

Tail asymptotics of the waiting time and the busy period for the $M/G/1/K$ queues with subexponential service times

Bin Liu · Jinting Wang · Yiqiang Q. Zhao

Received: 4 June 2012 / Revised: 1 February 2013 / Published online: 8 March 2013
© Springer Science+Business Media New York 2013

Abstract We study the asymptotic behavior of the tail probabilities of the waiting time and the busy period for the $M/G/1/K$ queues with subexponential service times under three different service disciplines: FCFS, LCFS, and ROS. Under the FCFS discipline, the result on the waiting time is proved for the more general $GI/G/1/K$ queue with subexponential service times and lighter interarrival times. Using the well-known Laplace–Stieltjes transform (LST) expressions for the probability distribution of the busy period of the $M/G/1/K$ queue, we decompose the busy period into a sum of a random number of independent random variables. The result is used to obtain the tail asymptotics for the waiting time distributions under the LCFS and ROS disciplines.

Keywords $M/G/1/K$ queue · $GI/G/1/K$ queue · Waiting time · Busy period · Subexponential distribution · Tail asymptotics

Mathematics Subject Classification (2010) 60K25 · 60F10 · 60G50

1 Introduction

In this paper, we study the tail asymptotic behavior of the distribution function (d.f.) of the busy period and of the waiting time for the $M/G/1$ queueing system with a fixed capacity $K \geq 1$ (including the service position). Specifically, we assume that the i.i.d.

B. Liu (✉)

Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614-0506, USA
e-mail: bin.liu@uni.edu

J. Wang

Department of Mathematics, Beijing Jiaotong University, 100044 Beijing, China
e-mail: jtwang@bjtu.edu.cn

Y. Q. Zhao

School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada
e-mail: zhao@math.carleton.ca

service time variables have a subexponential distribution. The family of subexponential distributions is an important class of heavy-tailed distributions, both in theory and in queueing applications.

The $M/G/1$ queueing system, with either a finite or infinite capacity, is a classical queueing system about which basic properties can be found in many queueing books. For the finite-capacity model, which is closely related to our study on the busy period and waiting time, we refer readers to Miller [25], and Takagi and LaMaire [31], among others, for standard results, and particularly to the book by Takagi [30] and the references therein. It is worthwhile mentioning that tail asymptotic properties are not a focus of the above-mentioned references.

Studying tail asymptotic properties has a twofold significance—its own theoretical importance and its applications. In either case, and particularly when studying its applications, people often search for simple explicit expressions for important performance measures such as the waiting time and the busy period. This is usually accomplished by employing the useful tool of transformations. Our focus is on finding a simple explicit characterization for the tail asymptotic of the distribution function of the busy period and the waiting time with various service disciplines for the $M/G/1/K$ queue when G is heavy-tailed. Tail asymptotic properties can often lead to approximations and performance bounds.

When G is heavy-tailed, comprehensive studies on asymptotic properties for the $M/G/1$ (and also for the more general $GI/G/1$) queueing system with an infinite system capacity can be found in the literature. Most of the literature references focus on the waiting time distribution. For example, Cohen [13] considered the $GI/G/1$ queue when the service time is regularly varying, Pakes [27] studied the $M/G/1$ queue when the service time is subexponential, and Veraverbeke [32] tackled the problem through Wiener–Hopf factorizations. Borst et al. [10] considered the impact of the service discipline on delay asymptotics and Boxma and Zwart [12] reviewed the impact of scheduling on tail behavior. Recently, Boxma and Denisov [11] considered the $GI/G/1$ queue with a regularly varying service time distribution of index α and proved that under a proper service discipline, the waiting time is regularly varying with any given index value between $-\alpha$ and $1 - \alpha$. The paper by Abate and Whitt [1] considered tail behavior for the waiting time of the $M/G/1$ priority queue (both heavy tail and light tail cases, and also closely related to the busy period).

Considering an infinite-capacity, but for the busy period, De Mayer and Teugels [14] obtained a tail asymptotic result for the $M/G/1$ queue with regularly varying service times. However, it was shown in Asmussen et al. [5] that the result proven in [14] cannot be true for the entire class of subexponential distributions. Zwart [38] extended the result in [14] to the $GI/G/1$ queue under the assumption that the tail of the service time distribution is of intermediate regular variation. His method is probabilistic and revealed an insightful relationship between the busy period and the cycle maximum. Jelenković and Momčilović [20] derived an asymptotic result for the busy period distribution in the stable $GI/G/1$ queue with square-root-insensitive service times, and Baltrunas et al. [9] considered the $GI/G/1$ queue with Weibull service times.

Also for the busy period, but for light-tailed behavior, Kyprianou [22] investigated the asymptotic for the $M/G/1$ queue where a service time has a meromorphic moment generating function and Palmowski and Rolski [28] studied the $GI/G/1$

work-conserving queues and provided exact tail asymptotics for the distribution of the busy period under light-tailed assumptions.

References on asymptotic properties, either light- or heavy-tailed, for the $M/G/1$ queue with a processor sharing server are also abundant, and mainly focus on the probability of the sojourn time, either conditional or unconditional. These include Zwart and Boxma [39], Egorova et al. [16], Mandjes and Zwart [24], Yashkov [33]; Egorova and Zwart [15], Zhen and Knessl [35], Yashkov and Yashkova [34], and Zhen and Knessl [36].

From the stochastic process point of view, finite-capacity models can be studied through reflected barriers at both 0 and K , such as reflected random walks for discrete models and reflected Brownian motion/Lévy processes for continuous models. For asymptotic properties of the $M/G/1/K$ (or for the more general single server) queues with a finite capacity K , much attention has been given to the blocking probability or the loss rate as K goes to infinity. For example, Baiocchi considered the $M/G/1/K$ and $GI/M/1/K$ models, and the $MAP/G/1/K$ model in [7] and [8], respectively; Jelenković [19] obtained subexponential loss rates for a $GI/G/1$ queue; Zwart [37] studied a $GI/G/1$ queue (and also a fluid model) with a subexponential input; Pihlsgård [29] obtained loss rate asymptotics for a $GI/G/1$ queue; Miyazawa et al. [26] studied asymptotic behavior for a finite buffer queue with QBD structure; Asmussen and Pihlsgård [6] studied the loss rate associated with a reflected Lévy process; Kim and Kim [21] carried out an asymptotic analysis for the loss probability of queues of the $GI/M/1$ type; Andersen and Asmussen [3] examined loss rate asymptotics for centered Lévy processes; Andersen [2] derived subexponential loss rate asymptotics for a reflected Lévy process when the mean is negative and the positive jumps are subexponential; and Liu and Zhao [23] considered an $M/G/1/N$ queue with server vacations and exhaustive service discipline.

In contrast to the above-mentioned studies, our focus in this paper is on tail asymptotic properties of the busy period and the waiting time of the $M/G/1/K$ queue when G is a subexponential random variable with distribution $B(x)$. Our main contributions in this paper include: (1) Based on the recursive formula with respect to the capacity K for the busy period, we connect the LST of the busy period distribution with a geometric sum of random variables. We derive the tail asymptotics for the busy period; (2) we study the tail asymptotic behavior for the waiting time distribution with the three service disciplines: first-come-first-served (FCFS), last-come-first-served (LCFS), and random-order-service (ROS). Under the FCFS discipline, the result is proved for the more general $GI/G/1$ queue with lighter interarrival times. The main results are reported in Theorems 2.1, 3.1, 4.1, and 5.1.

Throughout the paper, for a given non-negative r.v. X , its distribution function is denoted by $F(x) = P(X \leq x)$ and its tail (or, survival) function is denoted by $\bar{F}(x) = 1 - F(x) = P(X > x)$. A d.f. F has a light tail if there exists $\varepsilon > 0$ such that $E(e^{\varepsilon X}) < \infty$. Otherwise, the d.f. F is referred to as heavy-tailed. In this paper, we focus on a special class, \mathcal{S} , of heavy-tailed distributions, called subexponential distributions.

Definition 1.1 Let F^{*n} denote the n -fold convolution of F with corresponding tail $\overline{F^{*n}} = 1 - F^{*n}$, then the d.f. F is called subexponential (or $F \in \mathcal{S}$) if $\overline{F}(x) > 0$, $x \geq 0$, and for all $n \geq 2$,

$$\lim_{x \rightarrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n.$$

It can be shown that if the condition holds for $n = 2$ then it holds for all $n \geq 2$. The reader is referred to [17] or [18] for details and further references about subexponential distributions.

Also, throughout the paper the notations $a(x) \sim b(x)$ and $a(x) = o(b(x))$ mean that $a(x)/b(x) \rightarrow 1$ and $a(x)/b(x) \rightarrow 0$ as $x \rightarrow \infty$, respectively. For two random variables, X and Y , with d.f.s $F(x)$ and $G(x)$, respectively, we say X has a lighter tail if $\overline{F}(x) = o(\overline{G}(x))$ as $x \rightarrow \infty$. We denote by $\stackrel{d}{=}$ the equality in distribution, i.e., $X \stackrel{d}{=} Y$ means that the d.f.s of r.v.s X and Y are the same.

The rest of this paper is organized as follows. In Sect. 2, we obtain a tail asymptotic result on the waiting time of the $GI/G/1/K$ queue for the FCFS case. In Sect. 3, the tail asymptotic result is proved for the busy period of the $M/G/1/K$ queue. Sections 4 and 5 are devoted to the tail asymptotic results on the waiting time of the $M/G/1/K$ queue for the LCFS and ROS cases, respectively. Appendix contains literature results that are used in our analysis.

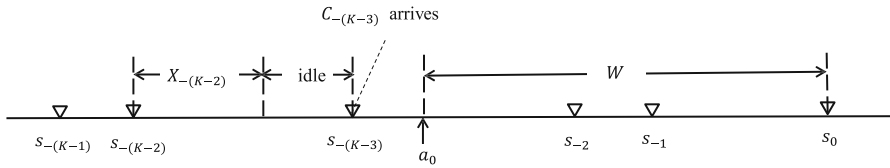
2 Waiting Time for FCFS System

In this section, we prove a tail asymptotic result on the waiting time distribution for the $GI/G/1/K$ system with subexponential service times and lighter interarrival times under the FCFS discipline. The system is assumed to be stable. Let W be the waiting time of an arriving customer (the tagged customer) who is accepted in to the system. We first prove the following fact:

Lemma 2.1 Let Z_1 and Z_2 be independent r.v.s with d.f.s F_1 and F_2 , respectively. Define $Z = Z_1 + Z_2$. Assume that Z has a d.f. $F \in \mathcal{S}$ and $\overline{F}_2(t) = o(\overline{F}(t))$. Then, $F_1 \in \mathcal{S}$ and $\overline{F}_1(t) \sim \overline{F}(t)$ as $t \rightarrow \infty$.

Proof We can write

$$\begin{aligned} \overline{F}(t) &= P(Z_1 + Z_2 > t) = P(Z_1 > t) + P(Z_2 > t, Z_1 \leq t) \\ &\quad + P(Z_1 + Z_2 > t, Z_2 \leq t, Z_1 \leq t) \\ &= \overline{F}_1(t) + \overline{F}_2(t)F_1(t) + \int_0^t (\overline{F}_2(t-y) - \overline{F}_2(t))dF_1(y). \end{aligned} \tag{2.1}$$



s_0 = the service starting time of customer C_0 ; s_{-j} = the service starting time of customer C_{-j} ;
 a_0 = the arrival time of customer C_0 (the tagged customer); X_{-j} = the length of service time of customer C_{-j} .

Fig. 1 The service starting times of $C_{-(K-1)}, \dots, C_0$ and the arriving time of C_0

Note that

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{1}{\bar{F}(t)} \int_0^t (\bar{F}_2(t-y) - \bar{F}_2(t)) dF_1(y) \\ & \leq \limsup_{t \rightarrow \infty} \int_0^\infty \left(\frac{\bar{F}_2(t-y)}{\bar{F}(t-y)} \cdot \frac{\bar{F}(t-y)}{\bar{F}(t)} - \frac{\bar{F}_2(t)}{\bar{F}(t)} \right) dF_1(y) = 0, \end{aligned} \quad (2.2)$$

where we have used the facts: $\lim_{t \rightarrow \infty} \bar{F}(t-y)/\bar{F}(t) = 1$ (by Lemma 6.1) and the dominated convergence theorem for interchanging the limit and the integral. Now, $\lim_{t \rightarrow \infty} \bar{F}_1(t)/\bar{F}(t) = 1$ follows from (2.1) and (2.2). \square

Remark 2.1 The above fact is useful, which is very likely available in the literature [18].

Theorem 2.1 For the $GI/G/1/K$ queueing system with a subexponential service time distribution $B(x)$ and a lighter interarrival time under the FCFS discipline, $P(W > x) \sim (K - 1)\bar{B}(x)$ as $x \rightarrow \infty$.

Proof We refer to the tagged customer as C_0 , and denote C_{-j} ($j = 1, 2, \dots$) by the j th last customer who enters the system before C_0 (blocked arrivals are not counted). Let s_{-j} represent the “service starting time” of C_{-j} ($j = 0, 1, \dots$), X_{-j} represent the length of service time of C_{-j} ($j = 1, 2, \dots$), and a_0 represent the arrival time of C_0 . Figure 1 depicts a typical order of service starting times of $C_{-(K-1)}, C_{-(K-2)}, \dots, C_0$ and the arriving time of C_0 . The following observations are straightforward.

- (1) The start of the service of customer $C_{-(K-1)}$ must be before the arrival of customer C_0 , i.e., $s_{-(K-1)} < a_0$. This is because $C_{-(K-1)}, C_{-(K-2)}, \dots, C_{-1}$ must have entered the system before C_0 does.
- (2) All customers that arrived during $(s_{-(K-1)}, a_0)$ should be accepted into the system without blocking. The number of arrivals during $(s_{-(K-1)}, a_0)$ is $K - \tau - 1$, where τ = the number of customers in the system at time $t = s_{-(K-1)}$ and $1 \leq \tau \leq K - 1$. Precisely, $C_{-(K-\tau-1)}, C_{-(K-\tau-2)}, \dots, C_{-1}$ arrive during $(s_{-(K-1)}, a_0)$.

By (2), we know that $a_0 - s_{-(K-1)}$ equals the sum of $K - \tau$ ($1 \leq \tau \leq K - 1$) interarrival times, where the first interarrival time should be interpreted as a *partial* interarrival time. Hence

$$\begin{aligned} s_0 - s_{-(K-1)} &= s_0 - a_0 + a_0 - s_{-(K-1)} \\ &= W + \text{sum of } K - \tau \text{ interarrival (or partial interarrival) times.} \end{aligned} \quad (2.3)$$

On the other hand, $s_{-(j-1)} - s_{-j}$ is the length of time between the service startings of customers C_{-j} and $C_{-(j-1)}$. It is worthwhile noticing that there may be an idle period between two consecutive services before $t = a_0$. So $s_{-(j-1)} - s_{-j} = X_{-j} +$ at most one idle period. Since $C_{-(K-1)}, C_{-(K-2)}, \dots, C_{K-\tau}$ are already in the system at time $t = s_{-(K-1)}$ and they are continuously served without any idle period in between, we have

$$\begin{aligned} s_0 - s_{-(K-1)} &= (s_0 - s_{-1}) + (s_{-1} - s_{-2}) + \dots + (s_{-(K-2)} - s_{-(K-1)}) \\ &= \sum_{j=1}^{K-1} X_{-j} + \text{sum of at most } K - \tau \text{ idle periods.} \end{aligned} \quad (2.4)$$

It follows from (2.3) and (2.4) that

$$W + Y_1 = \sum_{j=1}^{K-1} X_{-j} + Y_2, \quad (2.5)$$

where $Y_1 \triangleq$ sum of $K - \tau$ interarrival (or partial interarrival) times, and $Y_2 \triangleq$ sum of at most $K - \tau$ idle periods.

By Lemma 6.3 and Corollary 6.1, we have $P(\sum_{j=1}^{K-1} X_{-j} + Y_2 > t) \sim P(\sum_{j=1}^{K-1} X_{-j} > t) \sim (K-1)\bar{B}(t)$ because an idle period has a lighter tail than $\bar{B}(t)$. So, $P(W + Y_1 > t) \sim (K-1)\bar{B}(t)$ as $t \rightarrow \infty$. Again, by Lemma 6.3, $\lim_{t \rightarrow \infty} P(Y_1 > t)/\bar{B}(t) = 0$ because an interarrival time has a lighter tail than $\bar{B}(t)$. Applying Lemma 2.1, we have $P\{W > t\} \sim (K-1)\bar{B}(t)$ as $t \rightarrow \infty$. \square

3 Busy Period of $M/G/1/K$ Queue

In this section, we investigate the asymptotic behavior for the busy period for the $M/G/1/K$ system. Assume that customers arrive according to a Poisson process with rate λ . Let X be the length of a generic service time, and $N(X)$ be the number of arrivals during $(0, X)$. Let $c_k = P(N(X) = k) = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x)$, $k \geq 0$, which is the probability that k customers arrive during a service time. Further, we let $\bar{c}_k = P(N(X) \geq k) = 1 - \sum_{j=0}^{k-1} c_j$, $k \geq 1$, which is the probability that k or more customers arrive during a service time.

Define $h_1 = 1$ and h_n , recursively, by

$$h_n = c_0^{-1} \cdot \left[1 + \sum_{j=2}^{n-1} \bar{c}_{n-j+1} h_j \right], \quad n \geq 2. \quad (3.1)$$

The main result in this section is the following theorem.

Theorem 3.1 *Let A_n be the length of the busy period of the $M/G/1/n$ queueing system with a subexponential service time distribution $B(x)$. Then, $P(A_n > x) \sim h_n \bar{B}(x)$ as $x \rightarrow \infty$, where h_n is given by (3.1).*

To prove the above tail asymptotic result, let us use $\alpha_n(s)$ to represent the LST of the distribution function of A_n . It is well known that $\alpha_n(s)$ can be expressed in a recursive form with respect to the system capacity n as follows (see for example, p. 225 in Takagi [30], or Miller [25]):

$$\alpha_1(s) = \beta(s), \tag{3.2}$$

$$\alpha_n(s) = \frac{u_0(s)}{1 - \sum_{k=1}^{n-2} u_k(s) \prod_{j=n-k+1}^{n-1} \alpha_j(s) - [\sum_{k=n-1}^{\infty} u_k(s)] \prod_{j=2}^{n-1} \alpha_j(s)}, \quad n \geq 2, \tag{3.3}$$

where $\beta(s)$ is the LST of the service time distribution and

$$u_k(s) = \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-(\lambda+s)t} dB(t).$$

It is a convention that for $b < a$, $\sum_a^b \equiv 0$ and $\prod_a^b \equiv 1$.

The proof to Theorem 3.1 is carried out through the following steps with the use of properties in Appendix: (a) rewrite $\alpha_n(s)$ as the LST of the distribution function of the sum of two independent random variables, one of which is a geometric sum of i.i.d. random variables; (b) properly define random variables associated with the geometric sum; and (c) tail asymptotic analysis for components of the random variables defined in (b).

For the first step, we note that $u_k(s)$ is not the LST of a probability distribution function (since $u_k(0) \neq 1$). However, we can write $u_k(s) = c_k \beta_k(s)$, where $\beta_k(s)$ is the LST of the conditional distribution function

$$\begin{aligned} B_k(t) &\triangleq P(X < t | N(X) = k) \\ &= c_k^{-1} \int_0^t \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x), \quad k \geq 0. \end{aligned} \tag{3.4}$$

Now, we rewrite (3.3) in term of $\beta_k(s)$ as

$$\alpha_n(s) = \sigma_n(s) \cdot \beta_0(s), \quad n \geq 2, \tag{3.5}$$

where

$$\sigma_n(s) = \frac{1 - \bar{c}_1}{1 - \bar{c}_1 \delta_n(s)}, \tag{3.6}$$

$$\delta_n(s) = \sum_{k=1}^{n-2} \frac{c_k}{\bar{c}_1} \beta_k(s) \prod_{j=n-k+1}^{n-1} \alpha_j(s) + \frac{\bar{c}_{n-1}}{\bar{c}_1} \tau_{n-1}(s) \prod_{j=2}^{n-1} \alpha_j(s), \tag{3.7}$$

$$\tau_n(s) = \sum_{k=n}^{\infty} \frac{c_k}{\bar{c}_n} \beta_k(s). \tag{3.8}$$

Next, we will show that $\tau_n(s)$, $\delta_n(s)$, and $\sigma_n(s)$ can be viewed as the LSTs of probability distributions. To this end, let B_n be a random variable having distribution $B_n(x)$ (by abusing the notation). Define

$$T_n \stackrel{d}{=} B_k \text{ with probability } c_k/\bar{c}_n, \quad k \geq n \geq 1, \tag{3.9}$$

$$D_n \stackrel{d}{=} \begin{cases} B_k + \sum_{j=n-k+1}^{n-1} A_j \text{ with probability } c_k/\bar{c}_1, & 1 \leq k \leq n-2 \\ T_{n-1} + \sum_{j=2}^{n-1} A_j \text{ with probability } \bar{c}_{n-1}/\bar{c}_1, & n \geq 2. \end{cases} \tag{3.10}$$

By Lemma 6.6, it is easy to see that the LSTs of the d.f.s of random variables T_n and D_n are $\tau_n(s)$ and $\delta_n(s)$, respectively. According to Lemma 6.4, $\sigma_n(s)$ can be viewed as the LST of the d.f. of a geometric sum (denoted by S_n) with parameter \bar{c}_1 of i.i.d. random variables (each equal to D_n in distribution). By (3.5), $\alpha_n(s)$ is the LST of the d.f. of random variable $A_n \stackrel{d}{=} S_n + B_0$, the sum of the geometric sum S_n and the variable B_0 (which is independent of the geometric sum S_n).

Now, let us characterize the tail property for random variables B_k , T_n , and D_n . It is easy to verify that

$$P\{B_k > x\} = \int_x^{\infty} c_k^{-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) \leq c_k^{-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x} \quad \text{for } x \geq k/\lambda. \tag{3.11}$$

Namely, $B_k(x)$ ($k = 0, 1, 2, \dots$) is a light-tailed distribution. So, $\bar{B}_k(x) = o(1)\bar{B}(x)$ as $x \rightarrow \infty$.

It follows from the definition of T_n in (3.9) and (3.11) that

$$\begin{aligned} P\{T_n > x\} &= \sum_{k=n}^{\infty} \frac{c_k}{\bar{c}_n} P\{B_k > x\} = \frac{1}{\bar{c}_n} \int_x^{\infty} \left[1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right] dB(t) \\ &\sim \bar{B}(x)/\bar{c}_n, \quad \text{as } x \rightarrow \infty. \end{aligned} \tag{3.12}$$

For the tail behavior of D_n , it follows from its definition given in (3.10) that

$$P\{D_n > x\} = \sum_{k=1}^{n-2} \frac{c_k}{\bar{c}_1} P\left[\left(B_k + \sum_{j=n-k+1}^{n-1} A_j\right) > x\right] + \frac{\bar{c}_{n-1}}{\bar{c}_1} P\left[\left(T_{n-1} + \sum_{j=2}^{n-1} A_j\right) > x\right]. \tag{3.13}$$

We are now ready to prove the theorem, i.e., $P\{A_n > x\} \sim h_n \bar{B}(x)$ as $x \rightarrow \infty$ for $n \geq 1$.

Proof of Theorem 3.1 The proof of case $n = 1$ is immediate since $A_1 \stackrel{d}{=} X$ (a service time). We employ the induction on n to prove the general case of $n \geq 2$.

For the case of $n = 2$, $D_2 \stackrel{d}{=} T_1$ by the definition of D_n in (3.10). It follows from (3.12) that $P\{D_2 > x\} = P\{T_1 > x\} \sim \bar{B}(x)/\bar{c}_1$, which implies that D_2 has a subexponential distribution according to Lemma 6.2. Therefore, by Lemma 6.5,

$$P\{S_2 > x\} \sim \frac{\bar{c}_1}{1 - \bar{c}_1} \cdot \frac{1}{\bar{c}_1} \bar{B}(x) = \frac{\bar{B}(x)}{c_0} = h_2 \bar{B}(x).$$

Furthermore, since B_0 is light-tailed, $P\{A_2 > x\} \sim P\{B_0 + S_2 > x\} \sim h_2 \bar{B}(x)$ according to Corollary 6.1. The proof for the case of $n = 2$ is complete.

Assuming that $P\{A_k > x\} \sim h_k \bar{B}(x)$ for all k ($1 \leq k \leq n - 1$), let us verify that $P\{A_n > x\} \sim h_n \bar{B}(x)$ is also true. It follows from (3.13), Lemma 6.3, Corollary 6.1 and (3.12) that

$$\begin{aligned} P\{D_n > x\} &\sim \sum_{k=2}^{n-2} \frac{c_k}{\bar{c}_1} \sum_{j=n-k+1}^{n-1} h_j \bar{B}(x) + \frac{\bar{c}_{n-1}}{\bar{c}_1} \left[\frac{1}{\bar{c}_{n-1}} \bar{B}(x) + \sum_{j=2}^{n-1} h_j \bar{B}(x) \right] \\ &= \left[1 + \sum_{k=2}^{n-2} c_k \sum_{j=n-k+1}^{n-1} h_j + \bar{c}_{n-1} \sum_{j=2}^{n-1} h_j \right] \cdot \bar{B}(x)/\bar{c}_1. \end{aligned} \tag{3.14}$$

Recall that $A_n = S_n + B_0$. By (3.11), Corollary 6.1 and (3.14),

$$\begin{aligned} P\{A_n > x\} &\sim P\{S_n > x\} \sim \frac{\bar{c}_1}{1 - \bar{c}_1} \cdot P\{D_n > x\} \\ &= c_0^{-1} \cdot \left[1 + \sum_{k=2}^{n-2} c_k \sum_{j=n-k+1}^{n-1} h_j + \bar{c}_{n-1} \sum_{j=2}^{n-1} h_j \right] \bar{B}(x) \\ &= c_0^{-1} \cdot \left[1 + \sum_{j=2}^{n-1} \bar{c}_{n-j+1} h_j \right] \bar{B}(x) \\ &= h_n \bar{B}(x), \end{aligned}$$

where the second last equation holds after interchanging the summations: $\sum_{k=2}^{n-2} \sum_{j=n-k+1}^{n-1} = \sum_{j=3}^{n-1} \sum_{k=n-j+1}^{n-2}$ and noticing that: $\bar{c}_{n-j+1} = \bar{c}_{n-1} + \sum_{k=n-j+1}^{n-2} c_k$. The proof is complete. \square

4 Waiting Time for LCFS System

In this section, we prove a tail asymptotic result for the waiting time distribution for the $M/G/1/K$ system with non-preemptive LCFS queueing discipline. Clearly, if

$K = 1$ or 2 , the waiting time of a customer remains the same regardless of the service discipline. So, we assume that $K \geq 3$. The system is assumed to be stable. Definitions and notations used in the previous sections remain valid here, such as $W, X, N(X), A_n, B_n, B_k(t), c_k, \bar{c}_k, h_n$, and $\alpha_n(s)$. Unlike the case of the FCFS, the waiting time under the LCFS is closely related to the number of arrivals during the remaining service time, and therefore an explicit expression of the joint stationary distribution of the queue length and the remaining service time is required.

Consider the stationary system at an arbitrary time. Let L be the number of customers in the system, X_- be the elapsed service time, and X_+ be the remaining service time for the customer who is in service (if any). Denote $p_k \triangleq P(L = k), 0 \leq k \leq K$, and

$$p_k(x)dx \triangleq P(L = k, x \leq X_- < x + dx), \quad 1 \leq k \leq K \text{ and } x \geq 0.$$

By (1.16) on p. 202 and (1.60a) on p. 213 in Takagi [30],

$$p_K = 1 - (1 - p_0)/\rho, \tag{4.1}$$

$$p_k = p_0 c_k + \sum_{j=1}^{k+1} p_j c_{k-j+1}, \quad 0 \leq k \leq K - 2, \tag{4.2}$$

where $\rho = \lambda \int_0^\infty x dB(x)$. Note that $p_k (0 \leq k \leq K)$ are determined by independent

Eqs. (4.1) and (4.2) combined with the normalizing condition $\sum_{k=0}^K p_k = 1$.

By (1.61a) on p. 214 in Takagi [30], we have

$$p_k(x) = \lambda \bar{B}(x) e^{-\lambda x} \left[p_0 \frac{(\lambda x)^{k-1}}{(k-1)!} + \sum_{j=1}^k p_j \frac{(\lambda x)^{k-j}}{(k-j)!} \right], \quad 1 \leq k \leq K - 1 \text{ and } x \geq 0. \tag{4.3}$$

A formula for $k = K$ is also available, but we do not need it here.

We study $P(W > t)$, the tail probability of the waiting time of an arriving customer (the tagged customer) who is accepted in the system. Because Poisson arrivals see time average (PASTA), the tagged customer sees the system with $L (0 \leq L \leq K)$ customers (excluding itself) and the remaining service time X_+ if $L \neq 0$. Let W' be the waiting time of this customer before receiving its service (we define $W' = \infty$ if it is blocked). So $P(W > t) = P(W' > t | 0 \leq L \leq K - 1)$.

Let J be the number of customers that arrived (including those blocked) to the system during the remaining service time X_+ seen by the tagged customer. Define the following conditional distribution functions

$$\begin{aligned}
 G_k(t) &\triangleq P(X_+ \leq t | L = k), \quad 1 \leq k \leq K - 1, \\
 G_{k,j}(t) &\triangleq P(X_+ \leq t | L = k, J = j), \quad 1 \leq k \leq K - 1 \text{ and } j \geq 0, \\
 T_{k,n}(t) &\triangleq P(X_+ \leq t | L = k, J \geq n), \quad 1 \leq k \leq K - 1 \text{ and } n \geq 1.
 \end{aligned}$$

By the definition of $G_k(t)$, we can write

$$\begin{aligned}
 G_k(t) = p_k^{-1} P(L = k, X_+ \leq t) &= p_k^{-1} \int_0^\infty P(X_+ \leq t | X_- = x) p_k(x) dx \\
 &= p_k^{-1} \int_0^\infty \frac{B(x+t)}{B(x)} p_k(x) dx. \tag{4.4}
 \end{aligned}$$

Let

$$c_{k,j} = P(J = j | L = k) = \int_0^\infty \frac{(\lambda x)^j}{j!} e^{-\lambda x} dG_k(x), \quad j \geq 0, \quad 1 \leq k \leq K - 1, \tag{4.5}$$

$$\bar{c}_{k,j} = P(J \geq j | L = k) = \sum_{n=j}^\infty c_{k,n}, \quad j \geq 1, \quad 1 \leq k \leq K - 1. \tag{4.6}$$

We will prove the following result:

Theorem 4.1 *For the stable $M/G/1/K$ with subexponential service times and non-preemptive LCFS discipline, $P(W > x) \sim w_{lcf s} \cdot \bar{B}(x)$ as $x \rightarrow \infty$, where*

$$w_{lcf s} = \frac{1}{1 - p_K} \sum_{k=1}^{K-2} \left[p_k \sum_{n=2}^{K-k} h_n \bar{c}_{k, K-(k+n-1)} + \sum_{j=0}^k p_j \right] + 1.$$

It is obvious that $W' = X_+$ if $L = K - 1$ or $J = 0$. Otherwise, there are $J \geq 1$ arrivals (accepted or rejected to the system depending on whether or not the system is full) during the remaining service time. When $1 \leq L \leq K - 2$ and $J \geq K - L$, after the remaining service time X_+ the server will start serving the last accepted customer to the system, and after time A_2 the server will start serving the second last accepted customer to the system that arrived during the remaining service time. This will continue until all customers accepted to the system during the remaining service time have been served, and then the server will start serving the tagged customer. The total time is

$$W' = X_+ + A_2 + A_3 + \dots + A_{K-L}.$$

When $1 \leq L \leq K - 2$ and $J \leq K - L - 1$, the waiting time for the server to start serving the tagged customer can be similarly expressed in terms of the remaining service time X_+ and the busy period A_n . So, we have

$$W' = \begin{cases} 0, & \text{if } L=0, \\ X_+ + \sum_{n=K-k-j+1}^{K-k} A_n, & \text{if } L=k \ (1 \leq k \leq K-2) \text{ and } J=j \ (0 \leq j \leq K-k-1), \\ X_+ + \sum_{n=2}^{K-k} A_n, & \text{if } L=k \ (1 \leq k \leq K-2) \text{ and } J=j \ (j \geq K-k), \\ X_+, & \text{if } L=K-1, \\ \infty, & \text{if } L=K. \end{cases}$$

Let $T_{k,n}$, $G_{k,j}$ and G_k be random variables having d.f.s $T_{k,n}(t)$, $G_{k,j}(t)$ and $G_k(t)$, respectively. It follows from the above expression of W' that

$$P(W > t) = \sum_{k=1}^{K-2} \frac{p_k}{1-p_K} \left[\sum_{j=0}^{K-k-1} c_{k,j} P\left(G_{k,j} + \sum_{n=K-(k+j-1)}^{K-k} A_n > t\right) + \bar{c}_{k,K-k} P\left(T_{k,K-k} + \sum_{n=2}^{K-k} A_n > t\right) \right] + \frac{p_{K-1}}{1-p_K} P(G_{K-1} > t). \tag{4.7}$$

Next, we will study the tail probabilities of $T_{k,n}$, $G_{k,j}$ and G_k . For this purpose, we consider the probability of event $\{L = k, X_+ > t\}$, which will be used in the proof.

$$P(L = k, X_+ > t) = \int_0^\infty P(X_+ > t | X_- = x) p_k(x) dx = \int_0^\infty \frac{\bar{B}(x+t)}{\bar{B}(x)} p_k(x) dx. \tag{4.8}$$

For a fix $a > 0$, it follows from (4.8) that

$$\frac{P(L = k, X_+ > t)}{\bar{B}(t)} = \left(\int_0^a + \int_a^\infty \right) \frac{\bar{B}(x+t)}{\bar{B}(t)} \frac{p_k(x)}{\bar{B}(x)} dx. \tag{4.9}$$

For $t \geq 0$,

$$\int_a^\infty \frac{\bar{B}(x+t)}{\bar{B}(t)} \frac{p_k(x)}{\bar{B}(x)} dx \leq \int_a^\infty \frac{p_k(x)}{\bar{B}(x)} dx \rightarrow 0 \text{ as } a \rightarrow \infty. \tag{4.10}$$

According to Lemma 6.1, $\bar{B}(x+t)/\bar{B}(t) \rightarrow 1$ uniformly on $0 \leq x \leq a$ as $t \rightarrow \infty$. Therefore, from (4.9), (4.10) and (4.3)

$$\lim_{t \rightarrow \infty} \frac{P(L = k, X_+ > t)}{\bar{B}(t)} = \int_0^\infty \frac{p_k(x)}{\bar{B}(x)} dx = \sum_{j=0}^k p_j, \quad 1 \leq k \leq K-1. \tag{4.11}$$

By the definitions of $G_k(t)$, $G_{k,j}(t)$ and $T_{k,j}(t)$, we have, as $t \rightarrow \infty$,

$$P(G_k > t) = P(X_+ > t | L = k) = p_k^{-1} P(X_+ > t, L = k) \sim \left(p_k^{-1} \sum_{j=0}^k p_j \right) \bar{B}(t), \tag{4.12}$$

$$P(G_{k,j} > t) = P(X_+ > t | L = k, J = j) = o(1) \bar{B}(t), \tag{4.13}$$

$$\begin{aligned} P(T_{k,n} > t) &= P(X_+ > t | L = k, J \geq n) \\ &= (1/\bar{c}_{k,n}) P(X_+ > t, J \geq n | L = k) \\ &= (1/\bar{c}_{k,n}) \int_t^\infty \left[1 - \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x} \right] dG_k(x) \\ &\sim (1/\bar{c}_{k,n}) \bar{G}_k(t) \\ &\sim \left((1/\bar{c}_{k,n}) p_k^{-1} \sum_{j=0}^k p_j \right) \bar{B}(t). \end{aligned} \tag{4.14}$$

It follows from (4.7), Theorem 3.1 and (4.12–4.14) that

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{P(W > t)}{\bar{B}(t)} &= \frac{1}{1-p_K} \sum_{k=1}^{K-2} p_k \left[\sum_{j=1}^{K-k-1} c_{k,j} \sum_{n=K-(k+j-1)}^{K-k} h_n + \bar{c}_{k,K-k} \sum_{n=2}^{K-k} h_n + p_k^{-1} \sum_{j=0}^k p_j \right] + 1 \\ &= \frac{1}{1-p_K} \sum_{k=1}^{K-2} \left[p_k \sum_{n=2}^{K-k} h_n \bar{c}_{k,K-(k+n-1)} + \sum_{j=0}^k p_j \right] + 1, \end{aligned}$$

where the last equation holds after interchanging the summations: $\sum_{j=1}^{K-k-1} \sum_{n=K-(k+j-1)}^{K-k} = \sum_{n=2}^{K-k} \sum_{j=K-(k+n-1)}^{K-k-1}$ and noticing: $\bar{c}_{k,K-k-n+1} = \bar{c}_{k,K-k} + \sum_{j=K-(k+n-1)}^{K-k-1} c_{k,j}$.

5 Waiting Time for ROS System

In this section, we provide a tail asymptotic result on the waiting time for the $M/G/1/K$ system with ROS service discipline. The same method used for LCFS systems will be used here. Again, the system is assumed to be stable. Definitions and notations introduced so far remain valid in this section. For the same reason as in the previous section, we assume that $K \geq 3$.

To state the main theorem, let

$$w_0 = 0, \tag{5.1}$$

$$w_k = \frac{k}{k+1} \left[\left(\sum_{j=0}^{K-k-1} c_j w_{k+j-1} \right) + 1 + \bar{c}_{K-k} w_{K-2} \right], \quad 1 \leq k \leq K-2. \tag{5.2}$$

The main result is given in the following theorem.

Theorem 5.1 *For the stable $M/G/1/K$ with subexponential service times and ROS discipline, $P(W > x) \sim w_{ros} \cdot \bar{B}(x)$ as $x \rightarrow \infty$, where*

$$w_{ros} = \frac{1}{1 - p_K} \sum_{k=1}^{K-1} \left[p_k \sum_{j=0}^{K-k-1} c_{k,j} w_{k+j-1} + p_k \bar{c}_{K-k} w_{K-2} + \sum_{j=0}^k p_j \right]. \tag{5.3}$$

In a ROS system, suppose that the tagged customer sees the system with $L = k$. If $k \neq 0$ or K , and j customers arrived (not necessarily accepted to the system) during the remaining service time X_+ , then there are $\min(k + j - 1, K - 2)$ other customers (excluding the tagged one) in the system at the end of the remaining service time X_+ . Therefore, we can write $W' = X_+ + W_{\min(k+j-1, K-2)}$, where W_k is the length of the time period starting from the completion of a service to the end of the waiting time of the tagged customer, given that there are k other customers (excluding the tagged one) left behind in the system at the completion of that service. We therefore have

$$W' = \begin{cases} 0, & \text{if } L = 0, \\ X_+ + W_{k+j-1}, & \text{if } L = k \ (1 \leq k \leq K-1) \text{ and } J = j \ (0 \leq j \leq K-k-1), \\ X_+ + W_{K-2}, & \text{if } L = k \ (1 \leq k \leq K-1) \text{ and } J = j \ (j \geq K-k), \\ \infty, & \text{if } L = K. \end{cases} \tag{5.4}$$

To characterize the tail behavior of the waiting time: $P(W > t) = P(W' > t | 0 \leq L \leq K - 1)$, the key is to characterize the tail behavior of $P(W_k > t)$ for $0 \leq k \leq K - 2$. Because of the random selection, it is clear that $W_k = 0$ with probability $1/(k + 1)$ and $W_k > 0$ with probability $k/(k + 1)$. In the latter case, the tagged customer has to wait one service time before the next selection made by the server. Hence,

$$P(W_k > t) = \frac{k}{k+1} \left[\sum_{j=0}^{K-k-1} c_j P(B_j + W_{k+j-1} > t) + \bar{c}_{K-k} P(T_{K-k} + W_{K-2} > t) \right], \tag{5.5}$$

where $0 \leq k \leq K - 2$.

Proposition 5.1 (1) *For $x > 0$, $P\{W_0 > x\} = 0$; (2) For $1 \leq k \leq K - 2$, the asymptotic tail probability of W_k is given by*

$$P\{W_k > x\} \sim w_k \bar{B}(x), \quad \text{as } x \rightarrow \infty, \tag{5.6}$$

where the recursive expression for w_k is given in (5.2).

Proof Since $W_0 \equiv 0$, part (1) is obvious. To prove (5.6), the key is to verify the existence of $\lim_{x \rightarrow \infty} P(W_k > x) / \bar{B}(x)$ for all $1 \leq k \leq K - 2$, which can be done by the mathematical induction on K ($K \geq 3$). When $K = 3$, W_1 equals the sum of a light-tailed number of service times, so $\lim_{x \rightarrow \infty} P(W_1 > x) / \bar{B}(x)$ exists by Lemma 6.5. Now, assuming that all $\lim_{x \rightarrow \infty} P(W_k > x) / \bar{B}(x)$ exist when $K = n$, let us see the case of $K = n + 1$. Conditioning on the events $E = \{\text{no other customers in the system when the tagged customer starts its service}\}$ and its complement event E^c , we have $P(W_k > x) = P(E)P(W_k > x|E) + P(E^c)P(W_k > x|E^c)$. The existence of $\lim_{x \rightarrow \infty} P(W_k > x|E^c) / \bar{B}(x)$ follows from the induction assumption ($K = n$) because one position is always occupied by a customer other than the tagged customer before its service. The existence of $\lim_{x \rightarrow \infty} P(W_k > x|E) / \bar{B}(x)$ follows from the fact that $P(W_k > x|E) = P(A_{n+1-k} + A_{n+1-(k-1)} + \dots + A_n > x)$ and from Theorem 3.1, where A_m is the busy period of the $M/G/1/m$ queue. So, all $\lim_{x \rightarrow \infty} P(W_k > x) / \bar{B}(x)$ exist for the case $K = n + 1$.

Let $\lim_{x \rightarrow \infty} P(W_k > x) / \bar{B}(x) \triangleq w_k$. It follows from (5.5), (3.11) and (3.12) that w_k satisfies (5.2). □

Proof of Theorem 5.1 Recall that W is the waiting time of an arriving customer who is accepted to a steady state system and $P(W > t) = P(W' > t | 0 \leq L \leq K - 1)$. Following the same treatment as that in the previous section, we have

$$P(W > t) = \sum_{k=1}^{K-1} \frac{p_k}{1 - p_K} \left[\sum_{j=0}^{K-k-1} c_{k,j} P(G_{k,j} + W_{k+j-1} > t) + \bar{c}_{k,K-k} P(T_{k,K-k} + W_{K-2} > t) \right].$$

Finally, by Proposition 5.1, (4.13) and (4.14),

$$\lim_{t \rightarrow \infty} \frac{P(W > t)}{\bar{B}(t)} = \frac{1}{1 - p_K} \sum_{k=1}^{K-1} p_k \left[\left(\sum_{j=0}^{K-k-1} c_{k,j} w_{k+j-1} \right) + \left(p_k^{-1} \sum_{j=0}^k p_j \right) + \bar{c}_{k,K-k} w_{K-2} \right].$$

□

Acknowledgments The authors thank an anonymous referee for his/her suggestions, which inspired us to have a significantly improved proof to Theorem 2.1 and to generalize our result to the $G1/G/1/K$ queue. The authors also thank Drs. O. Boxma and B. Zwart for their suggestions of references. This research was supported in part through Summer Fellowships of the University of Northern Iowa, the National Natural Science Foundation of China (Grant No. 11171019) and an NSERC Discovery Grant.

6 Appendix: Useful Literature Results

For convenience, we collected useful literature results for this paper in this appendix.

The following two basic properties of the subexponential distribution can be found in Embrechts et al. [17].

Lemma 6.1 (Lemma 1.3.5 in [17]) *If $F \in \mathcal{S}$, let X be a random variable having F as its distribution. Then, (a)*

$$\lim_{x \rightarrow \infty} P(X > x + y | X > x) = \lim_{x \rightarrow \infty} \frac{\overline{F}(x + y)}{\overline{F}(x)} = 1, \text{ uniformly for any compact set of } y;$$

and (b)

$$\lim_{x \rightarrow \infty} e^{\epsilon x} P(X > x) = \infty = E(e^{\epsilon X}), \text{ for all } \epsilon > 0.$$

The following lemma says that the class \mathcal{S} is closed under tail-equivalence.

Lemma 6.2 (Lemma A 3.15 in [17]) *Suppose that F and G are distributions on $(0, \infty)$. If $F \in \mathcal{S}$ and*

$$\lim_{x \rightarrow \infty} \frac{\overline{G}(s)}{\overline{F}(x)} = c \in (0, \infty),$$

then, $G \in \mathcal{S}$.

The following result is useful for obtaining the asymptotic tail probability of the sum of independent random variables having subexponential distributions.

Lemma 6.3 (Lemma A 3.28 of in [17]) *Let Y_1, Y_2, \dots, Y_m be independent r.v.s and $F \in \mathcal{S}$. Assume*

$$\lim_{x \rightarrow \infty} \frac{P(Y_i > x)}{\overline{F}(x)} = a_i \in [0, \infty], \quad i = 1, 2, \dots, m.$$

Then,

$$\lim_{x \rightarrow \infty} \frac{P(\sum_{i=1}^m Y_i > x)}{\overline{F}(x)} = \sum_{i=1}^m a_i.$$

Corollary 6.1 *If F_1 and F_2 are two distributions such that $F_1 \in \mathcal{S}$ and $\overline{F}_2(x) = o(\overline{F}_1(x))$, then $F_1 * F_2 \in \mathcal{S}$ and $\overline{F}_1 * \overline{F}_2(x) \sim \overline{F}_1(x)$ as $x \rightarrow \infty$.*

Lemma 6.4 (Geometric Sum, p. 580 in [17]) *Let Y_1, Y_2, \dots be a sequence of i.i.d. r.v.s with a common d.f. $F(x)$. Let N be an integer-valued random variable, which is independent of the sequence Y_i and has a geometric distribution $P(N = k) =$*

$(1 - p)p^k, k = 0, 1, 2, \dots$. Define $S_0 \equiv 0$ and $S_n = Y_1 + Y_2 + \dots + Y_n$ for $n \geq 1$. Then,

$$G(x) \triangleq P(S_N \leq x) = (1 - p) \sum_{k=0}^{\infty} p^k F^{*k}(x),$$

where F^{*n} denotes the n -fold convolution of F (F^{*0} is defined as the Heaviside unit step function), and the LST of $G(x)$ is given by

$$\gamma(s) = \frac{1 - p}{1 - pf(s)},$$

where $f(s)$ is the LST of $F(x)$.

Lemma 6.5 (p. 296 in Asmussen [4]) Let Y_1, Y_2, \dots be i.i.d. r.v.s with a common subexponential distribution F and let N be an integer-valued random variable, independent of the sequence Y_i , with $Ez^N < \infty$ for some $z > 1$. Then,

$$P(Y_1 + Y_2 + \dots + Y_N > x) \sim EN\bar{F}(x).$$

Specifically, if $P(N = k) = (1 - p)p^k, k = 0, 1, 2, \dots$, then

$$P(Y_1 + Y_2 + \dots + Y_N > x) \sim \frac{p}{1 - p} \bar{F}(x) \text{ as } x \rightarrow \infty.$$

The following results can be easily verified.

Lemma 6.6 (1) Let Y_i be a r.v. whose d.f. has the LST $g_i(s), i = 1, 2, \dots, n$. Define

$$X = Y_i \text{ with probability } p_i, 1 \leq i \leq n,$$

where $\sum_{i=1}^n p_i = 1$. Then, the d.f. of X has the LST $g_X(s) = \sum_{i=1}^n p_i f_i(s)$;

(2) Let Y_1, Y_2, \dots, Y_n be independent r.v.s whose d.f.s have the LST $g_1(s), g_2(s), \dots, g_n(s)$, respectively. Then, $g(s) = \prod_{i=1}^n g_i(s)$ can be viewed as the LST of the d.f. of the r.v. $X \stackrel{d}{=} \sum_{i=1}^n Y_i$.

References

1. Abate, J., Whitt, W.: Asymptotics for M/G/1 low-priority waiting-time tail probabilities. *Queueing Syst.* **25**, 173–233 (1997)
2. Andersen, L.N.: Subexponential loss rate asymptotics for Lévy processes. *Math. Meth. Oper. Res.* **73**, 91–108 (2011)
3. Andersen, L.N., Asmussen, S.: Local time asymptotics for centered Lévy processes with two-sided reflection. *Stochastic Models* **27**, 202–219 (2011)
4. Asmussen, S.: *Applied Probability and Queues*. Springer, New York (2003)
5. Asmussen, S., Klüppelberg, C., Sigman, K.: Sampling at subexponential times, with queueing applications. *Stoch. Processes Their Appl.* **79**, 265–286 (1999)

6. Asmussen, S., Pihlsgård, M.: Loss rates for Lévy processes with two reflecting barriers. *Math. Oper. Res.* **32**, 308–321 (2007)
7. Baiocchi, A.: Asymptotic behavior of the loss probability of the $M/G/1/K$ and $G/M/1/K$ queues. *Queueing Syst.* **10**, 235–248 (1992)
8. Baiocchi, A.: Analysis of the loss probability of the $MAP/G/1/K$ queue Part I: asymptotic theory. *Commun. Stat. Stoch. Models* **10**, 867–893 (1994)
9. Baltrunas, A., Daley, D.J., Klüppelberg, C.: Tail behavior of the busy period of $GI/G/1$ queue with subexponential service times. *Stoch. Processes Appl.* **111**, 237–258 (2004)
10. Borst, S.C., Boxma, O.J., Núñez-Queija, R., Zwart, A.P.: The impact of the service discipline on delay asymptotics. *Perform. Eval.* **54**, 175–206 (2003)
11. Boxma, O., Denisov, D.: Sojourn time tails in the single server queue with heavy-tailed service times. *Queueing Syst.* **69**, 101–119 (2011)
12. Boxma, O., Zwart, B.: Tails in scheduling. *ACM SIGMETRICS Perform. Eval. Rev.* **34**, 13–20 (2007)
13. Cohen, J.W.: Some results on regular variation for distributions in queueing and fluctuation theory. *J. Appl. Probab.* **10**, 343–353 (1973)
14. De Meyer, A., Teugels, J.L.: On the asymptotic behavior of the distributions of the busy period and service time in $M/G/1$. *J. Appl. Probab.* **17**, 802–813 (1980)
15. Egorova, R., Zwart, B.: Tail behavior of conditional sojourn times in processor-sharing queues. *Queueing Syst.* **55**, 107–121 (2007)
16. Egorova, R., Zwart, A.P., Boxma, O.J.: Sojourn time tails in the $M/D/1$ processor sharing queue. *Probab. Eng. Inf. Sci.* **20**, 429–446 (2006)
17. Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg (1997)
18. Foss, S., Korshunov, D., Zachary, S.: *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer, New York (2011)
19. Jelenković, P.R.: Subexponential loss rates in a $GI/GI/1$ queue with applications. *Queues with heavy-tailed distributions*. *Queueing Syst.* **33**, 91–123 (1999)
20. Jelenković, P.R., Momčilović, P.: Large deviations of square root insensitive random sums. *Math. Oper. Res.* **29**(2), 398–406 (2004)
21. Kim, J., Kim, B.: Asymptotic analysis for loss probability of queues with finite $GI/M/1$ type structure. *Queueing Syst.* **57**, 47–55 (2008)
22. Kyprianou, E.K.: On the quasi-stationary distribution of the virtual waiting time in queues with poisson arrivals. *J. Appl. Probab.* **8**, 494–507 (1971)
23. Liu, Y., Zhao, Y.Q.: Asymptotic behavior of the loss probability for an $M/G/1/N$ queue with vacations. *Appl. Math. Model.* **37**, 1768–1780 (2013)
24. Mandjes, M.R.H., Zwart, A.P.: Large deviations of sojourn times in processor sharing queues. *Queueing Syst.* **52**, 237–250 (2006)
25. Miller, L.W.: A note on the busy period of an $M/G/1$ finite queue. *Oper. Res.* **23**(6), 1179–1182 (1975)
26. Miyazawa, M., Sakuma, Y., Yamaguchi, S.: Asymptotic behaviors of the loss probability for a finite buffer queue with QBD structure. *Stoch. Models* **23**, 79–95 (2007)
27. Pakes, A.G.: On the tails of waiting-time distributions. *J. Appl. Probab.* **12**, 555–564 (1975)
28. Palmowski, Z., Rolski, T.: On the exact asymptotics of the busy period in $GI/G/1$ queues. *Adv. Appl. Probab.* **38**(3), 792–803 (2006)
29. Pihlsgård, M.: Loss rate asymptotics in a $GI/G/1$ queue with finite buffer. *Stoch. Models* **21**, 913–931 (2005)
30. Takagi, H.: *Queueing analysis: A Foundation of Performance Evaluation*, vol. 2: Finite Systems. North-Holland, Amsterdam (1993)
31. Takagi, H., LaMaire, R.O.: On the busy period of an $M/G/1/K$ queue. *Oper. Res.* **42**(1), 192–193 (1994)
32. Veraverbeke, N.: Asymptotic behavior of Wiener–Hopf factors of a random walk. *Stoch. Processes Appl.* **5**, 27–37 (1997)
33. Yashkov, S.F.: On asymptotic property of the sojourn time in the $M/G/1$ -EPS queue. *Inf. Process* **6**, 256–257 (2006). (in Russian)
34. Yashkov, S.F., Yashkova, A.S.: Processor sharing: a survey of the mathematical theory. *Autom. Remote Control* **68**, 1662–1731 (2007)
35. Zhen, Q., Knessl, C.: Asymptotic expansions for the conditional sojourn time distribution in the $M/M/1$ -PS queue. *Queueing Syst.* **57**, 157–168 (2007)

36. Zhen, Q., Knessl, C.: Asymptotic expansions for the sojourn time distribution in the $M/G/1$ -PS queue. *Math. Meth. Oper. Res.* **71**, 201–244 (2010)
37. Zwart, A.P.: A fluid queue with a finite buffer and subexponential input. *Adv. Appl. Probab.* **32**, 221–243 (2000)
38. Zwart, A.P.: Tail asymptotic for the busy period in the $GI/G/1$ queue. *Math. Oper. Res.* **36**(3), 485–493 (2001)
39. Zwart, A.P., Boxma, O.J.: Sojourn time asymptotics in the $M/G/1$ processor sharing queue. *Queueing Syst.* **35**, 141–166 (2000)