The M/G/1 queue with two service speeds

O.J. Boxma[1,2,*]   I.A. Kurkova[1]

January 28, 2000

1. EURANDOM
P.O. Box 513, 5600 MB Eindhoven,
THE NETHERLANDS.

2. Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven
The NETHERLANDS

Abstract

We consider an $M/G/1$ queue with the special feature that the speed of the server alternates between two constant values s_L and $s_H > s_L$. The high-speed periods are exponentially distributed, and the low-speed periods have a general distribution. Our main results are: (i) For the case that the distribution of the low-speed periods has a rational Laplace-Stieltjes transform, we obtain the joint distribution of the buffer content and the state of the server speed. (ii) For the case that the distribution of the low-speed periods and/or the service request distribution is regularly varying at infinity, we obtain explicit asymptotics for the tail of the buffer content distribution. The two cases in which the offered traffic load is smaller respectively larger than the low service speed are shown to result in completely different asymptotics.

1991 Mathematics Subject Classification: 60K25, 68M20.

Key words & Phrases: $M/G/1$ queue, variable speed, buffer content, regular variation.

---

*also: CWI, P.O.Box 94079, 1090 GB Amsterdam, the Netherlands

1 Introduction and model description

This paper is devoted to the $M/G/1$ queue that alternates between two service speeds s_L and $s_H > s_L$. The high-speed periods are assumed to be negative exponentially distributed, and the low-speed periods have a general distribution $L(\cdot)$. We are mainly interested in the distribution of the workload. In case $L(\cdot)$ has a rational Laplace-Stieltjes Transform (LST), we obtain an exact expression for the LST of the workload distribution. Subsequently we turn to the case that both $L(\cdot)$ and the service request distribution may have a regularly varying tail at infinity (see Definitions of the Appendix). We determine the tail behaviour of the workload. It turns out that one has to distinguish between two cases, (i) $s_L < \rho$ and (ii) $s_L > \rho$; here ρ indicates the offered request load, viz., the arrival rate of customers times their mean service request. In case (ii) the slow speed is already enough to handle the offered traffic, and the tail behaviour of the workload is determined by the tail behaviour of the service request distribution. In case (i) the workload has a positive drift during the low-speed periods, and the tail behaviour of the workload distribution is determined by the heaviest of the tails of the service request distribution and $L(\cdot)$.

Our motivation for this study is two-fold. Firstly, the single-server queue with various speeds is a very natural and practically important model, that deserves a renewed interest w.r.t. the performance analysis of integrated-services communication networks. In such networks, the influence of high-priority traffic on low-priority traffic is often reflected in a variable capacity for low-priority traffic. For example, the available capacity for ABR or non real-time VBR traffic depends on the presence or absence of high-priority real-time VBR traffic. Other examples are provided by scheduling disciplines like GPS (Generalized Processor Sharing). As a design paradigm, GPS is at the heart of commonly-used scheduling algorithms for high-speed switches such as Weighted Fair Queueing. From a queueing point of view, GPS gives rise to the analysis of coupled servers, where the speed of one server varies depending on whether another server is busy or idle.

A second motivation for this study is the convincing evidence of long-tailed traffic characteristics in high-speed communication networks. Early indications of the long-range dependence of Ethernet traffic, attributed to long-tailed file size distributions, were reported in [18]. Long-tailed characteristics of the scene length distribution of MPEG video streams were inves-

tigated in [16, 17]. These and other empirical findings have triggered theoretical developments in the modelling and queueing analysis of long-tailed traffic phenomena. The influence of long-tailed service time distributions on waiting time and workload distributions of the single-server queue has been investigated in considerable detail; many results are gathered in the book [22]. For multi-server queues, very little is known. A recent study [6] has been devoted to an analysis of the workload tail behaviour of a heterogeneous two-server queue with one exponential and one heavy-tailed server. This tail behaviour is shown to exhibit completely different behaviour, depending on whether the exponential server alone does or does not have enough capacity to handle all offered traffic. In the framework of GPS, a similar phenomenon has been observed, cf. [4, 5]. In the present study we investigate the influence of long-tailed periods with slow service on the tail behaviour of the workload of a single-server queue. It should be noted that the low-speed period may correspond to a period in which another server in a GPS-operated switch is busy. Our model is in one sense more general: The low-speed periods do not necessarily correspond to busy periods of another queue.

There is a considerable literature on the single-server queue with several service speeds, but to the best of our knowledge the issue of regularly varying tail behaviour has not yet been discussed in this respect. An early paper is due to Yechiali and Naor [26]. They have studied the $M/M/1$ queue which alternates between two (see [27] for an extension) exponentially distributed phases, the arrival and service rates depending on the phase. Neuts [21] has generalized their study to the $M/G/1$ case. He deviates from the assumptions in [26] by assuming that the service time distribution of a customer depends only on the state of the phase process *at the time his service begins*. Halfin [14] analyzes the buffer content of an $M/G/1$ queue whose service rate varies according to a birth-and-death process with $c + 1$ states. A system of Volterra-type integral equations is derived for the joint distribution of the buffer content and the phase of the birth-and-death process, and is used for the numerical calculation of the distribution. Regterschot and De Smit [23] have studied an $M/G/1$ queue in which both the arrival rate and the service time distribution depend on the state of an underlying finite-state Markov chain. Dudin [11] and Dudin and Markov [12] have analysed the most general model: an M/G/1 queue in a semi-Markovian environment. In those papers the residence times in states of the semi-Markov process are assumed to have a limited PH-distribution. Several authors have considered queues

3

with service interruptions. In our setting this corresponds to taking $s_L = 0$. Some recent studies concerning such queues are Takine and Sengupta [25], Li et al. [19], and Núñez Queija [24].

The paper is organized as follows. In Section 2 we derive an expression for the joint steady-state distribution of the buffer content and the state of the server (low- or high speed). These expressions involve an unknown function $Q(x)$. In Section 3 we show how $Q(x)$ can be determined in case the low-speed period distribution has a rational LST. Unfortunately, we are not able to determine $Q(x)$ in the general case. However, in Section 4 we demonstrate how one can still determine the tail behaviour of the buffer content distribution, when $L(\cdot)$ and/or the service request distribution is regularly varying at infinity. We provide explicit asymptotics for this tail behaviour, distinguishing between $\rho < s_L$ and $\rho > s_L$. In the former case the tail of the unknown distribution is only determined by the tail of the service request distribution; in the latter case, it is determined by the heaviest of the tails of the service request distribution and the low-speed period distribution.

We end the present section with a detailed *model description*. We consider the $M/G/1$ queue with an infinite buffer, where customers arrive according to a Poisson process with rate λ. The required service times have distribution $B(t)$ with LST $\beta(s)$, Re $s \geq 0$, and mean β. Service is in order of arrival. The speed of the server alternates between two constant values s_L and $s_H > s_L$. The high-speed periods are distributed exponentially with mean $1/\nu$. The low-speed periods have distribution $L(t)$ with LST $\delta(s)$, Re $s \geq 0$, and mean δ. All interarrival intervals, service requests, lengths of high-speed periods and lengths of low-speed periods are independent.

2   The buffer content

Let $V(t)$ be the buffer content (workload) at time t. Let $X(t)$ be a random variable that alternates between the states H and L. $X(t)$ does not depend on $V(t)$. If $X(t) = H$ (respectively $X(t) = L$) then a customer in service is served with speed s_H (s_L). Observe that $\{(V(t), X(t)),\ t \geq 0\}$ is not a Markov process. A Markov process can be obtained by taking a third quantity into consideration, viz., the time $L_{\text{past}}(t)$ that has passed since the last change of service speed. We can study the joint distribution of the Markov process $\{(V(t), X(t), L_{\text{past}}(t)),\ t \geq 0\}$ in a similar way as Cohen [[8], Sec-

tion II.6.2] analyzes the $M/G/1$ queue (via the method of the supplementary variable). Introduce the following distribution functions, for $v \geq 0$, $t, \eta > 0$:

$$F_H(v,t) \quad := \quad P(V(t) \leq v, \; X(t) = H),$$
$$F_L(v,t,\eta)\mathrm{d}\eta \quad := \quad P(V(t) \leq v, X(t) = L, \; \eta < L_{\text{past}}(t) \leq \eta + \mathrm{d}\eta).$$

The ergodicity condition for this system is easily seen to be:

$$\lambda\beta < \frac{\delta}{\delta + 1/\nu}s_L + \frac{1/\nu}{\delta + 1/\nu}s_H. \tag{1}$$

A formal proof may be done using Lyapunov functions, cf. Fayolle et al. [13]; we refrain from giving this proof. If this ergodicity condition holds, then we shall denote the stationary workload by V and the corresponding steady-state distribution functions by $F_H(v)$ and $F_L(v,\eta)$.

It readily follows from the model specification that, for $v, t, \eta > 0$, $\Delta t \downarrow 0$:

$$F_H(v, t + \Delta t)$$

$$= \quad F_H(v + s_H\Delta t, t)(1 - \lambda\Delta t)(1 - \nu\Delta t) + \lambda\Delta t \int\limits_{x=0}^{v} B(v - x)dF_H(x,t)$$

$$+ (1 - \lambda\Delta t) \int\limits_{\eta=0}^{t} F_L(v + s_L\Delta t, \eta, t)P(\eta < L \leq \eta + \Delta t \mid L_{\text{past}} > \eta)\mathrm{d}\eta$$

$$+ o(\Delta t)$$

$$= \quad \Big(F_H(v,t) + \frac{\partial F_H(v,t)}{\partial v}s_H\Delta t\Big)(1 - \lambda\Delta t)(1 - \nu\Delta t)$$

$$+ \lambda\Delta t \int\limits_{x=0}^{v} B(v - x)\mathrm{d}F_H(x,t)$$

$$+ \Delta t \int\limits_{\eta=0}^{t} \frac{F_L(v + s_L\Delta t, \eta, t)\,\mathrm{d}L(\eta)}{1 - L(\eta)} + o(\Delta t), \tag{2}$$

$$F_L(v, \eta, t + \Delta t)$$

$$= \quad F_L(v + s_L\Delta t, \eta - \Delta t, t)\frac{1 - L(\eta)}{1 - L(\eta - \Delta t)}(1 - \lambda\Delta t)$$

5

$$+ \lambda \Delta t \int\limits_{x=0}^{v} B(v-x) \mathrm{d} F_L(x, \eta - \Delta t, t) + o(\Delta t)$$

$$= \left(F_L(v, \eta, t) + \frac{\partial F_L(v, \eta, t)}{\partial v} s_L \Delta t - \frac{\partial F_L(v, \eta, t)}{\partial \eta} \right) \frac{1 - L(\eta)}{1 - L(\eta - \Delta t)} (1 - \lambda \Delta t)$$

$$+ \lambda \Delta t \int\limits_{x=0}^{v} B(v-x) \mathrm{d} F_L(x, \eta, t) + o(\Delta t), \tag{3}$$

$$P(V(t) \le v, X(t) = L, L_{\mathrm{past}}(t) \le \Delta t) = \nu \Delta t F_H(v + s_H \Delta t) + o(\Delta t). \tag{4}$$

Let Δt tend to zero in (2), (3) and (4). Assume that the ergodicity condition (1) holds. Then we get the following equations for the steady-state distribution functions:

$$s_H \frac{\partial F_H(v)}{\partial v} = \nu F_H(v) + \lambda \int\limits_{x=0}^{v} (1 - B(v-x)) \mathrm{d} F_H(x)$$

$$- \int\limits_{\eta=0}^{\infty} F_L(v, \eta) \frac{\mathrm{d} L(\eta)}{1 - L(\eta)}, \tag{5}$$

$$\frac{\partial F_L(v, \eta)}{\partial \eta} = s_L \frac{\partial F_L(v, \eta)}{\partial v} - F_L(v, \eta) \frac{L'(\eta)}{1 - L(\eta)}$$

$$- \lambda F_L(v, \eta) + \lambda \int\limits_{x=0}^{v} B(v-x) \mathrm{d} F_L(x, \eta), \tag{6}$$

$$F_L(v, 0+) = \nu F_H(v). \tag{7}$$

Introduce the following LST:

$$\Phi_H(\omega) := \int\limits_{v=0-}^{\infty} \exp\{-\omega v\} \mathrm{d} F_H(v), \qquad \mathrm{Re}\,\omega \ge 0,$$

$$\Phi_L(\omega, \eta) := \int\limits_{v=0-}^{\infty} \exp\{-\omega v\} \mathrm{d} F_L(v, \eta), \qquad \mathrm{Re}\,\omega \ge 0.$$

Then by (5), (6) and (7) we have

$$s_H[\Phi_H(\omega) - F_H(0)] = \frac{\nu\Phi_H(\omega)}{\omega} + \lambda\frac{1-\beta(\omega)}{\omega}\Phi_H(\omega)$$

$$-\frac{1}{\omega}\int\limits_{\eta=0}^{\infty}\frac{\Phi_L(\omega,\eta)}{1-L(\eta)}dL(\eta), \tag{8}$$

$$\frac{1}{\omega}\frac{\partial\Phi_L(\omega,\eta)}{\partial\eta} = s_L[\Phi_L(\omega,\eta) - F_L(0,\eta)] - \frac{1}{\omega}\Phi_L(\omega,\eta)\frac{L'(\eta)}{1-L(\eta)}$$

$$-\lambda\frac{1-\beta(\omega)}{\omega}\Phi_L(\omega,\eta), \qquad \eta > 0, \tag{9}$$

$$\Phi_L(\omega,0+) = \nu\Phi_H(\omega). \tag{10}$$

Now define the function

$$\Psi_L(\omega,\eta) := \frac{\Phi_L(\omega,\eta)}{1-L(\eta)}, \quad \mathrm{Re}\,\omega \geq 0, \quad \eta \geq 0.$$

Then (9) implies the equation

$$\frac{\partial\Psi_L(\omega,\eta)}{\partial\eta} = (s_L\omega - \lambda(1-\beta(\omega)))\Psi_L(\omega,\eta) - s_L\omega\frac{F_L(0,\eta)}{1-L(\eta)}. \tag{11}$$

Writing

$$\Psi_L(\omega,\eta) = C(\omega,\eta)\exp\{(s_L\omega - \lambda(1-\beta(\omega)))\eta\},$$

the function $C(\omega,\eta)$ should satisfy the equation

$$\frac{\partial C(\omega,\eta)}{\partial\eta} = -s_L\omega\exp\{-(s_L\omega - \lambda(1-\beta(\omega)))\eta\}\frac{F_L(0,\eta)}{1-L(\eta)}.$$

Hence

$$C(\omega,\eta) = C(\omega,0) - s_L\omega\int\limits_{x=0}^{\eta}\exp\{-(s_L\omega - \lambda(1-\beta(\omega)))x\}\frac{F_L(0,x)}{1-L(x)}dx. \tag{12}$$

Moreover, by (10):

$$C(\omega,0) = \Psi_L(\omega,0+) = \Phi_L(\omega,0+) = \nu\Phi_H(\omega). \tag{13}$$

7

Thus

$$\Psi_L(\omega, \eta) = \exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\} \tag{14}$$
$$\times \Big[\nu\Phi_H(\omega) - s_L\omega \int\limits_{x=0}^{\eta} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0, x)}{1 - L(x)}dx\Big].$$

Finally we get the following equation for $\Phi_L(\omega, \eta)$:

$$\Phi_L(\omega, \eta) = (1 - L(\eta))\exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\} \tag{15}$$
$$\times \Big[\nu\Phi_H(\omega) - s_L\omega \int\limits_{x=0}^{\eta} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0, x)}{1 - L(x)}dx\Big],$$

and, using (8) and (14), the following equation for $\Phi_H(\omega)$:

$$\Phi_H(\omega)[\nu + \lambda(1 - \beta(\omega)) - s_H\omega]$$
$$= -s_H\omega F_H(0) + \int\limits_{\eta=0}^{\infty} \Psi_L(\omega, \eta)dL(\eta)$$
$$= -s_H\omega F_H(0) + \int\limits_{\eta=0}^{\infty} \exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\} \tag{16}$$
$$\times \Big[\nu\Phi_H(\omega) - s_L\omega \int\limits_{x=0}^{\eta} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0, x)}{1 - L(x)}dx\Big]dL(\eta).$$

We shall work out the obtained equations (15) and (16) in two particular cases.

1. Consider ω, $\mathrm{Re}\,\omega \geq 0$, such that

$$\delta\{\lambda(1 - \beta(\omega)) - s_L\omega\} < \infty. \tag{17}$$

Recall that $\delta(s)$ is the LST of the distribution $L(\cdot)$ of the low-speed periods. It follows from (16) that

$$\Phi_H(\omega)[\nu + \lambda(1 - \beta(\omega)) - s_H\omega - \nu\delta\{\lambda(1 - \beta(\omega)) - s_L\omega\}]$$
$$= -s_H\omega F_H(0) - s_L\omega R(\omega), \tag{18}$$

where

$$R(\omega) \; := \; \int\limits_{\eta=0}^{\infty} \exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\} \tag{19}$$

$$\times \int\limits_{x=0}^{\eta} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0, x)}{1 - L(x)}\mathrm{d}x\mathrm{d}L(\eta).$$

Note that (17) necessarily holds whenever

$$\mathrm{Re}\left(\lambda(1 - \beta(\omega)) - s_L\omega\right) \geq 0, \tag{20}$$

but may not hold when the inequality is reversed. If $\lambda\beta > s_L$, then (20), and hence also (18), is valid for all ω small enough with positive real part.

Note also, that whenever $s_L = 0$, (17) is true for all ω with $\mathrm{Re}\,\omega \geq 0$ and, moreover, the equation (18) is easily solvable. In fact, substituting $\omega = 0$ into (18), we get the constant

$$F_H(0) = \frac{\lambda\beta(1 + \nu\delta) - s_H}{1 + \nu\delta}$$

and we then derive $\Phi_H(\omega)$ and $\Phi_L(\omega, \eta)$ from (18) and (15) respectively.

2. Let us now suppose that

$$\mathrm{Re}\left(\lambda(1 - \beta(\omega)) - s_L\omega\right) < 0. \tag{21}$$

In this case we have

$$\begin{aligned}
0 \;\; &\leq \;\; \lim_{\eta\to\infty} |C(\omega, \eta)| = \lim_{\eta\to\infty} |\exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))\eta\}\Psi_L(\omega, \eta)| \\
&\leq \;\; \lim_{\eta\to\infty} |\exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))\eta\}\Psi_L(0, \eta)|, \;\; \mathrm{Re}\,\omega > 0.
\end{aligned}$$

Note that the latter limit equals zero, because of (21) and because of the fact that $\Psi_L(0, \eta)$ is constant, which is an immediate corollary of (11). Then using (12), (13) and (15),

$$C(\omega, \eta) \;\; = \;\; s_L\omega \int\limits_{x=\eta}^{\infty} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0, x)}{1 - L(x)}\mathrm{d}x,$$

9

$$\Phi_H(\omega) = \frac{s_L\omega}{\nu} \int\limits_{x=0}^{\infty} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0,x)}{1 - L(x)}\mathrm{d}x, \qquad (22)$$

$$\Phi_L(\omega,\eta) = s_L\omega(1 - L(\eta))\exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\} \qquad (23)$$

$$\times \int\limits_{x=\eta}^{\infty} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0,x)}{1 - L(x)}\mathrm{d}x,$$

and finally using (16) and (22), when (21) holds:

$$\Phi_H(\omega)[\nu + \lambda(1 - \beta(\omega)) - s_H\omega] \qquad (24)$$

$$= -s_H\omega F_H(0) + s_L\omega \int\limits_{\eta=0}^{\infty} \exp\{(s_L\omega - \lambda(1 - \beta(\omega)))\eta\}$$

$$\times \int\limits_{x=\eta}^{\infty} \exp\{-(s_L\omega - \lambda(1 - \beta(\omega)))x\}\frac{F_L(0,x)}{1 - L(x)}\mathrm{d}x\mathrm{d}L(\eta).$$

Note that, if $\lambda\beta < s_L$, then (21) is true for all ω small enough with $\operatorname{Re}\omega > 0$ and (23) and (24) hold for all such ω.

If the function $F_L(0,x)/(1 - L(x))$ were known, then $\Phi_H(\omega)$ and $\Phi_L(\omega,\eta)$ would also be known, yielding a complete solution of the problem. Let us clarify the meaning of this unknown function

$$Q(x) := \frac{F_L(0,x)}{1 - L(x)}.$$

We can write

$$Q(x) = \frac{F_L(0,x)}{F_L(\infty,x)}\frac{F_L(\infty,x)}{1 - L(x)},$$

where the second term

$$\frac{F_L(\infty,x)}{1 - L(x)} = \Psi_L(0,x) \equiv \mathrm{const}$$

by (11). (In fact, $\frac{F_L(\infty,x)}{1-L(x)}$ equals $\mathrm{P}(X = L)$ times the density of the excess life distribution of a low-speed period, divided by $1 - L(x)$, so this constant equals $\delta/(\delta + 1/\nu)$ times $1/\delta$, which is $1/(\delta + 1/\nu)$.) In view of (7),

$$\frac{F_L(\infty,x)}{1 - L(x)} \equiv F_L(\infty,0+) \equiv \nu F_H(\infty) = \frac{\nu}{1 + \nu\delta}.$$

10

The term

$$\frac{F_L(0,x)}{F_L(\infty,x)} = P_L(0 \mid x)$$

equals the probability $P(V = 0 \mid L_{\text{past}} = x, X = L)$ that the buffer is empty at a given moment in time, under condition that the system has been in state L for a time x.

We now express $Q(x)$ into the LST $\Phi_H(\cdot)$. Consider the M/G/1 queue with one service speed r. The LST of the conditional probability $P(0 \mid x,v)$ that the buffer is empty at time x if the initial workload is v, is well known (see (4.92), page 260 in [8], where the speed is assumed to be 1):

$$\int_{x=0}^{\infty} \exp\{-sx\}P(0 \mid x,v)\mathrm{d}x = \frac{\exp\{-r^{-1}(s + \lambda(1 - G(s)))v\}}{s + \lambda(1 - G(s))}, \qquad (25)$$

where $G(s)$ is the LST of the busy period of this system.

In our system the following then holds (use PASTA to equate the steady-state workload distribution during high-speed periods with the workload distribution at the end of such a period):

$$\int_{x=0}^{\infty} \exp\{-sx\}P_L(0 \mid x)\mathrm{d}x$$

$$= \frac{1}{F_H(\infty)} \int_{v=0}^{\infty} \int_{x=0}^{\infty} \exp\{-sx\}P_L(0 \mid x,v)\mathrm{d}x\mathrm{d}F_H(v)$$

$$= \frac{1}{F_H(\infty)} \int_{v=0}^{\infty} \frac{\exp\{-s_L^{-1}(s + \lambda(1 - G_L(s)))v\}}{s + \lambda(1 - G(s))}\mathrm{d}F_H(v)$$

$$= \frac{\Phi_H\big(s_L^{-1}(s + \lambda(1 - G_L(s)))\big)}{F_H(\infty)(s + \lambda(1 - G_L(s)))},$$

where $G_L(s)$ is the busy period of the ordinary $M/G/1$ system that always operates at speed s_L.

So we may introduce

$$q(s) := \int_{x=0}^{\infty} \exp\{-sx\}\frac{F_L(0,x)}{1 - L(x)}\mathrm{d}x = \frac{\nu\Phi_H\big(s_L^{-1}(s + \lambda(1 - G_L(s)))\big)}{s + \lambda(1 - G_L(s))}, \quad \mathrm{Re}\,s > 0.$$

$$(26)$$

11

Then

$$
\begin{aligned}
Q(x) &= \frac{F_L(0,x)}{1 - L(x)} = \frac{1}{2\pi i} \int_{\varepsilon - i\infty}^{\varepsilon + i\infty} \exp\{sx\} q(s) \mathrm{d}s \\
&= \frac{1}{2\pi i} \int_{\varepsilon - i\infty}^{\varepsilon + i\infty} \exp\{sx\} \frac{\nu \Phi_H\left(s_L^{-1}(s + \lambda(1 - G_L(s)))\right)}{s + \lambda(1 - G_L(s))} \mathrm{d}s, \quad \varepsilon > 0. (27)
\end{aligned}
$$

Thus the unknown function $Q(x)$ has been expressed in $\Phi_H(\cdot)$. In general, we don't know how to determine $Q(x)$. However, in the case of low-speed period distributions with a rational LST, $Q(x)$ can be determined; this is demonstrated in the next section. And in case the low-speed periods have a distribution that is regularly varying at infinity, one can determine the tail behaviour of the workload distribution; this is the subject of Section 4.

3 Rational LST of the low-speed period distribution

In this section we derive the solution of the main equations for the case in which the LST of the distribution of the low-speed periods is rational; viz., it has the form $\delta(s) = \delta_1(s)/\delta_2(s)$, where $\delta_1(s)$, $\delta_2(s)$ are relatively prime polynomials, the degree of $\delta_2(s)$ being higher than that of $\delta_1(s)$. Without loss of generality we can write

$$
\delta_2(s) = \prod_{j=1}^{n} (s - s_j)^{m_j},
$$

where $s_1, \ldots, s_n$ are different from each other, $m_j \in \{1, 2, \ldots\}$ and $\operatorname{Re} s_j < 0$, $j = 1, 2, \ldots, n$ (because $\delta(s)$ is analytic for $\operatorname{Re} s \geq 0$). The equation (18) holds for all ω with $\operatorname{Re} \omega \geq 0$, except (cf. (17)) those ω for which

$$
f(\omega) := s_L \omega - \lambda(1 - \beta(\omega)) = -s_j, \qquad j = 1, 2, \ldots, n.
$$

For all s with $\operatorname{Re} s > 0$ there exists exactly one ω with $\operatorname{Re} \omega \geq 0$ such that $f(\omega) = s$, see e.g. [8], p. 548. So (18) holds in the whole positive half-plane except for $\omega_1, \ldots, \omega_n$ such that $f(\omega_j) = -s_j$. It follows from (18) that

$$
\Phi_H(\omega) = \frac{-s_H F_H(0) - s_L R(\omega)}{\omega^{-1} k(\omega)}, \tag{28}
$$

12

where $R(\omega)$ is given in (19) and where, for Re $\omega \geq 0$,

$$k(\omega) := \nu + \lambda(1 - \beta(\omega)) - s_H\omega - \nu\delta\{-f(\omega)\}.$$

Let us compute the function $R(\omega)$. Using (27),

$$
\begin{aligned}
R(\omega) &= \frac{1}{2\pi i} \int\limits_{\eta=0}^{\infty} \exp\{f(\omega)\eta\} \int\limits_{x=0}^{\eta} \exp\{-f(\omega)x\} \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} \exp\{sx\}q(s)\,ds\,dx\,dL(\eta) \\
&= \frac{1}{2\pi i} \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} q(s) \int\limits_{\eta=0}^{\infty} \exp\{f(\omega)\eta\} \int\limits_{x=0}^{\eta} \exp\{(s - f(\omega))x\}\,dx\,dL(\eta)\,ds \\
&= \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} q(s) \int\limits_{\eta=0}^{\infty} \frac{\exp\{f(\omega)\eta\}}{s - f(\omega)} \Big(\exp\{(s - f(\omega))\eta\} - 1 \Big)\,dL(\eta)\,ds \\
&= \frac{1}{2\pi i} \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} \frac{q(s)(\delta\{-s\} - \delta\{-f(\omega)\})}{s - f(\omega)}\,ds.
\end{aligned}
$$

The integrand of the last integral has n poles in the right-half plane, these are zeros of $\delta_2\{-s\}$: $-s_1, -s_2, \ldots, -s_n$. Consider the semi-circle with centre in ε and radius R in the right-half plane. Choose R so large that all n above-mentioned poles are inside the semi-circle. Then the integral along the line segment from $\varepsilon - iR$ to $\varepsilon + iR$ and then along the semi-circle back to $\varepsilon - iR$ equals minus the sum of the residues of the integrand at those poles. Since the integral along the semi-circle disappears when $R \to \infty$ (this is directly seen for example from the representation (26) of q(s)), we have:

$$R(\omega) = \sum_{j=1}^{n} \frac{(-1)^{m_j-1}}{(m_j - 1)!} \frac{d^{m_j-1}}{da^{m_j-1}} \Big\{ \frac{q(a)}{a - f(\omega)} \frac{\delta_1\{-a\}}{\prod_{i\neq j}(-a - s_i)^{m_i}} \Big\}\Big|_{a=-s_j},$$

and via (28):

$$\Phi_H(\omega) \hspace{10cm} (29)$$

$$= \frac{-s_H F_H(0) - s_L \sum_{j=1}^{n} \frac{(-1)^{m_j-1}}{(m_j - 1)!} \frac{d^{m_j-1}}{da^{m_j-1}} \Big\{ \frac{q(a)}{a - f(\omega)} \frac{\delta_1\{-a\}}{\prod\limits_{i\neq j}(-a - s_i)^{m_i}} \Big\}\Big|_{a=-s_j}}{\omega^{-1}k(\omega)}.$$

13

The numerator of (29) has $\sum_{j=1}^{n} m_j + 1$ unknown constants: $F_H(0)$ and the $\sum_{j=1}^{n} m_j$ constants relating to the ith derivative of $q(s)$ at $s = s_j$, $i = 0, \ldots, m_j - 1$, $j = 1, \ldots, n$. So we observe that, due to (26), $\Phi_H(\omega)$ is completely determined by $F_H(0)$ and its values in a finite number of points (and, whenever the poles of $\delta\{s\}$ are of multiplicity more than 1, the values of its derivatives of the appropriate order at these points).

The following lemma shows the way to find these values.

Lemma 3.1 *Let* $\lambda\beta - s_H + \nu\delta(\lambda\beta - s_L) < 0$. *Then the function* $g(\omega) = \omega^{-1}k(\omega)$ *has exactly* $\sum_{j=1}^{n} m_j$ *zeros at* $\mathrm{Re}\,\omega > 0$.

Proof. First, let us note that the number of zeros of the function $g(\omega)$ is the same for all low speeds $s'_L \in [s_L, s_H]$. In fact, otherwise (since zeros of $g(\omega)$ depend continuously on the parameters) for some $s'_L \in [s_L, s_H]$ there would be a zero ω with $\mathrm{Re}\,\omega = 0$. But this is impossible, because of

$$g(0) = \lambda\beta - s_H + \nu\delta(\lambda\beta - s'_L) < 0 \quad \text{for all } s'_L \in [s_L, s_H], \qquad (30)$$

and, for real ϕ,

$$|g(i\phi)| \geq \phi^{-1}(|\nu + \lambda(1 - \beta(i\phi)) - s_H i\phi| - \nu|\delta\{\lambda(1 - \beta(\omega)) - s_L\}|) > 0. \quad (31)$$

So, it suffices to prove that $g(\omega)$ has exactly $\sum_{j=1}^{n} m_j$ zeros, when $s_L = s_H$. In the latter case we have

$$f(\omega) = s_L\omega - \lambda(1 - \beta(\omega)) = s_H\omega - \lambda(1 - \beta(\omega)).$$

The function $(\nu - f(\omega))\delta_2\{-f(\omega)\}$ has $\sum_{j=1}^{n} m_j + 1$ zeros at $\mathrm{Re}\,\omega \geq 0$. In fact, the zeros of this function are at points $f(\omega) = \nu$ or $f(\omega) = -s_j$. Since $\mathrm{Re}\,\nu > 0$, $\mathrm{Re}(-s_j) > 0$, again by [8], each of these last equations has exactly one zero of multiplicity 1 at $\mathrm{Re}\,\omega \geq 0$. Then by Rouché's theorem the function $(\nu - f(\omega))\delta_2\{-f(\omega)\} - \nu\delta_1\{-f(\omega)\}$ has $\sum_{j=1}^{n} m_j + 1$ zeros at $\mathrm{Re}\,\omega \geq 0$. Then $k(\omega) = (\nu - f(\omega)) - \nu\delta\{-f(\omega)\}$ has $\sum_{j=1}^{n} m_j + 1$ zeros at $\mathrm{Re}\,\omega \geq 0$. Note that one of these zeros is $\omega = 0$ and

$$k(\omega) = (\lambda\beta - s_H)(1 + \nu\delta)\omega(1 + 0(1)) \quad \text{as } \omega \to 0.$$

Hence $g(\omega) = \omega^{-1}k(\omega)$ has $\sum_{j=1}^{n} m_j$ zeros at $\mathrm{Re}\,\omega \geq 0$. Moreover, it follows from (30) and (31) that none of these zeros is situated on $\mathrm{Re}\,\omega = 0$. The lemma is proved.

14

Let us point out now that since the left-hand side of (29) is analytic in the *whole* half-plane $\mathrm{Re}\,\omega > 0$ and never equals zero, the right-hand side must satisfy the same properties. Then the zeros of the denominator of the right-hand side should also be the zeros of the numerator of the right-hand side, and of the same multiplicity. Substituting them into the numerator we obtain $\sum_{j=1}^{n} m_j$ equations. We get one more equation if we substitute $\omega = 0$ in (29) and take into account (30) and the fact that $\Phi_H(0) = 1/(1 + \nu\delta)$. Thus we have a system of $\sum_{j=1}^{n} m_j + 1$ linear equations for as many unknown constants, which has one solution. Substituting these constants again into (29), we find $\Phi_H(\omega)$.

Note that $\Phi_L(\omega, \eta)$ now may be found from (15), if we substitute instead of $\frac{F_L(0,x)}{1-L(x)}$ its representation (27):

$$
\Phi_L(\omega, \eta) = (1 - L(\eta)) \exp\{f(\omega)\eta\}\Big[\nu\Phi_H(\omega)
$$
$$
- \frac{\nu s_L \omega}{2\pi i} \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} \frac{\Phi_H(s_L^{-1}(s + \lambda(1 - G_L(s))))(\exp\{\eta(s - f(\omega))\} - 1)}{(s + \lambda(1 - G_L(s)))(s - f(\omega))}\,\mathrm{d}s\Big].
$$

Moreover, the integral $\int_{\eta=0}^{\infty} \Phi_L(\omega, \eta)\mathrm{d}\eta$ may be expressed in already known functions and constants:

$$
\int\limits_{\eta=0}^{\infty} \Phi_L(\omega, \eta)\mathrm{d}\eta = \frac{\delta\{-f(\omega)\} - 1}{f(\omega)} \nu\Phi_H(\omega)
$$
$$
- \frac{s_L \omega}{2\pi i f(\omega)} \int\limits_{\varepsilon-i\infty}^{\varepsilon+i\infty} \frac{q(s)(\delta\{-s\}f(\omega) - \delta\{-f(\omega)\}s + s - f(\omega))}{s(s - f(\omega))}\,\mathrm{d}s
$$
$$
= \frac{\delta\{-f(\omega)\} - 1}{f(\omega)} \nu\Phi_H(\omega)
$$
$$
- s_L \omega \sum_{j=1}^{n} \frac{(-1)^{m_j-1}}{(m_j - 1)!} \frac{d^{m_j-1}}{da^{m_j-1}} \Big\{ \frac{q(a)}{a(a - f(\omega))} \frac{\delta_1\{-a\}}{\prod_{i\neq j}(-a - s_i)^{m_i}} \Big\}\Big|_{a=-s_j}.
$$

Remark 3.1. As already said in the Introduction, the papers [11, 12, 23] are considering a similar model to the one in the present section. The model of [23] is more general in the sense that it allows several server states; it is less general in the sense that it assumes an underlying Markov modulated

15

process. The papers [11, 12] study a very general $M/G/1$ model in which the arrival rate, service speed and allocated service requests depend on the state of a semi-Markov process. In the case of limited PH-distributions for residence times in states of that semi-Markov process, the way of obtaining the workload LST via the computation of zeros of the determinant is indicated. The case that we have studied in the present section is more restricted, but therefore allows a very detailed and complete analysis.

3.1   Example: exponential low-speed periods

We will show how this procedure works on a simple example. Let the distribution of the low-speed periods be exponential with parameter ζ. Then the LST $\delta(s) = \frac{\zeta}{\zeta+s}$ has one pole $s = -\zeta$ and $\delta = 1/\zeta$. The equation (29) reduces to

$$\Phi_H(\omega) = \frac{-s_H F_H(0) - \dfrac{s_L q(\zeta)\zeta}{\zeta - f(\omega)}}{\omega^{-1}k(\omega)}. \tag{32}$$

By Lemma 3.1 the function $g(\omega) = \omega^{-1}k(\omega)$ has one zero ω_1 of multiplicity 1 at $\mathrm{Re}\,\omega > 0$. Substituting it into (32), we get

$$-s_H F_H(0) - \frac{s_L q(\zeta)\zeta}{\zeta - f(\omega_1)} = 0. \tag{33}$$

Now, substitution of $\omega = 0$ into (32) gives

$$-s_H F_H(0) - s_L q(\zeta) = \frac{\lambda\beta(\zeta + \nu) - \zeta s_H - \nu s_L}{\nu + \zeta}. \tag{34}$$

It follows from (33) and (34) that

$$q(\zeta) = \frac{(\lambda\beta(\nu + \zeta) - \zeta s_H - \nu s_L)(\zeta - f(\omega_1))}{s_L(\nu + \zeta)f(\omega_1)}$$

$$F_H(0) = \frac{-\zeta(\lambda\beta(\zeta + \nu) - s_H\zeta - s_L\nu)}{s_H(\nu + \zeta)f(\omega_1)}.$$

Finally, substituting the constants $F_H(0)$ and $q(\zeta)$ just found into (32), we obtain the workload transforms for both the high- and low-speed periods:

$$\Phi_H(\omega) = \frac{1/\delta(\lambda\beta(\nu + 1/\delta) - s_H 1/\delta - s_L\nu)}{(s_L\omega_1 - \lambda(1 - \beta(\omega_1)))(\nu + 1/\delta)}$$

16

$$\times \frac{\omega(\lambda(\beta(\omega_1) - \beta(\omega)) + s_L(\omega_1 - \omega))}{(\nu + \lambda(1 - \beta(\omega)) - s_H\omega)(1/\delta + \lambda(1 - \beta(\omega)) - s_L\omega) - \nu/\delta},$$

$$\Phi_L(\omega) = \int_{\eta=0}^{\infty} \Phi_L(\omega, \eta)\mathrm{d}\eta = \frac{\nu(\lambda\beta(\nu + 1/\delta) - s_H/\delta - s_L\nu)}{(s_H\omega_1 - \lambda(1 - \beta(\omega_1)))(\nu + 1/\delta)}$$

$$\times \frac{\omega(\lambda(\beta(\omega_1) - \beta(\omega)) + s_H(\omega_1 - \omega))}{(\nu + \lambda(1 - \beta(\omega)) - s_H\omega)(1/\delta + \lambda(1 - \beta(\omega)) - s_L\omega) - \nu/\delta}.$$

4 Buffer content asymptotics

In this section we discuss the asymptotic behaviour of the workload distribution under the assumption that $B(t)$ and/or $L(t)$ has a regularly varying tail. We distinguish between the cases $\lambda\beta > s_L$ and $\lambda\beta < s_L$. The more delicate case $\lambda\beta = s_L$ is not discussed. We restrict ourselves to the theoretically interesting case of regular variation with an index between 1 and 2; however, larger values can also be handled.

Theorem 4.1 *Let*

$$\begin{aligned}
1 - B(t) &= [C_B + o(1)]t^{-\nu_1}l_1(t), & t \to \infty, \\
1 - L(t) &= [C_L + o(1)]t^{-\nu_2}l_2(t), & t \to \infty,
\end{aligned}$$

where $1 < \nu_1, \nu_2 < 2$, $l_1(t), l_2(t)$ *are slowly varying functions.*
(i) Let $\lambda\beta > s_L$. *Then*

$$\frac{1}{1 + \nu\delta} - F_H(v) = [D_H + o(1)]v^{1-\min(\nu_1,\nu_2)}l_H(v), \tag{35}$$

$$\frac{\nu(1 - L(\eta))}{1 + \nu\delta} - F_L(v, \eta) = \nu(1 - L(\eta))[D_H + o(1)]v^{1-\min(\nu_1,\nu_2)}l_H(v) \tag{36}$$

$$\frac{\nu\delta}{1 + \nu\delta} - \int_{\eta=0}^{\infty} F_L(v, \eta)\mathrm{d}\eta = [D_L + o(1)]v^{1-\min(\nu_1,\nu_2)}l_L(v), \tag{37}$$

17

as $v \to \infty$, where $l_H(v)$, $l_L(v)$ are slowly varying functions and

$$
D_H = \begin{cases}
\dfrac{\lambda C_B}{(\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L)} \dfrac{1}{1 - \nu_1}, & \text{if } \nu_1 < \nu_2, \\[3mm]
\dfrac{\nu(\lambda\beta - s_L)^{\nu_2} C_L}{(1 + \nu\delta)((\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L))} \dfrac{1}{1 - \nu_2}, & \text{if } \nu_2 < \nu_1, \\[3mm]
\dfrac{\lambda C_B}{(\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L)} \dfrac{1}{1 - \nu_1} \\[3mm]
\quad + \dfrac{\nu(\lambda\beta - s_L)^{\nu_2} C_L}{(1 + \nu\delta)((\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L))} \dfrac{1}{1 - \nu_2}, & \text{if } \nu_1 = \nu_2;
\end{cases}
$$

$$
D_L = \begin{cases}
\dfrac{\lambda\nu\delta C_B}{(\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L)} \dfrac{1}{1 - \nu_1}, & \text{if } \nu_1 < \nu_2, \\[3mm]
\dfrac{\nu(\lambda\beta - s_L)^{\nu_2 - 1}(s_H - \lambda\beta) C_L}{(1 + \nu\delta)((\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L))} \dfrac{1}{1 - \nu_2}, & \text{if } \nu_2 < \nu_1, \\[3mm]
\dfrac{\lambda\nu\delta C_B}{(\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L)} \dfrac{1}{1 - \nu_1} \\[3mm]
\quad + \dfrac{\nu(\lambda\beta - s_L)^{\nu_2 - 1}(s_H - \lambda\beta) C_L}{(1 + \nu\delta)((\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L))} \dfrac{1}{1 - \nu_2}, & \text{if } \nu_1 = \nu_2.
\end{cases}
$$

(ii) *Let $\lambda\beta < s_L$. Then*

$$
\frac{1}{1 + \nu\delta} - F_H(v) = [D_H + o(1)]v^{1 - \nu_1} l_H(v), \tag{38}
$$

$$
\frac{\nu(1 - L(\eta))}{1 + \nu\delta} - F_L(v, \eta) = \nu(1 - L(\eta))[D_H + o(1)]v^{1 - \nu_1} l_H(v), \tag{39}
$$

$$
\frac{\nu\delta}{1 + \nu\delta} - \int_{\eta = 0}^{\infty} F_L(v, \eta)\mathrm{d}\eta, = [D_L + o(1)]v^{1 - \nu_1} l_L(v), \tag{40}
$$

as $v \to \infty$, where $l_H(v)$, $l_L(v)$ are slowly varying functions and

$$
D_H = \frac{\lambda C_B}{(\lambda\beta - s_H) + \delta\nu(\lambda\beta - s_L)} \frac{1}{1 - \nu_1},
$$

18

$$D_L = \frac{\lambda \nu \delta C_B}{(\lambda \beta - s_H) + \delta \nu (\lambda \beta - s_L)} \frac{1}{1 - \nu_1}.$$

Remark 4.1. The theorem indicates that (i) if $\lambda \beta > s_L$, then the tail of the workload distribution is determined by the heaviest of the tails of the service request and low-speed period distributions, and (ii) if $\lambda \beta < s_L$ then the tail of the workload distribution is only determined by the tail of the service request distribution. Moreover, the sum of the workload tails in (38) and (40) behaves as if the system is always operating with the speed $(s_H + \nu \delta s_L)/(1 + \nu \delta)$ which is the average speed of the server. In fact, consider a stable M/G/1 system with the required service time distribution $B(\cdot)$ and one constant speed of service r. Let us denote the distribution function of its workload by $W^r(v)$. Then by [7], the assumption

$$1 - B(t) = [C_B + o(1)]t^{-\nu} l(t), \quad t \to \infty,$$

with $C_B > 0$, $1 < \nu < 2$, $l(t)$ slowly varying, implies that

$$1 - W^r(v) = [D_r + o(1)]v^{1-\nu} l(v), \quad v \to \infty, \tag{41}$$

where

$$D_r = \frac{\lambda C_B}{r - \lambda \beta} \frac{1}{\nu - 1}.$$

Let us compare our two-speeds system with the M/G/1, having the same arrival intensity and required service time distribution, but one constant service speed s_H. Indeed, this system is always stable, whenever our two-speeds system is stable. In addition to this, for any given realisation of arrivals and services, a workload of this one-speed system is never greater than the one of our two-speeds system. Thus, by simple arguments of stochastic ordering, we get the following lower bound:

$$1 - \frac{F_H(v)}{F_H(\infty)} \geq 1 - W^{s_H}(v) = [D_{s_H} + o(1)]v^{1-\nu_1} l_1(v), \quad v \to \infty. \tag{42}$$

Moreover, in case (ii), the low speed s_L is sufficient for the M/G/1 queue to be stable. Comparing in the same way our system with the $M/G/1$, having all the same characteristics but a constant service speed s_L, we get the upper bound:

$$1 - \frac{F_H(v)}{F_H(\infty)} \leq 1 - W^{s_L}(v) = [D_{s_L} + o(1)]v^{1-\nu_1} l_2(v), \quad v \to \infty. \tag{43}$$

19

Thus, in case (ii), we have bounds for the tail of the distribution function $F_H(v)$ from both sides, and both of them are of order $v^{1-\nu_1}$. Similar bounds are, of course, true for $F_L(v, \eta)$ and $\int_{\eta=0}^{\infty} F_L(v, \eta) d\eta$.

Remark 4.2. It is well known [7] that (41) can also be written as

$$1 - W^r(v) \sim \frac{\lambda\beta}{r - \lambda\beta} P(B^{\text{res}} > v) = \frac{\lambda\beta}{r - \lambda\beta} \int_v^{\infty} \frac{1 - B(u)}{\beta} du$$

($f(v) \sim g(v)$ denotes $f(v)/g(v) \to 1$ for $v \to \infty$). Then (38) and (40) can be rewritten in a similar way; e.g. (38) becomes

$$\frac{1}{1 + \nu\delta} - F_H(v) \sim \frac{\lambda\beta}{(s_H - \lambda\beta) + \nu\delta(s_L - \lambda\beta)} P(B^{\text{res}} > v), \quad v \to \infty.$$

Remark 4.3 In [5] the impact of regular variation in the coupled processors model is investigated. The service times are assumed to be regularly varying with the indexes ν_1 and ν_2 for the first and the second processor respectively. The second processor always operates at one constant speed, while the first processor operates at a lower speed when the second processor is busy and at a higher speed when the second processor is free. The tail of the workload distribution is found. It is shown to be regularly varying with the index ν_1 if the low speed is larger than the offered traffic load and with the index $\min(\nu_1, \nu_2)$ if the low speed is smaller than that. We would like to remark that this result is a particular case of Theorem 4.1 just stated. In fact, the high-speed periods of the first processor are exponential and the low-speed periods correspond to the busy periods of the second processor, whence by [20] they are regularly varying with the index ν_2. The asymptotic constants found in [5] coincide with ours in this particular case. But our result is more general, since in our model the low speed periods do not necessarily correspond to the busy periods of another queueing system.

Proof of Theorem 4.1. The proof will essentially rely on Lemma 6.1 from the Appendix. This lemma implies

$$
\begin{aligned}
\beta(\omega) &= 1 - \beta\omega - [C_B + o(1)]\Gamma(1 - \nu_1)\omega^{\nu_1} l_1(1/\omega), \quad \omega \downarrow 0 & (44)\\
\delta(\omega) &= 1 - \delta\omega - [C_L + o(1)]\Gamma(1 - \nu_2)\omega^{\nu_2} l_2(1/\omega), \quad \omega \downarrow 0. & (45)
\end{aligned}
$$

Case (i): $\lambda\beta > s_L$.
We shall find the asymptotics of $\Phi_H(\omega)$ when $\omega \downarrow 0$ using equation (18),

20

which holds for all ω sufficiently small with $\mathrm{Re}\,\omega > 0$ in this case. Using (44) and (45), the asymptotics of the second term in the left-hand side of (18) is seen to be:

$$
\begin{aligned}
k(\omega) &= \nu + \lambda(1 - \beta(\omega)) - s_H\omega - \nu\delta\{\lambda(1 - \beta(\omega)) - s_L\omega\} \\
&= \quad [(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)]\omega \\
&\quad + [\lambda C_B(1 + \nu\delta)\Gamma(1 - \nu_1)\omega^{\nu_1}l_1(1/\omega) \\
&\quad\quad + \nu C_L(\lambda\beta - s_L)^{\nu_2}\Gamma(1 - \nu_2)\omega^{\nu_2}l_2(1/\omega)](1 + o(1)), \quad \omega \downarrow 0.
\end{aligned}
\tag{46}
$$

Let us consider the right-hand side of (18) and prove that its asymptotics is $C\omega + O(\omega^2)$ as $\omega \downarrow 0$, where

$$
C = -s_H P(V = 0, x = H) - s_L P(V = 0, x = L).
$$

This is equivalent with the fact that (remember that $f(\omega) = s_L\omega - \lambda(1 - \beta(\omega))$):

$$
\begin{aligned}
R(\omega) &= \int_{\eta=0}^{\infty} \exp\{f(\omega)\eta\} \int_{x=0}^{\eta} \exp\{-f(\omega)x\}\frac{F_L(0, x)}{1 - L(x)}\mathrm{d}x\,\mathrm{d}L(\eta) \\
&= P(V = 0, x = L) + O(\omega).
\end{aligned}
\tag{47}
$$

Let us show this. Since $\lambda\beta > s_L$, for all sufficiently small s with $\mathrm{Re}\,s < 0$ there exist $\omega_1(s)$ and $\omega_2(s)$ at $\mathrm{Re}\,\omega \geq 0$ such that $f(\omega_1(s)) \equiv f(\omega_2(s)) \equiv s$ (cf. [9], p. 297). Moreover $\omega_1(s) \to 0$ and $\omega_2(s) \to \omega_2(0) > 0$ as $s \to 0$. The function $f(\omega)$ being analytic at $\omega_2(0) > 0$, by the implicit function theorem

$$
\omega_2(s) = \omega_2(0) + O(s), \qquad \text{as } s \uparrow 0.
\tag{48}
$$

Let us define $\varepsilon(\omega) = \omega_2(f(\omega))$ for sufficiently small ω with $\mathrm{Re}\,\omega > 0$. Due to the expansion $f(\omega) = (s_L - \lambda\beta)\omega(1 + o(1))$ as $\omega \downarrow 0$, we have

$$
\varepsilon(\omega) = \varepsilon(0) + O(\omega), \qquad \text{as } \omega \downarrow 0.
\tag{49}
$$

where $\varepsilon(0) = \omega_2(0) > 0$. The definition of $\varepsilon(\omega)$ provides the equality $f(\omega) \equiv f(\varepsilon(\omega))$. Then, using (18):

$$
R(\omega) \equiv R(\varepsilon(\omega)) = \frac{-\Phi_H(\varepsilon(\omega))k(\varepsilon(\omega)) - s_H\varepsilon(\omega)F_H(0)}{s_L\varepsilon(\omega)},
\tag{50}
$$

21

for all sufficiently small ω with $\text{Re}\,\omega > 0$. Finally (49), (50) together with the analyticity of the functions $\Phi_H(\omega)$ and $k(\omega)$ at $\varepsilon(0) > 0$ imply the required expansion

$$R(\omega) = R(0) + O(\omega), \qquad \text{as } \omega \downarrow 0.$$

Moreover

$$R(0) = \int\limits_{\eta=0}^{\infty} \int\limits_{x=0}^{\eta} \frac{F_L(0,x)}{1-L(x)}\,\mathrm{d}x\,\mathrm{d}L(\eta) = \int\limits_{x=0}^{\infty} F_L(0,x)\,\mathrm{d}x = P(V=0, x=L),$$

and (47) is proved.

Now, substituting (46) and (47) into (18), we have

$$\Phi_H(\omega)$$

$$= \frac{-[P(v=0,x=H)s_H + P(v=0,x=L)s_L]\omega(1+O(\omega))}{[\lambda\beta - s_H + \nu\delta(\lambda\beta - s_L)]\omega + [K_1\omega^{\nu_1}l_1(1/\omega) + K_2\omega^{\nu_2}l_2(1/\omega)](1+o(1))}$$

$$= \frac{-[P(v=0,x=H)s_H + P(v=0,x=L)s_L]}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)}(1+O(\omega))$$

$$\times \left[1 - \frac{K_1\omega^{\nu_1-1}l_1(1/\omega)(1+o(1))}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)} - \frac{K_2\omega^{\nu_2-1}l_2(1/\omega)(1+o(1))}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)}\right],$$

where

$$\begin{aligned}
K_1 &= \lambda C_B(1+\nu\delta)\Gamma(1-\nu_1), \\
K_2 &= \nu C_L(\lambda\beta - s_L)^{\nu_2}\Gamma(1-\nu_2).
\end{aligned}$$

Moreover, substituting $\omega = 0$ into (18), we obtain the constant

$$\frac{-P(v=0,x=H)s_H - P(v=0,x=L)s_L}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)} = \Phi_H(0) = F_H(\infty) = \frac{1}{1+\nu\delta}. \tag{51}$$

Finally, Lemma 6.1 in the appendix applies and we get (35).

Case (ii): $\lambda\beta < s_L$.
We start with the equation (24), which holds for all ω such that $\text{Re}\,(s_L\omega - \lambda(1 - \beta(\omega))) > 0$ and consequently in this case for all sufficiently small ω

with positive real part. Adding the term $\nu\Phi_H(\omega)(-1 - \delta f(\omega))$ to both sides of (24) and using the representation (22), we get:

$$\Phi_H(\omega)[-\nu\delta f(\omega) + \lambda(1 - \beta(\omega)) - s_H\omega] \tag{52}$$

$$= -s_H\omega F_H(0)$$

$$+ s_L\omega \int_{x=0}^{\infty} \exp\{-f(\omega)x\}\Big(\int_{\eta=0}^{x} \exp\{f(\omega)\eta\}dL(\eta) - 1 - f(\omega)\delta\Big)$$

$$\times \frac{F_L(0, x)}{1 - L(x)}dx.$$

By (44) the asymptotics of the second term in the left-hand side of (52) is the following:

$$-\nu\delta f(\omega) + \lambda(1 - \beta(\omega)) - s_H\omega \tag{53}$$

$$= [\lambda\beta - s_H + \nu\delta(\lambda\beta - s_L)]\omega + \lambda(1 + \nu\delta)C_B\Gamma(1 - \nu_1)\omega^{\nu_1}l_1(1/\omega)(1 + o(1))$$

as $\omega \downarrow 0$.

Let us prove that the right-hand side of (52) is

$$\omega\Big[-s_H P(V = 0, X = H) - s_L P(V = 0, X = L) + O(\omega^{\nu_1 - 1 + \varepsilon_0})\Big] \tag{54}$$

for some $\varepsilon_0 > 0$ when $\omega \downarrow 0$. As shown in p. 548 of [8], for all s with $\mathrm{Re}\, s \geq 0$ there exists exactly one $\omega(s)$ with $\mathrm{Re}\,\omega(s) \geq 0$ such that $f(\omega(s)) = s$. Moreover $\omega(s)$ is regular at $\mathrm{Re}\, s > 0$ and continuous at $\mathrm{Re}\, s \geq 0$, $\omega(0) = 0$. Then the function

$$\tilde{R}(s) := \int_{x=0}^{\infty} \exp\{-sx\}\Big(\int_{\eta=0}^{x} \exp\{s\eta\}dL(\eta) - 1 - s\delta\Big)\frac{F_L(0, x)}{1 - L(x)}dx \tag{55}$$

is well defined and continuous in $\mathrm{Re}\, s > 0$. In fact, due to (52),

$$\tilde{R}(s) \equiv \tilde{R}(f(\omega(s)))$$
$$= \Phi_H(\omega(s))[-\nu\delta f(\omega(s)) + \lambda(1 - \beta(\omega(s))) - s_H\omega(s)]\omega^{-1}(s)s_L^{-1}$$
$$+ s_H s_L^{-1}F_H(0),$$

where all functions at the right-hand side are well defined and continuous. We will prove that for some $\varepsilon_0 > 0$

$$\tilde{R}(s) + P(V = 0, X = L) = O(s^{\nu_1 - 1 + \varepsilon_0}) \quad s \downarrow 0. \tag{56}$$

23

First of all, remember that

$$Q(x) = \frac{F_L(0, x)}{1 - L(x)} = \frac{\nu}{1 + \nu\delta} P_L(0 \mid x),$$

as was noticed in Section 2. Here the conditional probability $P_L(0 \mid x)$ equals the probability that at time x the buffer is empty in the M/G/1 system with the same arrival rate λ and the service time distribution $B(\cdot)$ but with one constant service speed s_L and the initial workload distribution as the one at high-speed periods in our two-speeds system. Since $\lambda\beta < s_L$, this M/G/1 system with one service speed s_L is stable. Then, as $x \to \infty$, $P_L(0 \mid x)$ tends to the stationary probability of zero workload in this system, which is $1 - \lambda\beta/s_L$. Thus we can write

$$Q(x) = \frac{\nu(s_L - \lambda\beta)}{s_L(1 + \nu\delta)} + \bar{Q}(x), \tag{57}$$

where $\bar{Q}(x) \to 0$ as $x \to \infty$. We show in Lemma 6.2 in the Appendix that for *all* $\varepsilon > 0$,

$$|\bar{Q}(x)| = o(x^{1-\nu_1+\varepsilon}), \qquad x \to \infty. \tag{58}$$

We have

$$
\begin{aligned}
&\tilde{R}(s) + P(V = 0, X = L) \\
&= \int_{x=0}^{\infty} \exp\{-sx\} \Big(\int_{\eta=0}^{x} \exp\{\eta s\} \, dL(\eta) - 1 \Big) Q(x) \, dx \\
&\quad + \int_{x=0}^{\infty} (1 - L(x)) Q(x) \, dx - s \int_{\eta=0}^{\infty} (1 - L(\eta)) \, d\eta \int_{x=0}^{\infty} \exp\{-sx\} \, Q(x) \, dx \\
&= s \int_{\eta=0}^{\infty} (1 - L(\eta)) \int_{x=0}^{\infty} \exp\{-sx\} (Q(x+\eta) - Q(x)) \, dx \, d\eta \\
&= s \int_{\eta=0}^{\infty} (1 - L(\eta)) \int_{x=0}^{\infty} \exp\{-sx\} (\bar{Q}(x+\eta) - \bar{Q}(x)) \, dx \, d\eta \\
&= I_1 + I_2 - I_3,
\end{aligned}
$$

24

where

$$
I_1 = s \int\limits_{\eta=0}^{1/|s|} (1 - L(\eta))(\exp\{s\eta\} - 1) \int\limits_{x=\eta}^{\infty} \exp\{-sx\}\bar{Q}(x)\,\mathrm{d}x\,\mathrm{d}\eta,
$$

$$
I_2 = s \int\limits_{\eta=1/|s|}^{\infty} (1 - L(\eta)) \int\limits_{x=0}^{\infty} \exp\{-sx\}(\bar{Q}(x+\eta) - \bar{Q}(x))\,\mathrm{d}x\,\mathrm{d}\eta,
$$

$$
I_3 = s \int\limits_{\eta=0}^{1/|s|} (1 - L(\eta)) \int\limits_{x=0}^{\eta} \exp\{-sx\}\bar{Q}(x)\,\mathrm{d}x.
$$

We will find the asymptotics of each of these terms I_1, I_2 and I_3 as $s \downarrow 0$. Let us fix an arbitrary $\varepsilon > 0$. To estimate I_1, we observe that due to (58) (put $sx = t$):

$$
\left| s \int\limits_{x=\eta}^{\infty} \exp\{-sx\}\bar{Q}(x)\,\mathrm{d}x \right| \leq s \int\limits_{x=0}^{\infty} \exp\{-sx\}|\bar{Q}(x)|\,\mathrm{d}x
$$

$$
\leq O(s) + s^{\nu_1 - 1 - \varepsilon} \int\limits_{t=0}^{\infty} \exp\{-t\}t^{1-\nu_1+\varepsilon}\,\mathrm{d}t = O(s^{\nu_1 - 1 - \varepsilon}),
$$

uniformly for all $\eta > 0$. Moreover, due to the assumption of the theorem on the distribution function $L(\cdot)$,

$$
\left| \int\limits_{\eta=0}^{1/s} (1 - L(\eta))(\exp\{s\eta\} - 1)\,\mathrm{d}\eta \right| \leq 2s \int\limits_{\eta=0}^{1/s} (1 - L(\eta))\eta\,\mathrm{d}\eta = O(s^{\nu_2 - 1 - \varepsilon}).
$$

Then $I_1 = O(s^{\nu_1 + \nu_2 - 2 - 2\varepsilon})$. To treat the second term I_2, we note again that due to (58)

$$
\left| s \int\limits_{x=0}^{\infty} \exp\{-sx\}(\bar{Q}(x+\eta) - \bar{Q}(x))\,\mathrm{d}x \right|
$$

$$
\leq O(s) + 2s^{\nu_1 - 1 - \varepsilon} \int\limits_{t=0}^{\infty} \exp\{-t\}t^{1-\nu_1+\varepsilon}\,\mathrm{d}t = O(s^{\nu_1 - 1 - \varepsilon}),
$$

25

uniformly for all $\eta > 0$ and that due to the assumption of the theorem

$$\int\limits_{\eta=1/s}^{\infty} (1 - L(\eta)) \, d\eta = O(s^{\nu_2 - 1 - \varepsilon}).$$

So, $I_2 = O(s^{\nu_1 + \nu_2 - 2 - 2\varepsilon})$. Finally, to estimate I_3, we have by (58) for all sufficiently large η

$$\left| (1 - L(\eta)) \int\limits_{x=0}^{\eta} \exp\{-sx\} \bar{Q}(x) \, dx \right| \leq (1 - L(\eta)) \int\limits_{x=0}^{\eta} |\bar{Q}(x)| \, dx \leq C_1 \eta^{2 - \nu_1 - \nu_2 + 2\varepsilon},$$

where $C_1 > 0$ is a constant. Then $I_3 = O(s) + O(s^{\nu_1 + \nu_2 - 2 - 2\varepsilon})$. Thus $I_1 + I_2 + I_3 = O(s) + O(s^{\nu_1 - 1 + \nu_2 - 1 - 2\varepsilon})$ for *all* $\varepsilon > 0$. Let us choose $\varepsilon < (\nu_2 - 1)/2$. Then (56) is fulfilled. Since $f(\omega) = (s_L - \lambda\beta)\omega(1 + o(1))$, as $\omega \downarrow 0$, (54) follows.

Substituting the expansions (53) and (54) into (52), we have

$$\Phi_H(\omega) \tag{59}$$
$$= \frac{-[P(V = 0, X = H)s_H + P(V = 0, X = L)s_L]}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)} (1 + O(\omega^{\nu_1 - 1 + \varepsilon_0}))$$
$$\times \left[1 - \frac{\lambda C_B(1 + \nu\delta)\Gamma(1 - \nu_1)\omega^{\nu_1 - 1} l_1(1/\omega)(1 + o(1))}{(\lambda\beta - s_H) + \nu\delta(\lambda\beta - s_L)} \right].$$

Moreover, substitution of $\omega = 0$ into (59) gives the unknown constant as in (51). Applying Lemma 6.1, we get (38).

The results (36), (37) and (39), (40) come from (15) and (23), the previous results (35) and (38) and again Lemma 6.1.

5   Conclusion

We have considered an $M/G/1$ queue with the special feature that the speed of the server alternates between two constant values s_L and $s_H > s_L$. For the case that the distribution of the low-speed periods has a rational Laplace-Stieltjes transform, we have obtained the joint distribution of the buffer content and the state of the server speed. For the case that the distribution

of the low-speed periods and/or the service request distribution is regularly varying at infinity, we have obtained explicit asymptotics for the tail of the buffer content distribution. The two cases in which the offered traffic load is smaller respectively larger than the low service speed were shown to result in completely different asymptotics. When it is smaller, the tail of the workload distribution is only determined by the tail of the service request distribution. The case in which the offered traffic load equals the low service speed is more delicate; we leave its analysis as an open problem.

Another interesting problem for future research is the generalization of the present model to one with $K \geq 2$ different service speeds.

Finally we would like to observe a relation between the present model and an $M/G/1$ queue with one service speed but with the feature that the arrival intensity alternates between two constant values λ_L and $\lambda_H > \lambda_L$. For the ordinary $M/G/1$ queue, it is clear that multiplication of the arrival rate by a factor c has exactly the same effect on the steady-state workload distribution as the division of the service speed by c. In fact, given the same initial buffer content, the workload distribution of the former model at time t coincides with the one of the latter model at time ct for any $t \geq 0$. In the model of the present paper, the service speed during the H-periods is multiplied by a factor s_H/s_L compared to the service speed during the L-periods. Instead, we could also have divided the arrival rate by s_H/s_L during the H-periods, and simultaneously have multiplied the mean length $1/\nu$ of the H-periods by that factor s_H/s_L. By taking appropriate weight factors w.r.t. the L- and H-periods, we could then determine the workload distribution in the model with two arrival rates from that in the model with two service speeds – and vice versa.

Acknowledgement

6 Appendix

Definition 6.1 *A measurable positive function $f(t)$ is called regularly varying of index ν if for all $x > 0$*

$$\frac{f(xt)}{f(t)} \to x^\nu, \qquad as \ t \to \infty.$$

Definition 6.2 *A measurable positive function $l(t)$ is called slowly varying if it is regularly varying of index $\nu = 0$.*

We also use without mention the following fact: for all slowly varying function $l(t)$ and *all $\varepsilon > 0$*

$$t^{-\varepsilon}l(t) \to 0, \quad t^\varepsilon l(t) \to \infty, \quad as \ t \to \infty.$$

Lemma 6.1 *Let Z be a non-negative random variable with LST $\Phi(\omega)$, $l(t)$ a slowly varying function, $\nu \in (n, n+1)$, $(n \in \mathbf{N})$ and $C > 0$. Then the following are equivalent:*

(i) $P(Z > t) = \frac{(-1)^n}{\Gamma(1-\nu)}[C + o(1)]t^{-\nu}l(t), \quad t \to \infty;$

(ii) $E[Z^n] < \infty$ and

$$\Phi(\omega) = \sum_{j=0}^n \frac{E[Z^j](-\omega)^j}{j!} + (-1)^{n+1}[C + o(1)]\omega^\nu l(1/\omega), \quad \omega \downarrow 0.$$

Proof. See Theorem 8.1.6 of [3].

Lemma 6.2 *Let $\lambda\beta < s_L$. For all $\varepsilon > 0$*

$$|\bar{Q}(x)| = o(x^{1-\nu_1+\varepsilon}), \qquad x \to \infty. \tag{60}$$

Proof. Remember that

$$\bar{Q}(x) = \frac{\nu}{1+\nu\delta}[P(V = 0 \mid L_{\text{past}} = x, X = L) - (1 - \lambda\beta/s_L)],$$

where $P(V = 0 \mid L_{\text{past}} = x, X = L)$ is the probability that at a given moment of time the buffer is empty under condition that by this moment the system

28

has been operating at low speed for a time x. Let us introduce the interval of time $\tau_{\text{past}}(t)$ since the last change of speeds until the first moment when the buffer is empty:

$$\tau_{\text{past}}(t) := \min\{s \mid s \geq t - L_{\text{past}}(t),\ V(s) = 0\} - (t - L_{\text{past}}(t)).$$

Denote by $N(t)$ the number of customers in the system at time t. Let us also introduce the random variables $V_{\text{past}}(t)$ and $N_{\text{past}}(t)$ which are the workload and the number of customers in the buffer at the moment of the last change of speeds:

$$V_{\text{past}}(t) := V(t - L_{\text{past}}(t)); \qquad N_{\text{past}}(t) := N(t - L_{\text{past}}(t)).$$

Then by the formula of total probability

$$
\begin{aligned}
|\bar{Q}(x)| \ \leq\ & \frac{\nu}{1+\nu\delta}|P(V = 0 \mid L_{\text{past}} = x, X = L, \tau_{\text{past}} \leq x/2) - (1 - \lambda\beta/s_L)| \\
& \times P(\tau_{\text{past}} \leq x/2 \mid L_{\text{past}} = x, X = L) \\
& + \frac{2\nu}{1+\nu\delta}P(\tau_{\text{past}} > x/2 \mid L_{\text{past}} = x, X = L) \\
\leq\ & \frac{\nu}{1+\nu\delta}J_1 + \frac{2\nu}{1+\nu\delta}J_2 + \frac{2\nu}{1+\nu\delta}J_3,
\end{aligned}
$$

where

$$
\begin{aligned}
J_1 &= |P(V = 0 \mid L_{\text{past}} = x, X = L, \tau_{\text{past}} \leq x/2) - (1 - \lambda\beta/s_L)|, \\
J_2 &= P(\tau_{\text{past}} > x/2 \mid L_{\text{past}} = x, X = L, N_{\text{past}} \leq [x(s_L - \lambda\beta)/(8\beta)]), \\
J_3 &= P(N_{\text{past}} > [x(s_L - \lambda\beta)/(8\beta)] \mid L_{\text{past}} = x, X = L),
\end{aligned}
$$

($[\cdot]$ denotes an integer part).

Let us fix an arbitrary $\varepsilon > 0$. We show that J_1, J_2, J_3 are all $o(x^{1-\nu_1+\varepsilon})$. The term J_1 refers to the situation where the buffer has already become empty before $x/2$. Note that, whenever $\tau_{\text{past}} = t < x$, the probability $P(V = 0 \mid L_{\text{past}} = x, X = L, \tau_{\text{past}} = t)$ equals the probability that at a given moment of time the workload is zero under condition that by this moment the system has been working at the low speed for a time $x - t$ and that at the moment of the last change of speeds the workload was zero:

$$
\begin{aligned}
P(V = 0 \mid L_{\text{past}} &= x, X = L, \tau_{\text{past}} = t) \\
&= P(V = 0 \mid L_{\text{past}} = x - t, X = L, V_{\text{past}} = 0).
\end{aligned}
$$

29

But the latter probability equals the probability that the buffer is empty at the moment of time $x - t$ in the $M/G/1$ system having the same arrival and service time characteristics but only one constant service speed s_L and zero workload at the initial moment. The speed of convergence of this probability to the stationary probability $1 - \lambda\beta/s_L$ in this one-speed queueing system was found in [2], it is of order $O((x - t)^{1-\nu_1}l(x - t))$, as $x - t \to \infty$, where $l(x)$ is a slowly varying function. Then

$$
\begin{aligned}
|J_1| &\leq \sup_{0 \leq t \leq x/2} |P(V = 0 \mid L_{\text{past}} = x - t, X = L, V_{\text{past}} = 0) - (1 - \lambda\beta/s_L)| \\
&= o(x^{-\nu_1+1+\varepsilon}).
\end{aligned}
$$

The term J_2 refers to the situation that it takes a very long time ($> x/2$) before the buffer first empties, although the queue length at the last change of speeds is at most $[x(s_L - \lambda\beta)/(4\beta)]$. We observe that

$$
\begin{aligned}
|J_2| &\leq P(G_1 + G_2 + \cdots + G_{[x(s_L-\lambda\beta)/(4\beta)]} > x/2) \\
&\leq P(G_1 + G_2 + \cdots + G_{[x(s_L-\lambda\beta)/(4\beta)]} - [x(s_L - \lambda\beta)/(4\beta)]EG_i > x/4).
\end{aligned}
$$

Here the random variables $G_1, G_2, \ldots$ are independent and distributed as the busy period in the $M/G/1$ system with an arrival intensity λ, service time distribution $B(\cdot)$ and one constant service speed s_L. Their mean $EG_i = \beta(s_L - \lambda\beta)^{-1}$. It was shown in [20] that the G_i are regularly varying with the index ν_1. Then by the principle of large deviations for regularly varying random variables with index $\in (1, 2)$, see e.g. [15] or [10], there exists a constant $C > 0$ such that

$$
|J_2| \leq C[x(s_L - \lambda\beta)/(4\beta)]P(G_1 - EG_1 > x/4)
$$

for all sufficiently large x. Then $J_2 = o(x^{1-\nu_1+\varepsilon})$ for all $\varepsilon > 0$ as $x \to \infty$.

Finally, J_3 is the probability that at the moment of switch from the high to the low speed, the number of customers in our two-speeds system exceeds the number $[x(s_L-\lambda\beta)/(4\beta)]$. By PASTA this equals the probability that at high speed periods the number of customers exceeds this number. Let us again compare our system with the $M/G/1$ system having the same arrival and service characteristics and the same initial distribution but only one constant service speed s_L. Denote its number of customers at time t by $N^{s_L}(t)$. Then for any given realisation of arrivals and services the number of customers in

our two-speeds system is never greater than the numbre of customers in this one-speed system, i.e. $N(t) \leq N^{s_L}(t)$. As in the Remark 4.1, by elementary arguments of stochastic ordering we get:

$$
\begin{aligned}
J_3 &= P(N > [x(s_L - \lambda\beta)/(4\beta)] \mid X = H) \\
&\leq P(N^{s_L} > [x(s_L - \lambda\beta)/(4\beta)]).
\end{aligned}
$$

It follows from [1] that the latter probability is regularly varying of index $\nu_1 - 1$. Then $J_3 = o(x^{1-\nu_1+\varepsilon})$ for all $\varepsilon > 0$ and the lemma is proved.

References

[1] S. Asmussen, C. Klüppelberg, K. Sigman (1999). Sampling at subexponential times with queueing applications. *Stoch. Proc. Appl.* **79**, 265–286.

[2] S. Asmussen and J. Teugels (1996). Convergence rates for M/G/1 queues and ruin problems with heavy tails. *J. Appl. Probab.* **33**, 1181-1190.

[3] N.H. Bingham, C.M. Goldie, and J.L. Teugels (1989). *Regular Variation.* Cambridge Univ. Press, Cambridge, England.

[4] S.C. Borst, O.J. Boxma and P.R. Jelenković (1999). Generalized processor sharing with long-tailed traffic sources. In: *Teletraffic Engineering in a Competitive World, Proc. ITC-16*, Edinburgh, UK, eds. P. Key and D. Smith (North-Holland, Amsterdam), 345-354.

[5] S.C. Borst, O.J. Boxma and P.R. Jelenković (1999). Coupled processors with regularly varying service times. Technical Report COSOR 99-11, Eindhoven University of Technology. To appear in the proceedings of Infocom 2000.

[6] O.J. Boxma, Q. Deng and A.P. Zwart (1999). Waiting-time asymptotics for the M/G/2 queue with heterogeneous servers. Technical Report COSOR 99-20, Eindhoven University of Technology.

[7] J.W. Cohen (1973). Some results on regular variation for distributions in queueing and fluctuation theory. *Journal Appl. Probab.* **10**, 343–353.

[8] J.W. Cohen (1982). *The Single Server Queue.* North-Holland Publ. Co., Amsterdam.

[9] J.W. Cohen and O.J. Boxma (1983). *Boundary Value Problems in Queueing System Analysis.* North-Holland Publ. Co., Amsterdam.

[10] R.A. Davis and T. Hsing (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879-917.

[11] A. Dudin. Queue M/G/1 in a semi-Markovian random environment. Preprint, personal communication.

[12] A. Dudin and A. Markov (1990). Calculation of the characteristics of the queue in the semi-Markovian cyclic RE. XV-th All-union workshop in computer networks, part 2, 241–246.

[13] G. Fayolle, V.A. Malyshev and M.V. Menshikov (1995). *Constructive Theory of Markov Chains.* Cambridge Univ. Press, Cambridge.

[14] S. Halfin (1972). Steady-state distribution for the buffer content of an $M/G/1$ queue with varying service rate. *SIAM J. Appl. Math.* **23**, 356-363.

[15] C.C. Heyde (1967). On large deviation problems for sums of independent random variables not attached to the normal law. *Ann. Math. Statist.* **38**, 1575-1578.

[16] D.P. Heyman and T.V. Lakshman (1996). Source models for VBR broadcast-video traffic. *IEEE/ACM Trans. Netw.* **4**, 40-48.

[17] P.R. Jelenković, A.A. Lazar and N. Semret (1997). The effect of multiple time scales and subexponentiality of MPEG video streams on queueing behaviour. *IEEE J. Sel. Areas Commun.* **15**, 1052-1071.

[18] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson (1993). On the self-similar nature of Ethernet traffic. In: *Proc. SIGCOMM '93*, pp. 183-193.

[19] W. Li, D. Shi and X. Chao (1997). Reliability analysis of $M/G/1$ queueing systems with server breakdowns and vacations. *J. Appl. Probab.* **34**, 546-555.

[20] A. de Meyer and J.L. Teugels (1980). On the asymptotic behaviour of the distributions of the busy period and service time in M/G/1. *J. Appl. Probab.*, **17**, 802-813.

[21] M.F. Neuts (1971). A queue subject to extraneous phase changes. *Adv. Appl. Probab.* **3**, 78-119.

[22] K. Park and W. Willinger, eds. (2000). *Self-Similar Network Traffic and Performance Evaluation*. Wiley, New York.

[23] G.J.K. Regterschot and J.H.A. de Smit (1986). The queue M/G/1 with Markov modulated arrivals and service. *Mathematics of Operations Research* **11**, 465-483.

[24] R. Núñez Queija (1998). Sojourn times in a processor-sharing queue with service interruptions. CWI Report PNA-R9807, Amsterdam; accepted for publication in *Queueing Systems*.

[25] T. Takine and B. Sengupta (1997). A single server queue with service interruptions. *Queueing Systems* **26**, 285-300.

[26] U. Yechiali and P. Naor (1971). Queueing problems with heterogeneous arrival and service. *Oper. Res.* **19**, 722-734.

[27] U. Yechiali (1973). A queueing-type birth-and-death process defined on a continuous-time Markov chain. *Oper.Res.* **21**, 604-609.