

$$\boldsymbol{Q} = \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(x) \exp(\boldsymbol{Q}x), \tag{6}$$

and can be computed as the limit $\boldsymbol{Q}_\infty$ in the following recursion

$$\boldsymbol{Q}_{n+1} = \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(x) \exp(\boldsymbol{Q}_n x), \qquad n = 0, 1, \ldots, \tag{7}$$

with $\boldsymbol{Q}_0 = \boldsymbol{C}$ [16]. Let $\boldsymbol{\kappa}$ denote a probability vector which satisfies $\boldsymbol{\kappa}\boldsymbol{Q} = \boldsymbol{0}$.

Recently, the single-server queue with MAS input was revisited in [20]. We define $\boldsymbol{R}(x)$ as an $M \times M$ matrix which is given by

$$\boldsymbol{R}(x) = \int_0^x dw \int_w^\infty d\boldsymbol{D}(y) \exp(\boldsymbol{Q}(y - w)). \tag{8}$$

Then the following formula for $\boldsymbol{v}(x)$ was shown in [20].

$$\boldsymbol{v}(x) = (1 - \rho)\boldsymbol{\kappa} \sum_{n=0}^\infty \boldsymbol{R}^{(n)}(x), \tag{9}$$

where $\boldsymbol{R}^{(0)}(x) = \boldsymbol{I}$ and $\boldsymbol{R}^{(n)}(x)$ $(n = 1, 2, \ldots)$ denotes the n-fold convolution of $\boldsymbol{R}(x)$ with itself, i.e.,

$$\boldsymbol{R}^{(n)}(x) = \int_0^x \boldsymbol{R}(x-y)d\boldsymbol{R}^{(n-1)}(y).$$

We now assume that customers are served on a FIFO basis. For $i, j \in \mathcal{M}$ such that $D_{i,j} > 0$, let $W_{i,j}$ and $T_{i,j}$ denote generic random variables representing the actual waiting time and the sojourn time (i.e. the sum of the waiting time and the service time) of customers who arrive with state transition from state i to state j. We define $w_{i,j}(x)$ and $t_{i,j}(x)$ $(i, j \in \mathcal{M}$ such that $D_{i,j} > 0)$ as the distribution functions of $W_{i,j}$ and $T_{i,j}$, respectively. Applying the conditional PASTA [8], we have

$$w_{i,j}(x) = \frac{v_i(x)D_{i,j}}{\pi_i D_{i,j}}, \qquad t_{i,j}(x) = \int_0^x w_{i,j}(x-y)dH_{i,j}(y). \tag{10}$$

Let W and T denote generic random variables representing the overall actual waiting time and the overall sojourn time, respectively. We then define $\boldsymbol{w}(x)$ and $\boldsymbol{t}(x)$ as $1 \times M$ vectors whose jth elements represent $E[W \leq x, S_A^+ = j]$ and $E[T \leq x, S_A^+ = j]$, respectively, where S_A^+ denotes a generic random variable representing the state of the underlying Markov chain immediately after arrivals. It then follows from (10) that

$$\boldsymbol{w}(x) = \frac{\boldsymbol{v}(x)\boldsymbol{D}}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}, \qquad \boldsymbol{t}(x) = \frac{1}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}} \int_0^x \boldsymbol{v}(x-y)d\boldsymbol{D}(y). \tag{11}$$

In what follows, we summarize some known properties of $\boldsymbol{R}(x)$.

Proposition 1 ([20])

(a). We denote $\boldsymbol{R}(\infty)$ by $\boldsymbol{R}$ with the (i,j)th element $R_{i,j}$. Then

$$\boldsymbol{R} = \boldsymbol{I} + (\boldsymbol{\beta} - \boldsymbol{e})\boldsymbol{\kappa} + (\boldsymbol{C} + \boldsymbol{D})(\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q})^{-1}. \tag{12}$$

(b). $\boldsymbol{R}$ is a nonnegative matrix whose spectral radius is less than one.

(c). $\boldsymbol{\pi}$ and $\boldsymbol{\kappa}$ are related by

$$\boldsymbol{\pi} = (1-\rho)\boldsymbol{\kappa}(\boldsymbol{I} - \boldsymbol{R})^{-1}. \tag{13}$$

(d). For each $i \in \mathcal{M}$, if $D_{i,j} = 0$ for all $j \in \mathcal{M}$, $R_{i,j} = 0$ for all $j \in \mathcal{M}$. On the other hand, if $D_{i,j} \geq 0$ for some $i, j \in \mathcal{M}$, $R_{i,j} > 0$ for all $j \in \mathcal{M}$.

We then divide $\mathcal{M}$ into two disjoint subsets $\mathcal{M}_0$ and $\mathcal{M}_1$ in such a way that $i \in \mathcal{M}_0$ if $\boldsymbol{D}_{i,j} = 0$ for all $j \in \mathcal{M}$, and otherwise $i \in \mathcal{M}_1$.

(e). Let $R_{i,j}(x)$ denote the (i,j)th element of $\boldsymbol{R}(x)$. For $i \in \mathcal{M}_1$ and for all $j \in \mathcal{M}$, $R_{i,j}(x)/R_{i,j}$ is a proper distribution function.

4 Subexponential asymptotic formulas

We first consider a single-server queue with MAS input described in section 3. For a while, we assume that there exists at least one pair (k,l) $(k,l \in \mathcal{M})$ such that $D_{k,l} > 0$ and $\overline{H_{k,l}}(x) > 0$ for all $x > 0$. We then have for all $i,j \in \mathcal{M}$,

$$\lim_{x\to\infty} \frac{\overline{H_{i,j}}(x)}{\overline{H_{k,l}}(x)} \in [0,\infty].$$

Thus there exists a pair (m,n) $(m,n \in \mathcal{M}$ such that $D_{m,n} > 0)$ such that

$$\lim_{x\to\infty} \frac{\overline{H_{i,j}}(x)}{\overline{H_{m,n}}(x)} \in [0,1], \qquad \text{for all } i,j \in \mathcal{M}. \tag{14}$$

We call $H_{m,n}(x)$ satisfying (14) the service time distribution function with the fattest tail.

Definition 1 *Let $H(x)$ denote the service time distribution function with the fattest tail among $H_{i,j}(x)$'s $(i,j \in \mathcal{M})$. We denote its mean by h $(< \infty)$:*

$$h = \int_0^\infty x dH(x).$$

Further let $h^*_{i,j}$ be

$$\lim_{x\to\infty} \frac{\overline{H_{i,j}}(x)}{\overline{H}(x)} = h^*_{i,j} \in [0,1], \qquad \textit{for all } i,j \in \mathcal{M}. \tag{15}$$

We now consider the asymptotic formula for $\boldsymbol{v}(x)$ under the following assumption.

Assumption 1 *The equilibrium distribution function $H_e(x)$ of the fattest-tailed service time distribution $H(x)$ is subexponential.*

Sufficient conditions so that $H_e(x)$ is subexponential are investigated in [5]. Note that the subexponentiality of $H_e(x)$ implies that $H_e(x)$ is heavy-tailed, from which one can readily show $\overline{H}(x) > 0$ for all $x > 0$. Thus Assumption 1 implies that there exists a pair (k,l) $(k,l \in \mathcal{M})$ such that $\overline{H_{k,l}}(x) > 0$ for all $x > 0$, as well.

Let $\overline{\boldsymbol{D}}(x)$ denote an $M \times M$ matrix whose (i,j)th element $\overline{D_{i,j}}(x)$ is given by $D_{i,j}\overline{H_{i,j}}(x)$, i.e., $\overline{\boldsymbol{D}}(x) = \boldsymbol{D} - \boldsymbol{D}(x)$. It then follows from Definition 1 that

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{D}}(x)}{\overline{H}(x)} = \frac{\boldsymbol{D}^*}{h} \qquad i,j \in \mathcal{M}, \tag{16}$$

where $\boldsymbol{D}^*$ denotes an $M \times M$ matrix whose (i,j)th element $D^*_{i,j}$ is given by

$$D^*_{i,j} = hD_{i,j}h^*_{i,j}, \qquad i,j \in \mathcal{M}. \tag{17}$$

Note here that $0 \leq D^*_{i,j} < hD_{i,j} < \infty$ and $\boldsymbol{D}^* \neq \boldsymbol{O}$. Let $\overline{\boldsymbol{v}}(x)$ denote the tail distribution of $\boldsymbol{v}(x)$, i.e.,

$$\overline{\boldsymbol{v}}(x) = \boldsymbol{v}(\infty) - \boldsymbol{v}(x) = \boldsymbol{\pi} - \boldsymbol{v}(x).$$

Theorem 1 *Under Assumption 1,* $\overline{\boldsymbol{v}}(x)$ *satisfies*

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{v}}(x)}{\overline{H_e}(x)} = \frac{\boldsymbol{\pi}\boldsymbol{D}^*\boldsymbol{e}}{1-\rho}\boldsymbol{\pi}. \tag{18}$$

The proof of Theorem 1 is provided in section 5. From Theorem 1 and relationships in (10) and (11), we obtain asymptotic formulas for actual waiting time and sojourn time distributions. Let $\overline{\boldsymbol{w}}(x) = \boldsymbol{w}(\infty) - \boldsymbol{w}(x)$ and $\overline{\boldsymbol{t}}(x) = \boldsymbol{t}(\infty) - \boldsymbol{t}(x)$. Note that

$$\boldsymbol{w}(\infty) = \boldsymbol{t}(\infty) = \frac{\boldsymbol{\pi}\boldsymbol{D}}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}.$$

Corollary 1 *Under Assumption 1, for all* $i, j \in \mathcal{M}$ *such that* $D_{i,j} > 0$, $\overline{w_{i,j}}(x)$ *and* $\overline{t_{i,j}}(x)$ *satisfy*

$$\lim_{x\to\infty} \frac{\overline{w_{i,j}}(x)}{\overline{H_e}(x)} = \lim_{x\to\infty} \frac{\overline{t_{i,j}}(x)}{\overline{H_e}(x)} = \frac{\boldsymbol{\pi}\boldsymbol{D}^*\boldsymbol{e}}{1-\rho},$$

and $\overline{\boldsymbol{w}}(x)$ *and* $\overline{\boldsymbol{t}}(x)$ *satisfy*

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{w}}(x)}{\overline{H_e}(x)} = \lim_{x\to\infty} \frac{\overline{\boldsymbol{t}}(x)}{\overline{H_e}(x)} = \frac{\boldsymbol{\pi}\boldsymbol{D}^*\boldsymbol{e}}{1-\rho} \cdot \frac{\boldsymbol{\pi}\boldsymbol{D}}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}.$$

The results for the actual waiting time distribution are the immediate consequence of Theorem 1, (10) and (11). On the other hand, the results for the sojourn time distribution come from (10), (11) and Lemmas 1 and 3 (a). Lemma 1 implies that

$$\lim_{x\to\infty} \frac{\overline{H_{i,j}}(x)}{\overline{H_e}(x)} = \lim_{x\to\infty} \frac{\overline{H_{i,j}}(x)}{\overline{H}(x)} \cdot \frac{\overline{H}(x)}{\overline{H_e}(x)} = 0.$$

Thus applying the above equation together with Lemma 3 (a) and Theorem 1 to (10) and (11), we obtain the desired results.

We now define H^G as a generic random variable representing a service time of a randomly chosen customer, whose distribution function is denoted by $H^G(x)$. We then have

$$\overline{H^G}(x) = \frac{\boldsymbol{\pi}\overline{\boldsymbol{D}}(x)\boldsymbol{e}}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}},$$

and the mean $h^G = E[H^G]$ is given by (see (5))

$$h^G = \frac{\rho}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}.$$

Lemma 4

$$\lim_{x\to\infty}\int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy = \boldsymbol{D}^*.$$

The proof is given in Appendix A. It follows from Lemma 4 that

$$\begin{aligned}
\lim_{x\to\infty}\frac{\overline{H_e^G}(x)}{\overline{H_e}(x)} &= \lim_{x\to\infty}\frac{1}{\overline{H_e}(x)}\int_x^\infty \frac{\overline{H^G}(y)}{h^G}dy \\
&= \lim_{x\to\infty}\frac{1}{\overline{H_e}(x)}\int_0^\infty \frac{\overline{H^G}(x+y)}{h^G}dy \\
&= \frac{\boldsymbol{\pi}}{h^G\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}\left(\lim_{x\to\infty}\int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy\right)\boldsymbol{e} \\
&= \frac{\boldsymbol{\pi}\boldsymbol{D}^*\boldsymbol{e}}{\rho} > 0. \qquad (19)
\end{aligned}$$

Thus Assumption 1 is equivalent to the following:

Assumption 2 *The equilibrium distribution function $H_e^G(x)$ of the overall (i.e., customer-average) service time distribution $H^G(x)$ is subexponential.*

The equivalence comes from (19) and Lemma 2, because for some constant $0 < c < \infty$

$$\lim_{x\to\infty}\frac{\overline{G}(x)}{\overline{H_e}(x)} = c \qquad \Leftrightarrow \qquad \lim_{x\to\infty}\frac{\overline{H_e}(x)}{\overline{G}(x)} = c^{-1}.$$

Note also that if for some constant c $(0 < c < \infty)$,

$$\lim_{x\to\infty}\frac{\overline{G}(x)}{\overline{H_e}(x)} = c,$$

then

$$\lim_{x\to\infty}\frac{\overline{G}(x)}{\overline{H_e^G}(x)} = \lim_{x\to\infty}\frac{\overline{H_e}(x)}{\overline{H_e^G}(x)}\cdot\frac{\overline{G}(x)}{\overline{H_e}(x)} = \frac{\rho c}{\boldsymbol{\pi}\boldsymbol{D}^*\boldsymbol{e}}.$$

As a result, we have the following corollary.

Corollary 2 *Under Assumption 2, we have*

$$\begin{aligned}
\lim_{x\to\infty}\frac{\overline{\boldsymbol{v}}(x)}{\overline{H_e^G}(x)} &= \frac{\rho}{1-\rho}\boldsymbol{\pi}, \\
\lim_{x\to\infty}\frac{\overline{w_{i,j}}(x)}{\overline{H_e^G}(x)} &= \lim_{x\to\infty}\frac{\overline{t_{i,j}}(x)}{\overline{H_e^G}(x)} = \frac{\rho}{1-\rho}, \\
\lim_{x\to\infty}\frac{\overline{\boldsymbol{w}}(x)}{\overline{H_e^G}(x)} &= \lim_{x\to\infty}\frac{\overline{\boldsymbol{t}}(x)}{\overline{H_e^G}(x)} = \frac{\rho}{1-\rho}\cdot\frac{\boldsymbol{\pi}\boldsymbol{D}}{\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}}.
\end{aligned}$$

Corollary 2 implies that

$$\lim_{x\to\infty}\frac{\overline{\boldsymbol{v}}(x)\boldsymbol{e}}{\overline{H_e^G}(x)} = \lim_{x\to\infty}\frac{\overline{\boldsymbol{w}}(x)\boldsymbol{e}}{\overline{H_e^G}(x)} = \lim_{x\to\infty}\frac{\overline{\boldsymbol{t}}(x)\boldsymbol{e}}{\overline{H_e^G}(x)} = \frac{\rho}{1-\rho}. \qquad (20)$$

Namely, all three quantities V, W and T are subexponential (see Lemma 2) and have the same asymptotic constant given only in terms of the utilization factor ρ, independent of the structure of the joint process of arrivals and services (i.e., $\boldsymbol{C}$ and $\boldsymbol{D}(x)$). Therefore, the same results hold for the single-server queue with MPP arrivals, as well, because MAP with state-dependent service times described in section 3 is dense in stationary marked point processes (MPPs) [1]. This is formally stated as follows.

We consider a stationary work-conserving FIFO G/G/1 queue. The input process is characterized by an ergodic stationary sequence of $\{(A_n, H_n); n = 0, 1, \ldots\}$, where A_n and H_n represent the arrival epoch and service time of the nth customer, respectively. Let A and H^G denote generic random variables for $\{A_n; n = 1, 2, \ldots\}$ and $\{H_n; n = 1, 2, \ldots\}$, respectively. We define $H^G(x)$ as $\Pr(H^G \leq x)$. Let $\rho = E[H^G]/E[A] < 1$. We define V, W and T as generic random variables representing stationary virtual waiting time, actual waiting time and sojourn time, respectively, in the above FIFO G/G/1 queue. It then follows from Corollary 3 in [1] that there exists a sequence of $\{(\boldsymbol{C}^{(m)}, \boldsymbol{D}^{(m)}(x));\ m = 1, 2, \ldots\}$ such that the $W^{(m)}$ converges to W in distribution, where $W^{(m)}$ denotes the stationary actual waiting time in the mth FIFO single-server queue with arrivals represented by $(\boldsymbol{C}^{(m)}, \boldsymbol{D}^{(m)}(x))$. The same result for V and T can be obtained by small modifications [3]. Thus from Lemma 2 and (20), we have the following theorem.

Theorem 2 *In the ergodic and stationary FIFO G/G/1 queue satisfying Assumption 2, all V, W and T are subexponential, and*

$$\lim_{x\to\infty} \frac{\Pr(V > x)}{\overline{H_e^G}(x)} = \lim_{x\to\infty} \frac{\Pr(W > x)}{\overline{H_e^G}(x)} = \lim_{x\to\infty} \frac{\Pr(T > x)}{\overline{H_e^G}(x)} = \frac{\rho}{1-\rho}. \tag{21}$$

Before proving Theorem 1, we provide some remarks on Theorem 2. For the actual waiting time in the stationary GI/GI/1 queue, the limit in (21) was shown by Pakes [12] under Assumption 2. Further in [15], Sigman pointed out that the limit also holds for the sojourn time distribution, too. In the FIFO G/GI/1 queue, the following invariance relationship holds [14]

$$\Pr(V \leq x) = 1 - \rho + \rho \int_0^x \Pr(W \leq x - y) dH_e^G(y),$$

from which and Lemma 3 (a), one can show that the limit in (21) holds for the virtual waiting time, too. For the actual waiting time in the Markov-modulated G/G/1 queue, Jelenković and Lazar showed the limit in (21) holds under the assumption that $H^G(x)$ is heavy-tailed together with Assumption 2. Because the subexponential $H_e^G(x)$ does not implies the heavy-tailed $H^G(x)$ (see Remark 6.2 in [15]), Theorem 2 is more general in this regard, too.

5 Proof of Theorem 1

We define $\overline{\boldsymbol{R}}(x)$ $\overline{\boldsymbol{R}^{(n)}}(x)$ $(n = 1, 2, \ldots)$ as $\boldsymbol{R} - \boldsymbol{R}(x)$ and $\boldsymbol{R}^n - \boldsymbol{R}^{(n)}(x)$, respectively. Note that $\overline{\boldsymbol{R}^{(1)}}(x) = \overline{\boldsymbol{R}}(x)$. It then follows from (9) and (13) that

$$\begin{aligned}\boldsymbol{v}(x) &= (1-\rho)\boldsymbol{\kappa} + (1-\rho)\boldsymbol{\kappa}\sum_{n=1}^{\infty}\left(\boldsymbol{R}^n - \overline{\boldsymbol{R}^{(n)}}(x)\right) \\ &= \boldsymbol{\pi} - (1-\rho)\boldsymbol{\kappa}\sum_{n=1}^{\infty}\overline{\boldsymbol{R}^{(n)}}(x).\end{aligned}$$

Thus

$$\overline{\boldsymbol{v}}(x) = (1-\rho)\boldsymbol{\kappa}\sum_{n=1}^{\infty}\overline{\boldsymbol{R}^{(n)}}(x),$$

Proposition 2 *For each $\epsilon > 0$ and $n = 1, 2, \ldots$, there exists some constant $K < \infty$ such that*

$$\frac{\overline{\boldsymbol{R}^{(n)}}(x)}{\overline{H_e}(x)} \le K(1+\epsilon)^n\boldsymbol{R}^n, \qquad \textit{for all} \quad x \ge 0, \tag{22}$$

where K is independent of n. Further

$$\lim_{x\to\infty}\frac{\overline{\boldsymbol{R}^{(n)}}(x)}{\overline{H_e}(x)} = \sum_{m=0}^{n-1}\boldsymbol{R}^m\boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}\boldsymbol{R}^{n-1-m}. \tag{23}$$

Once Proposition 2 is proved, Theorem 1 is shown as follows. Because the spectral radius of nonnegative matrix $\boldsymbol{R}$ is less than one (Proposition 1 (b)), we can choose ϵ in such a way that the spectral radius of $(1+\epsilon)\boldsymbol{R}$ is less than one. It then follows from (22) and dominated convergence theorem that

$$\lim_{x\to\infty}\frac{\overline{\boldsymbol{v}}(x)}{\overline{H_e}(x)} = (1-\rho)\boldsymbol{\kappa}\sum_{n=1}^{\infty}\lim_{x\to\infty}\frac{\overline{\boldsymbol{R}^{(n)}}(x)}{\overline{H_e}(x)}$$

Then applying (23) to the above equation, we obtain

$$\begin{aligned}\lim_{x\to\infty}\frac{\overline{\boldsymbol{v}}(x)}{\overline{H_e}(x)} &= (1-\rho)\boldsymbol{\kappa}\sum_{n=1}^{\infty}\sum_{m=0}^{n-1}\boldsymbol{R}^m\boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}\boldsymbol{R}^{n-1-m} \\ &= (1-\rho)\boldsymbol{\kappa}\sum_{m=0}^{\infty}\boldsymbol{R}^m\boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}\sum_{n=m+1}^{\infty}\boldsymbol{R}^{n-1-m} \\ &= (1-\rho)\boldsymbol{\kappa}(\boldsymbol{I}-\boldsymbol{R})^{-1}\boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}(\boldsymbol{I}-\boldsymbol{R})^{-1}.\end{aligned}$$

Theorem 1 immediately follows from the above equation and (13). Thus the remaining is to prove Proposition 2. We now start with the proof of the second part.

Lemma 5

$$\overline{\boldsymbol{R}}(x) = \int_0^{\infty}\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)dy.$$

The proof is given in Appendix B.

Lemma 6

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{R}}(x)}{\overline{H_e}(x)} = \boldsymbol{D}^* \boldsymbol{e}\boldsymbol{\kappa}, \qquad i,j \in \mathcal{M}.$$

Proof. We first note that for any $n > 0$

$$\begin{aligned}
\frac{\overline{\boldsymbol{R}}(x)}{\overline{H_e}(x)} &= \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy \\
&= \int_0^n \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy + \int_n^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy \\
&\le \frac{\overline{\boldsymbol{D}}(x)}{\overline{H_e}(x)} \int_0^n \exp(\boldsymbol{Q}y) dy + \int_n^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy \\
&= \frac{\overline{H}(x)}{\overline{H_e}(x)} \cdot \frac{\overline{\boldsymbol{D}}(x)}{\overline{H}(x)} \int_0^n \exp(\boldsymbol{Q}y) dy + \int_n^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy.
\end{aligned}$$

Therefore, we obtain for any $n > 0$

$$\begin{aligned}
\limsup_{x\to\infty} \frac{\overline{\boldsymbol{R}}(x)}{\overline{H_e}(x)} &\le 0 \cdot \frac{\boldsymbol{D}^*}{h} \int_0^n \exp(\boldsymbol{Q}y) dy + \limsup_{x\to\infty} \int_n^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy \\
&= \limsup_{x\to\infty} \int_n^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy. \qquad (24)
\end{aligned}$$

where we use Lemma 1.

Note here that $\exp(\boldsymbol{Q}x)$ is a stochastic matrix with an invariant probability vector $\boldsymbol{\kappa}$, so that

$$\lim_{x\to\infty} \exp(\boldsymbol{Q}x) = \boldsymbol{e}\boldsymbol{\kappa}.$$

Thus for any $\epsilon > 0$, we can choose an integer n^* large enough so that

$$\exp(\boldsymbol{Q}x) \le (1+\epsilon)\boldsymbol{e}\boldsymbol{\kappa}, \qquad \text{for all } x \ge n^*. \qquad (25)$$

It then follows from (24) and (25) that

$$\begin{aligned}
\int_{n^*}^\infty \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy &= \sum_{n=n^*}^\infty \int_n^{n+1} \frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)} dy \\
&\le (1+\epsilon) \sum_{n=n^*}^\infty \frac{\overline{\boldsymbol{D}}(x+n)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x)} \\
&\le (1+\epsilon) \int_{n^*-1}^\infty \frac{\overline{\boldsymbol{D}}(x+y)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x)} dy \\
&\le (1+\epsilon) \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x)} dy.
\end{aligned}$$

Then using Lemma 4, we have

$$\limsup_{x\to\infty}\int_{n^*}^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy \le \boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}.$$

On the other hand, the lower bound is obtained as follows. We first note that

$$\frac{\overline{\boldsymbol{R}}(x)}{\overline{H_e}(x)} \ge \int_m^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy, \qquad \text{for all } m=0,1,\ldots,$$

and for any $\epsilon_1 > 0$, there exists an integer m^* large enough so that

$$\exp(\boldsymbol{Q}x) \ge (1-\epsilon_1)\boldsymbol{e}\boldsymbol{\kappa}, \qquad \text{for all } \ x \ge m^*-1.$$

Therefore it suffices to show that

$$\liminf_{x\to\infty}\int_{m^*-1}^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy \ge \boldsymbol{D}^*\boldsymbol{e}\boldsymbol{\kappa}, \tag{26}$$

Note here that

$$\begin{aligned}
&\int_{m^*-1}^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy \\
&\quad= \sum_{m=m^*-1}^{\infty}\int_m^{m+1}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy \\
&\quad\ge (1-\epsilon_1)\sum_{m=m^*-1}^{\infty}\frac{\overline{\boldsymbol{D}}(x+m+1)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x)}dy \\
&\quad\ge (1-\epsilon_1)\int_{m^*}^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x)}dy \\
&\quad= (1-\epsilon_1)\int_0^{\infty}\frac{\overline{H_e}(x+m^*)}{\overline{H_e}(x)}\cdot\frac{\overline{\boldsymbol{D}}(x+m^*+y)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x+m^*)}dy.
\end{aligned}$$

Further for any $\epsilon_2 > 0$, there exists x^* large enough so that

$$\frac{\overline{H_e}(x+m^*)}{\overline{H_e}(x)} \ge 1-\epsilon_2, \qquad \text{for all } x \ge x^*,$$

because of the heavy-tailed property (1). Thus we have for all $x > x^*$

$$\begin{aligned}
&\int_{m^*-1}^{\infty}\frac{\overline{\boldsymbol{D}}(x+y)\exp(\boldsymbol{Q}y)}{\overline{H_e}(x)}dy \\
&\qquad\ge (1-\epsilon_1)(1-\epsilon_2)\int_0^{\infty}\frac{\overline{\boldsymbol{D}}(x+m^*+y)\boldsymbol{e}\boldsymbol{\kappa}}{\overline{H_e}(x+m^*)}dy.
\end{aligned} \tag{27}$$

Because ϵ_1 and ϵ_2 are arbitrary, (26) follows from Lemma 4 and (27), which completes the proof. □

We now define random variables $G_{i,j}$ $(i,j \in \mathcal{M})$ whose distribution functions $G_{i,j}(x)$ are given by

$$G_{i,j}(x) = \begin{cases} \dfrac{R_{i,j}(x)}{R_{i,j}}, & \text{if } i \in \mathcal{M}_1, \\ 1, & \text{if } i \in \mathcal{M}_0. \end{cases}$$

Recall that $G_{i,j}(x)$ is a proper distribution function (see Proposition 1 (e)). It then follows from Lemma 6 that

$$\lim_{x\to\infty} \frac{\overline{G_{i,j}}(x)}{\overline{H}_e(x)} = G^*_{i,j}, \qquad i,j = 1,\ldots,M,$$

where

$$G^*_{i,j} = \begin{cases} \displaystyle\sum_{k=1}^{M} \frac{D^*_{i,k}\kappa_j}{R_{i,j}}, & \text{if } i \in \mathcal{M}_1, \\ 0, & \text{if } i \in \mathcal{M}_0. \end{cases}$$

Thus we have

$$\frac{\left[\overline{\boldsymbol{R}^{(n)}}(x)\right]_{i,j}}{\overline{H}_e(x)} = \sum_{k_1\in\mathcal{M}} \cdots \sum_{k_{n-1}\in\mathcal{M}} R_{i,k_1}R_{k_1,k_2}\cdots R_{k_{n-2},k_{n-1}}R_{k_{n-1},j} \cdot \frac{\Pr(G_{i,k_1}+G_{k_1,k_2}+\cdots+G_{k_{n-2},k_{n-1}}+G_{k_{n-1},j} > x)}{\overline{H}_e(x)}. \tag{28}$$

Applying Lemma 3 (a) to (28), we have

$$\lim_{x\to\infty} \frac{\left[\overline{\boldsymbol{R}^{(n)}}(x)\right]_{i,j}}{\overline{H}_e(x)} = \sum_{k_1\in\mathcal{M}} \cdots \sum_{k_{n-1}\in\mathcal{M}} R_{i,k_1}R_{k_1,k_2}\cdots R_{k_{n-2},k_{n-1}}R_{k_{n-1},j} \cdot \left(G^*_{i,k_1}+G^*_{k_1,k_2}+\cdots+G^*_{k_{n-2},k_{n-1}}+G^*_{k_{n-1},j}\right).$$

Therefore, noting

$$R_{i,j}G^*_{i,j} = \sum_{k=1}^{M} D^*_{i,k}\kappa_j, \text{for all } i,j \in \mathcal{M},$$

we have (23).

Finally, the former part of Proposition 2 is obtained, by applying Lemma 3 (b) to (28), where K depends only on ϵ and $G^* = \max(G^*_{i,j}; i,j \in \mathcal{M})$, and independent of n. □

A Proof of Lemma 4

(16) implies that for any $\epsilon > 0$, there exists $x^* > 0$ enough large so that

$$\frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H}(x+y)} \le (1+\epsilon)\frac{\boldsymbol{D}^*}{h} \qquad \text{for all } x \ge x^*, y \ge 0,$$

from which it follows that for all $x \geq x^*$,

$$\begin{aligned}\int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy &= \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H}(x+y)} \cdot \frac{\overline{H}(x+y)}{\overline{H_e}(x)}dy \\ &\leq (1+\epsilon)\frac{\boldsymbol{D}^*}{h}\int_0^\infty \frac{\overline{H}(x+y)}{\overline{H_e}(x)}dy \\ &= (1+\epsilon)\boldsymbol{D}^*,\end{aligned}$$

because

$$\overline{H_e}(x) = h^{-1}\int_0^\infty \overline{H}(x+y)dy.$$

Since ϵ is arbitrary, we have

$$\limsup_{x\to\infty} \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy \leq \boldsymbol{D}^*. \tag{29}$$

On the other hand, the lower bound is obtained as follows. We first note that

$$\int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy \geq \int_z^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy, \qquad \text{for all } z \geq 0.$$

Thus it suffices to show that

$$\liminf_{x\to\infty} \int_z^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy \geq \boldsymbol{D}^*, \qquad \text{for some } z \geq 0. \tag{30}$$

Note that

$$\begin{aligned}\int_z^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy &= \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y+z)}{\overline{H_e}(x)}dy \\ &= \int_0^\infty \frac{\overline{\boldsymbol{D}}(x+y+z)}{\overline{H}(x+y+z)} \cdot \frac{\overline{H}(x+y+z)}{\overline{H_e}(x)}dy.\end{aligned} \tag{31}$$

It follows from (16) that for any $\epsilon > 0$, there exists $z^* > 0$ such that

$$\frac{\overline{\boldsymbol{D}}(x+z^*)}{\overline{H}(x+z^*)} \geq (1-\epsilon)\frac{\boldsymbol{D}^*}{h}, \qquad \text{for all } x \geq 0. \tag{32}$$

Thus using (31) and (32), we obtain that

$$\begin{aligned}\int_{z^*}^\infty \frac{\overline{\boldsymbol{D}}(x+y)}{\overline{H_e}(x)}dy &\geq (1-\epsilon)\boldsymbol{D}^*\int_0^\infty \frac{\overline{H}(x+y+z^*)}{h\overline{H_e}(x)}dy \\ &= (1-\epsilon)\boldsymbol{D}^*\frac{\overline{H_e}(x+z^*)}{\overline{H_e}(x)}.\end{aligned} \tag{33}$$

Since ϵ is arbitrary, (30) immediately follows from (33) and the heavy-tailed property (1), which completes the proof. □

B Proof of Lemma 5

Note first that

$$\begin{aligned}\overline{\boldsymbol{R}}(x) &= \int_x^\infty dw \int_w^\infty d\boldsymbol{D}(y) \exp(\boldsymbol{Q}(y-w)) \\ &= \int_x^\infty d\boldsymbol{D}(y) \int_x^y \exp(\boldsymbol{Q}(y-w))dw \\ &= \int_x^\infty d\boldsymbol{D}(y) \int_0^{y-x} \exp(\boldsymbol{Q}(w))dw. \end{aligned} \tag{34}$$

Post-multiplying both sides of (34) by $\boldsymbol{e\kappa} - \boldsymbol{Q}$, noting $\exp(\boldsymbol{Q}x)\boldsymbol{e} = \boldsymbol{e}$ and calculating the second integral, we have

$$\overline{\boldsymbol{R}}(x)(\boldsymbol{e\kappa} - \boldsymbol{Q}) = \int_x^\infty d\boldsymbol{D}(y)[(y-x)\boldsymbol{e\kappa} + \boldsymbol{I} - \exp(\boldsymbol{Q}(y-x))].$$

Partial integration then yields

$$\begin{aligned}\overline{\boldsymbol{R}}(x)(\boldsymbol{e\kappa} - \boldsymbol{Q}) &= \int_x^\infty \overline{\boldsymbol{D}}(y)[\boldsymbol{e\kappa} - \exp(\boldsymbol{Q}(y-x))\boldsymbol{Q}]dy. \\ &= \int_0^\infty \overline{\boldsymbol{D}}(x+y)[\boldsymbol{e\kappa} - \exp(\boldsymbol{Q}y)\boldsymbol{Q}]dy.\end{aligned}$$

Because $(\boldsymbol{e\kappa} - \boldsymbol{Q})$ is nonsingular and $\boldsymbol{\kappa}(\boldsymbol{e\kappa} - \boldsymbol{Q})^{-1} = \boldsymbol{\kappa}$, we have

$$\begin{aligned}\overline{\boldsymbol{R}}(x) &= \int_0^\infty \overline{\boldsymbol{D}}(x+y)[\boldsymbol{e\kappa} - \exp(\boldsymbol{Q}y)\boldsymbol{Q}(\boldsymbol{e\kappa} - \boldsymbol{Q})^{-1}]dy \\ &= \int_0^\infty \overline{\boldsymbol{D}}(x+y)[\boldsymbol{e\kappa} - \exp(\boldsymbol{Q}y)(\boldsymbol{e\kappa} - \boldsymbol{I})]dy,\end{aligned}$$

from which Lemma 5 follows. □

Acknowledgement

The author is thankful to Søren Asmussen for his valuable comment on the existence of a sequence of MASs, in connection with Theorem 2. This research was supported in part by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science under Grant No. 12650380.

REFERENCES

[1] Asmussen, S. and Koole, G. (1993) Marked point processes as limits of Markovian arrival streams. *J. Appl. Prob.* **30**, 365–372.

[2] Asmussen, S., Henriksen, L. F. and Klüppelberg, C. (1994) Large claims approximations for risk processes in a Markovian environment. *Stoch. Process. Appl.* **54**, 29–43.

[3] Asmussen, S. (2000) Private communications.

[4] Embrechts, P. and Goldie, C. M. (1982) On convolution tails. *Stoch. Process. Appl.* **13**, 263–278.

[5] Embrechts, P. and Omey, E. (1984) A property of longtailed distributions. *J. Appl. Prob.* **21**, 80–87.

[6] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

[7] Jelenković, P. R. and Lazar, A. A. (1998) Subexponential asymptotics of a Markov-modulated random walk with queueing applications. *J. Appl. Prob.* **35**, 325–347.

[8] König, D. and Schmidt, V. (1990) Extended and conditional versions of the PASTA property. *Adv. Appl. Prob.* **22** 510–512.

[9] Loynes, R. M. (1962) The stability of a queue with non-independent interarrival and service times. *Proc. Cambridge Philos. Soc.* **58** 497–520.

[10] Lucantoni, D. M., Meier-Hellstern, K. S. and Neuts, M. F. (1990) A single-server queue with server vacations and a class of non-renewal arrival processes. *Adv. Appl. Prob.* **22**, 676–705.

[11] Neuts, M. F. (1981) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach* (The Johns Hopkins University Press, Baltimore).

[12] Pakes, A. G. (1975) On the tails of waiting time distribution. *J. Appl. Prob.* **12**, 555–564.

[13] Seneta, E. (1976) *Regularly Varying Functions.* Lecture Notes in Math. **508**, Springer-Verlag, Berlin.

[14] Sengupta, B. (1989) An invariance relationship for the G/G/1 queue. *Adv. Appl. Prob.* **21**, 956-957.

[15] Sigman, K. (1999) Appendix: A primer on heavy-tailed distribution, *QUESTA* **33**, 261–275.

[16] Takine, T. and Hasegawa, T. (1994) The workload in the MAP/G/1 queue with state-dependent services: Its application to a queue with preemptive resume priority. *Stoch. Mod.* **10**, 183–204.

[17] Takine, T. (1996) A nonpreemptive priority MAP/G/1 queue with two classes of customers. *J. Opns. Res. Soc. Japan* **39** 266–290.

[18] Takine, T. (1996) A continuous version of matrix-analytic methods with the skip-free to the left property. *Stoch. Mod.* **12**, 673–682.

[19] Takine, T. (1999) The nonpreemptive priority MAP/G/1 queue. *Opns. Res.* **47** 917–927.

[20] Takine, T. (1999) Matrix product-form solution for an LCFS-PR single-server queue

with multiple arrival streams governed by a Markov chain. submitted for publication.