MATHEMATICS OF OPERATIONS RESEARCH
Vol. 29, No. 2, May 2004, pp. 398–406
ISSN 0364-765X | EISSN 1526-5471 | 04 | 2902 | 0398
DOI 10.1287/moor.1030.0082

Large Deviations of Square Root Insensitive Random Sums

Predrag R. Jelenković
Department of Electrical Engineering, Columbia University, New York, New York 10027, predrag@ee.columbia.edu
Petar Momčilović
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, petar@us.ibm.com

We provide a large deviation result for a random sum $\sum_{n=0}^{N_x} X_n$, where N_x is a renewal counting process and $\{X_n\}_{n\geq 0}$ are i.i.d. random variables, independent of N_x, with a common distribution that belongs to a class of square root insensitive distributions. Asymptotically, the tails of these distributions are heavier than $e^{-\sqrt{x}}$ and have zero relative decrease in intervals of length $\sqrt{x}$, hence square root insensitive. Using this result we derive the asymptotic characterization of the busy period distribution in the stable GI/G/1 queue with square root insensitive service times; this characterization further implies that the tail behavior of the busy period exhibits a functional change for distributions that are lighter than $e^{-\sqrt{x}}$.

Key words: large deviation; random sum; busy period; GI/G/1 queue; subexponential distribution; square root insensitivity
MSC2000 subject classification: Primary: 60K25
OR/MS subject classification: Primary: Queues/busy period analysis
History: Received June 3, 2002; revised June 27, 2003.

1. Introduction. We study probabilities of large deviations for random sums of variables that belong to a general subclass of subexponential distributions. This question is central to understanding many important problems of probability theory and has been extensively investigated over the years, originating with the classical results of Nagaev (1969, 1977), Heyde (1967), and Nagaev (1979). Recently, in Klüppelberg and Mikosch (1997), the authors consider large deviations of random renewal sums of variables with polynomially decaying distributions; see also Klüppelberg and Mikosch (1997) for additional references on large deviations of heavy-tailed sums. In this paper we explore the questions of Klüppelberg and Mikosch (1997) for random variables with tails lighter than any polynomial but heavier than $e^{-\sqrt{x}}$.

The Weibull tail $e^{-\sqrt{x}}$ represents a natural condition, since easy arguments show that our large deviation results do not hold for distributions lighter than $e^{-\sqrt{x}}$. The criticality of $e^{-\sqrt{x}}$ has appeared in a variety of settings, starting with early large deviation results of Nagaev (1969) and more recent analyses in Asmussen et al. (1999), Foss and Korshunov (2000), Jelenković and Momčilović (2003), and Jelenković et al. (forthcoming). This phenomenon arises from a requirement that a distribution has to tolerate Gaussian deviations of order $\sqrt{x}$ which we refer to as square root insensitivity; see Jelenković et al. (forthcoming).

The next section contains the definitions and main results of the paper. In §3 we use these results to examine the tail of the busy period in the GI/G/1 queue. The busy period is one of the primary quantities of the fundamental GI/G/1 queueing model. Its understanding is essential in addressing a long list of queueing systems, including the processor sharing (Jelenković and Momčilović 2003), generalized processor sharing (Borst et al. 2003), coupled processors (Borst et al. 2000), static priority (Abate and Whitt 1997), and fluid (Boxma and Dumas 1998) queues, as well as in estimating ruin probabilities (Asmussen and Teugels 1996) in insurance risk theory. Furthermore, our large deviation results can be applied to problems discussed in Klüppelberg and Mikosch (1997). The paper is concluded with the proof of our main result in §4.

2. Large deviations. This section contains the main results of the paper stated in Proposition 1 and Theorem 1. We consider sums of independent and identically distributed (i.i.d.) random variables $\{X, X_n,\ n \geq 0\}$ and focus on the following class of subexponential distributions $\mathcal{SC}$, first introduced in Nagaev (1977). Definitions of related classes $\mathcal{S}$ and $\mathcal{S}^*$ are given in the appendix.

Definition 1. A nonnegative random variable X (or its hazard function) belongs to class $\mathcal{SC}$ (subexponential concave) if its hazard function $Q(x) \triangleq -\log \mathbb{P}[X > x]$ is *eventually concave*, such that, $Q(x)/\log x \to \infty$ as $x \to \infty$ and for $x \geq x_0$, $\beta x \leq u \leq x$,

$$\frac{Q(x) - Q(u)}{Q(x)} \leq \alpha \frac{x-u}{x},$$

for some fixed $x_0 > 0$, $0 < \alpha < 1$ and $0 < \beta < 1$.

It is easy to see that random variables with hazard functions $(\log x)^\gamma$, $\gamma > 1$, and x^α, $0 < \alpha < 1$, i.e., lognormal and Weibull distributions, belong to $\mathcal{SC}$. We note that the assumption $Q(x)/\log x \to \infty$ ensures the finiteness of all moments for X. Basic properties of random variables in $\mathcal{SC}$ were derived in Lemma 3.1 of Jelenković and Momčilović (2003) which, for convenience, we restate here.

Throughout the paper, for any two real functions $f(x)$ and $g(x)$, we use the standard notation $f(x) \sim g(x)$ as $x \to \infty$ to denote $\lim_{x\to\infty} f(x)/g(x) = 1$.

Lemma 1. *Let $X \in \mathcal{SC}$ and Q be its hazard function; then*
(i) $Q(x) \leq Q(u)(x/u)^\alpha$ *for all* $x_0 \leq u \leq x$;
(ii) $\mathbb{P}[X > x - x^\delta] \sim \mathbb{P}[X > x]$ *as* $x \to \infty$ *for any* $0 \leq \delta < 1 - \alpha$;
(iii) $X \in \mathcal{S}^* \subseteq \mathcal{S}$;
(iv) *for any* $0 < \xi < 1$ *there is* $\delta > 0$ *such that for some* $\varepsilon > 0$ *and sufficiently large* x,

$$Q((\xi - \delta)x) + Q((1-\xi)x) \geq (1+\varepsilon)Q(x).$$

Clearly, for $\alpha < 1/2$, part (ii) of the preceding lemma implies $\mathbb{P}[X > x - \sqrt{x}] \sim \mathbb{P}[X > x]$ as $x \to \infty$; this was termed square root insensitivity in Jelenković et al. (forthcoming). Next, let $\{A, A_i, i \geq 1\}$ be a sequence of nonnegative i.i.d. random variables independent of $\{X_n\}$ with $\mathbb{E}A = \lambda^{-1}$, $\mathbb{E}A^2 < \infty$ and define N_x to be a counting process

$$N_x = \max\left\{n\colon \sum_{i=1}^{n} A_i < x\right\}. \tag{1}$$

At this point we arrive at our main result, which will be used in §3 for deriving the asymptotics of the busy period. The operators $\vee$ and $\wedge$ denote maximum and minimum, respectively.

Proposition 1. *If $\mathbb{E}A^2 < \infty$ and $X \in \mathcal{SC}$ with $\alpha < 1/2$, then, for any $0 < \delta < 1/2 - \alpha$ and $\nu > 0$, as $x \to \infty$,*

$$\mathbb{P}\left[\sum_{n=0}^{N_x} X_n - \lambda x \mathbb{E}X > \nu x,\ \bigvee_{n=0}^{N_x} X_n \leq \nu x - x^{1/2+\delta}\right] = o(\mathbb{P}[X > \nu x]).$$

Proof. Presented in §4. □

Using the preceding proposition, the next large deviation theorem follows.

Theorem 1. *If $\mathbb{E}A^2 < \infty$ and $X \in \mathcal{SC}$ with $\alpha < 1/2$, then for $\nu > 0$, as $x \to \infty$,*

$$\mathbb{P}\left[\sum_{n=0}^{N_x} X_n - \lambda x \mathbb{E}X > \nu x\right] \sim \lambda x \mathbb{P}[X > \nu x].$$

Remark 1. (i) Straightforward examination of the proof shows that the result holds assuming that the first renewal interval is almost surely finite, $A_1 < \infty$, while the remaining intervals $\{A_i, i \geq 2\}$ are i.i.d. with $\mathbb{E}A_2^2 < \infty$, and independent of A_1.

(ii) N_x does not have to be a renewal as long as its right tail is exponentially bounded, i.e., it is necessary that N_x satisfies the bound of Lemma 2 in §4.

(iii) Using the same arguments as in the following proof of the lower bound, one can show that this result fails to hold for distributions with tails lighter than $e^{-\sqrt{x}}$, i.e., the distributions that are not square root insensitive.

Proof. The upper bound is a direct consequence of Proposition 1 and square root insensitivity, i.e., Lemma 1(ii). In particular, for $0 < \delta < 1/2 - \alpha$,

$$\begin{aligned}\mathbb{P}\left[\sum_{n=0}^{N_x} X_n - \lambda x\mathbb{E}X > \nu x\right] &\leq \mathbb{P}\left[\bigvee_{n=0}^{N_x} X_n > \nu x - x^{1/2+\delta}\right]\\ &\quad + \mathbb{P}\left[\sum_{n=0}^{N_x} X_n - \lambda x\mathbb{E}X > \nu x, \bigvee_{n=0}^{N_x} X_n \leq \nu x - x^{1/2+\delta}\right]\\ &\leq (\mathbb{E}N_x + 1)\mathbb{P}[X > \nu x - x^{1/2+\delta}] + o(\mathbb{P}[X > \nu x]),\end{aligned}$$

as $x \to \infty$; in the last inequality we used the union bound as well. In proving the lower bound, for $\eta > 0$ and $x_\eta \triangleq \nu x + 2\eta\mathbb{E}X\sqrt{x}$ we derive

$$\begin{aligned}(2)\quad \mathbb{P}\left[\sum_{n=0}^{N_x} X_n - \lambda x\mathbb{E}X > \nu x\right] &\geq \mathbb{P}[N_x \geq \lceil \lambda x - \eta\sqrt{x}\rceil]\mathbb{P}\left[\sum_{n=0}^{\lceil \lambda x - \eta\sqrt{x}\rceil} X_n - \lambda x\mathbb{E}X > \nu x\right]\\ &\geq \mathbb{P}[N_x \geq \lceil \lambda x - \eta\sqrt{x}\rceil](\lambda x - \eta\sqrt{x})\mathbb{P}[X > x_\eta]\\ &\quad \cdot \mathbb{P}\left[\sum_{n=1}^{\lceil \lambda x - \eta\sqrt{x}\rceil} X_n - \lambda x\mathbb{E}X > \nu x - x_\eta, \bigvee_{n=1}^{\lceil \lambda x - \eta\sqrt{x}\rceil} X_n \leq x_\eta\right].\end{aligned}$$

Since $\mathbb{E}X^2 < \infty$, by Markov's inequality one has

$$\begin{aligned}\mathbb{P}\left[\bigvee_{n=1}^{\lceil \lambda x - \eta\sqrt{x}\rceil} X_n \leq x_\eta\right] &= (1 - \mathbb{P}[X > x_\eta])^{\lceil \lambda x - \eta\sqrt{x}\rceil}\\ &\geq \left(1 - \frac{\mathbb{E}X^2}{x_\eta^2}\right)^{\lceil \lambda x - \eta\sqrt{x}\rceil} \longrightarrow 1,\end{aligned}$$

as $x \to \infty$. Taking $\underline{\lim}$ as $x \to \infty$ in (2), using Lemma 1(ii), the Central Limit Theorem, the preceding limit, and passing $\eta \to \infty$ yields the lower bound. □

3. Busy period of the GI/G/1 queue. Investigation of the busy period of the M/G/1 queue with exponentially bounded service distributions has a long history; for recent results see Abate and Whitt (1997) and the references therein. The first analysis involving the heavy-tailed regularly varying service times has appeared in de Meyer and Teugels (1980). The derivation in de Meyer and Teugels (1980) made use of Karamata Tauberian Theory (Bingham et al. 1987) and the Poisson arrival structure. In Zwart (2001) this result was generalized for the GI/G/1 queue by developing a sample path technique that exploits the relationship between the busy period and cycle maxima. Furthermore, it was shown in Asmussen et al. (1999) that results obtained in de Meyer and Teugels (1980) and Zwart (2001) do not hold for distributions lighter than $e^{-\sqrt{x}}$.

Here we resolve the question that was left open in Zwart (2001) and Asmussen et al. (1999) by deriving the tail of the busy period distribution for a class of subexponential service times with tails heavier than $e^{-\sqrt{x}}$ but lighter than any polynomial. In addition, our result, in conjunction with Asmussen et al. (1999), shows that the asymptotic behavior of the busy period exhibits a transition in its qualitative behavior depending on the relationship of the service distribution to the Weibull tail $e^{-\sqrt{x}}$.

Without loss of generality we assume that the first (0th) customer arrives to the empty queue at time $t = 0$. Denote by B_i the service requirement of the ith customer and by A_i the interarrival time between the ith and $(i+1)$th customers. Random sequences $\{A, A_i, \ i \geq 0\}$ and $\{B, B_i, \ i \geq 0\}$ are respectively i.i.d. and independent of each other. Let $\mathbb{E}A^2 < \infty$ and N_x be a counting process as defined earlier in (1).

The amount of unfinished work in the queue at time t is denoted by V_t; for the exact definition of V_t, see, e.g., Cohen (1982). The busy period is a stopping time at which the queue becomes empty for the first time after $t = 0$, i.e.,

$$P = \inf\{t > 0\colon \ V_t = 0\}.$$

The traffic load ρ is equal to $\mathbb{E}B/\mathbb{E}A < 1$. Let K be the number of customers served during the busy period. Note that, since $\sum_{i=0}^{K-1} B_i = P$, by Wald's lemma, $\mathbb{E}K = \mathbb{E}P/\mathbb{E}B$. The expected number of customers served during the busy period can be also represented as (Cohen 1982, p. 286):

$$\mathbb{E}K = e^{\sum_{n=1}^{\infty}(1/n)\mathbb{P}[S_n > 0]},$$

where $S_n = \sum_{i=1}^{n}(B_i - A_i)$. In the case of the M/G/1 queue $\mathbb{E}K = (1-\rho)^{-1}$.

THEOREM 2. *If $\mathbb{E}A^2 < \infty$ and $B \in \mathcal{SC}$ with $\alpha < 1/2$, then as $x \to \infty$,*

$$\mathbb{P}[P > x] \sim \mathbb{E}K\,\mathbb{P}[B > (1-\rho)x].$$

REMARK 2. It is interesting to observe that the asymptotic behavior of the busy period in the M/G/∞ queue is the same for the whole class of subexponential distributions, irrespective of the relationship of the service distribution to $e^{-\sqrt{x}}$, as proved in Theorem 3.5 of Jelenković and Lazar (1999).

PROOF. The proof of the lower bound was given earlier in Zwart (2001). Thus, it remains to prove the upper bound. Denote by S the cycle maximum, i.e., $S = \sup\{V_t, \ 0 \leq t \leq P\}$. Then, following the approach in Zwart (2001), for some $0 < \delta < 1/2 - \alpha$,

$$\begin{aligned}
(3)\quad \mathbb{P}[P > x] &\leq \mathbb{P}[S > (1-\rho)x - x^{1/2+\delta}] + \mathbb{P}[P > x, S \leq (1-\rho)x - x^{1/2+\delta}] \\
&\leq \mathbb{P}[S > (1-\rho)x - x^{1/2+\delta}] + \mathbb{P}\left[\sum_{i=0}^{N_x} B_i > x, \bigvee_{i=0}^{N_x} B_i \leq (1-\rho)x - x^{1/2+\delta}\right],
\end{aligned}$$

where the second inequality follows from the facts that: (i) $\{S \leq x\}$ implies $\{B_i \leq x\}$ for all $0 \leq i \leq N_P$, (ii) $N_P \geq N_x$ on $\{P > x\}$ and (iii) $\{P > x\}$ implies, by work conservation, that $\{\sum_{i=0}^{N_x} B_i > x\}$. Next, for $B \in \mathcal{S}^*$ the distribution of the cycle maximum S is shown (Asmussen 1998) to satisfy (see also Asmussen et al. 2002), $\mathbb{P}[S > x] \sim \mathbb{E}K\,\mathbb{P}[B > x]$ as $x \to \infty$. Hence, using this fact and Lemma 1(ii), (iii), the first term in (3) satisfies

$$\overline{\lim_{x\to\infty}} \frac{\mathbb{P}[S > (1-\rho)x - x^{1/2+\delta}]}{\mathbb{P}[B > (1-\rho)x]} \leq \mathbb{E}K.$$

Thus, to complete the proof, one needs to show that the second term in (3) is $o(\mathbb{P}[B > (1-\rho)x])$ as $x \to \infty$. However, that is immediate from Proposition 1. $\square$

4. Proof of Proposition 1. The following uniform bounds play an important role in the proof of Proposition 1. In this paper C denotes a sufficiently large positive constant, while c represents a sufficiently small positive constant. The values of C and c may vary in different places, i.e., $C/2 = C$, $C^2 = C$, $C + 1 = C$, etc.

THEOREM 3. *Let $Q \in \mathcal{S}\mathcal{C}$ and $\mathbb{P}[X > x] \le Cxe^{-Q(x)}$. Then*
(i) *For all x and u,*

$$\mathbb{P}\left[\sum_{i=1}^{u} X_i - u\mathbb{E}X > x\right] \le C\big(e^{-cx^2/u} + ue^{-(1/2)Q(x)}\big).$$

(ii) *For any positive integer k there exists $0 < \gamma < 1$ such that for all $1 \le n \le Cx$,*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \wedge \gamma x - n\mathbb{E}X > x\right] \le Ce^{-kQ(x)}.$$

PROOF. See Theorem 3.2 of Jelenković and Momčilović (2003). □

LEMMA 2. *Let N_x be defined by* (1) *with $\mathbb{E}A^2 < \infty$. Then, there exists $\delta > 0$ such that for all x and $0 \le u \le \delta x$,*

$$\mathbb{P}[N_x - \lambda x > u] \le Ce^{-cu^2/x}.$$

PROOF. See Lemma 6 of Jelenković et al. (forthcoming). □

LEMMA 3. *Let Q be the hazard function of $X \in \mathcal{S}\mathcal{C}$. There exists $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$, all $n, x \ge \varepsilon n$ and $u \le (1-\varepsilon)x$, the following inequality holds*:

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \wedge u - n\mathbb{E}X > x\right] \le Ce^{-(1+\varepsilon)Q(x)}.$$

PROOF. In view of Theorem 3(ii), it is sufficient to consider only $\varepsilon x \le u \le (1-\varepsilon)x$ since otherwise the statement holds. Markov's inequality yields for $s > 0$,

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \wedge u - n\mathbb{E}X > x\right] \le e^{-s(n\mathbb{E}X + x)}\big(\mathbb{E}e^{s(X\wedge u)}\big)^n. \tag{4}$$

Next, for some $1 < \zeta < \varepsilon^{\alpha-1}$ we set $s = \zeta Q(x)/x$ and estimate the expectation in (4) as a sum of three terms:

$$\begin{aligned} \mathbb{E}e^{s(X\wedge u)} &= \int_0^{1/s} e^{sz}\, d\mathbb{P}[X \le z] + \int_{1/s}^{u} e^{sz}\, d\mathbb{P}[X \le z] + e^{su}\mathbb{P}[X > u] \\ &\le 1 + s\mathbb{E}X + s^2\mathbb{E}X^2 + \int_{1/s}^{u} e^{sz}\, d\mathbb{P}[X \le z] + e^{su - Q(u)}, \end{aligned} \tag{5}$$

where we used $e^x \le 1 + x + x^2$ on $[0, 1]$. Now, the assumption on the range of u implies $\varepsilon \le u/x \le 1 - \varepsilon$ and, hence, by the choice of s,

$$\begin{aligned} su - Q(u) &= \zeta \frac{u}{x} Q(x) - Q(u) \\ &\le \left[\zeta \frac{u}{x} - \left(\frac{u}{x}\right)^{\alpha}\right] Q(x) < -cQ(x), \end{aligned} \tag{6}$$

where the second inequality is due to Lemma 1(i) and the last bound follows from the range of ζ. The last inequality, for all u in the assumed interval, leads to $e^{su-Q(u)} \le s^2 s^{-2} e^{-cQ(x)} \le Cs^2$ (recall that by definition $Q(x)/\log x \to \infty$ as $x \to \infty$). On the other hand, integration by parts, Markov's inequality and concavity of $Q(\cdot)$ result in a bound on the integral in (5):

$$\begin{aligned}\int_{1/s}^{u} e^{sz}\, d\mathbb{P}[X \le z] &\le e\mathbb{P}[X > 1/s] + s\int_{1/s}^{u} e^{sz-Q(z)}\, dz\\ &\le s^2 e\mathbb{E}X^2 + Csx(e^{su-Q(u)} + e^{1-Q(1/s)})\\ &\le Cs^2\big(1 + x^2(e^{su-Q(u)} + e^{1-Q(1/s)})\big);\end{aligned}$$

note that the concavity of any $f(x) \ge 0$ implies $\sup_{a\le x\le b} f(x) \le f(a) + f(b)$. Hence, due to (6), the choice of s and $X \in \mathcal{SC}$ ($Q(x)/\log x \to \infty$ and $Q(x) = O(x^{\alpha})$, $0 < \alpha < 1$), the right-hand side of the preceding inequality is bounded by Cs^2. The obtained bounds, in connection with (5), yield $\mathbb{E}e^{s(X\wedge u)} \le 1 + s\mathbb{E}X + C^* s^2$, for some constant C^* and all u in the given interval. Then, by replacing this estimate in (4), using $1 + x \le e^x$ for all $x > 0$ and the definition of s, we obtain

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{n} X_i \wedge u - n\mathbb{E}X > x\right] &\le e^{-sx+nC^* s^2}\\ &\le e^{-\zeta Q(x)[1-(n/x)C^*\zeta Q(x)/x]} \le Ce^{-(1+\varepsilon)Q(x)},\end{aligned}$$

since $\zeta > 1$ and $Q(x)/\log x \to \infty$ as $x \to \infty$; this concludes the proof. □

Finally, we provide the proof of Proposition 1.

Proof of Proposition 1. In order to simplify the notation we define

$$f \triangleq \mathbb{P}\left[\sum_{i=0}^{N_x} X_i - \lambda x\mathbb{E}X > \nu x,\ \bigvee_{i=0}^{N_x} X_i \le \nu x - x^{1/2+\delta}\right].$$

The following straightforward identity represents the basis of our analysis

$$\sum_{i=0}^{N_x} X_i = \sum_{i=0}^{N_x}\left[X_i \wedge \left(\bigvee_{j=0}^{N_x} X_j\right)\right] = \bigvee_{j=0}^{N_x}\left\{\sum_{i=0}^{N_x} X_i \wedge X_j\right\}.$$

This identity, the union bound, and conditioning on X_0 yield

$$\begin{aligned}(7)\quad f &\le \mathbb{P}\left[\bigvee_{j=0}^{N_x}\left\{\sum_{i=0}^{N_x} X_i \wedge X_j \mathbf{1}_{\{X_j \le \nu x - x^{1/2+\delta}\}}\right\} - \lambda x\mathbb{E}X > \nu x\right]\\ &\le Cx\mathbb{P}\left[\sum_{i=0}^{N_x} X_i \wedge X_0 \mathbf{1}_{\{X_0 \le \nu x - x^{1/2+\delta}\}} - \lambda x\mathbb{E}X > \nu x\right] + \mathbb{P}[N_x > Cx]\\ &\le Cx\int_0^{\nu x - x^{1/2+\delta}} \mathbb{P}\left[\sum_{i=1}^{N_x} X_i \wedge u - \lambda x\mathbb{E}X > \nu x - u\right] d\mathbb{P}[X \le u] + o(\mathbb{P}[X > \nu x]),\end{aligned}$$

as $x \to \infty$; the last inequality is also due to Lemma 2 and Lemma 1(i). Next, we upper bound the integrand in the preceding inequality for all u in the interval of integration. To ease the notation, let $g(x, u) \triangleq (\nu x - u)x^{-\delta/2}$. Then, for any $0 < \xi < 1/2$ and $\varepsilon > 0$, invoking Lemma 3 (when $u \le \xi\nu x$) and Theorem 3(i) (when $u > \xi\nu x$) yields

$$\begin{aligned}(8)\ \mathbb{P}&\left[\sum_{i=1}^{N_x} X_i \wedge u - \lambda x\mathbb{E}X > \nu x - u\right]\\ &\le \mathbb{P}\left[\sum_{i=1}^{\lceil \lambda x + g(x,u)\rceil} X_i \wedge u - \lambda x\mathbb{E}X > \nu x - u\right] + \mathbb{P}[N_x > \lceil \lambda x + g(x,u)\rceil]\\ &\le C\mathbf{1}_{\{u\le \xi\nu x\}} e^{-Q(\nu x - u)} + C\mathbf{1}_{\{u > \xi\nu x\}}\big(e^{-c(\nu x - u)^2/x} + xe^{-(1/2)Q(\nu x - u - g(x,u)\mathbb{E}X - \mathbb{E}X)}\big)\\ &\quad + \mathbb{P}[N_x > \lceil \lambda x + g(x,u)\rceil]\\ &\le C\mathbf{1}_{\{u\le \xi\nu x\}} e^{-Q(\nu x - u)} + C\mathbf{1}_{\{u > \xi\nu x\}}\big(e^{-c(\nu x - u)^2/x} + xe^{-((1-\varepsilon)/2)Q(\nu x - u)}\big) + Ce^{-cg^2(x,u)/x},\end{aligned}$$

where in the last inequality we used Lemma 1(i) for the second term inside the brackets and Lemma 2 for the last term. Now, note that in (8), by the definition of $g(x,u)$, the first term in the brackets and the last term are ordered as

$$e^{-c(\nu x-u)^2/x} \le Ce^{-c(\nu x-u)^2/x^{1+\delta}} = Ce^{-cg^2(x,u)/x};$$

hence, these two terms can be combined into one. Therefore, in conjunction with (8) and (7), the upper bound on f is as follows:

$$\begin{aligned}(9)\quad f &\le Cx\int_0^{\nu x-x^{1/2+\delta}} \left(e^{-c(\nu x-u)^2/x^{1+\delta}} + \mathbf{1}_{\{u>\xi\nu x\}} x e^{-((1-\varepsilon)/2)Q(\nu x-u)} + \mathbf{1}_{\{u\le\xi\nu x\}} e^{-Q(\nu x-u)}\right) d\mathbb{P}[X\le u]\\ &\quad + o(\mathbb{P}[X>\nu x])\\ &\triangleq f_1+f_2+f_3+o(\mathbb{P}[X>\nu x]),\quad \text{as } x\to\infty.\end{aligned}$$

Integration by parts yields a bound on f_1:

$$\begin{aligned} f_1 &\le Cxe^{-cx^{1-\delta}} + Cx\int_0^{\nu x-x^{1/2+\delta}} \mathbb{P}[X>u]\, e^{-c(\nu x-u)^2/x^{1+\delta}}\, du\\ &= Cxe^{-cx^{1-\delta}} + Cxe^{-Q(\nu x)}\int_0^{\nu x-x^{1/2+\delta}} e^{Q(\nu x)-Q(u)-c(\nu x-u)^2/x^{1+\delta}}\, du.\end{aligned}$$

To establish $f_1 = o(\mathbb{P}[X>\nu x])$, in view of Lemma 1(i), it is enough to show that the exponent in the last integral is upper bounded by $-cx^\delta$ for all given u. To this end, by definition of $\mathscr{SC}$ and Lemma 1(i), for all large x,

$$\begin{aligned}(10)\qquad Q(\nu x)-Q(u)-c\frac{(\nu x-u)^2}{x^{1+\delta}} &\le Cx^\alpha\frac{\nu x-u}{x} - c\frac{(\nu x-u)^2}{x^{1+\delta}}\\ &\le Cx^{-(1/2-\alpha)+\delta} - cx^\delta,\end{aligned}$$

since for all x large enough the right-hand side of the first inequality is increasing in u and $u\le \nu x - x^{1/2+\delta}$. Now, since $\delta<1/2-\alpha$ by assumption, it follows that (10) is bounded by $-cx^\delta$.

As far as f_2 is concerned, discretizing the integral results in

$$\begin{aligned} f_2 &\le Cx^2\sum_{i=1}^{\lceil(1-\xi)\nu x^{1/2-\delta}\rceil}\int_{\nu x-(i+1)x^{1/2+\delta}}^{\nu x-ix^{1/2+\delta}} e^{-((1-\varepsilon)/2)Q(\nu x-u)}\, d\mathbb{P}[X\le u]\\ &\le Cx^2\sum_{i=1}^{\lceil(1-\xi)\nu x^{1/2-\delta}\rceil} e^{-((1-\varepsilon)/2)Q(ix^{1/2+\delta})-Q(\nu x-(i+1)x^{1/2+\delta})}\\ &\le Cx^{5/2-\delta}e^{-((1-\varepsilon)/2)Q(x^{1/2+\delta})-Q(\nu x-2x^{1/2+\delta})} \vee Cx^{5/2-\delta}e^{-((1-\varepsilon)/2)Q((1-\xi)\nu x)-Q(\xi\nu x-2x^{1/2+\delta})},\end{aligned}$$

where the last inequality follows from the concavity property of $Q(\cdot)$; i.e., the maximum of all summands is equal to either the first or the last summand. Thus, Lemma 1(i) and (ii) imply that the first term in the maximum is $o(\mathbb{P}[X>\nu x])$ as $x\to\infty$; the exponent of the second term is by Lemma 1(i) bounded by (for large x)

$$\frac{1-\varepsilon}{2}Q((1-\xi)\nu x)+Q(\xi\nu x-2x^{1/2+\delta}) \ge Q(\nu x)\left(\frac{1-\varepsilon}{2}(1-\xi)^\alpha+(\xi-\varepsilon)^\alpha\right).$$

Next, it is easy to verify that for any $\xi>(3/5)^2$ (recall that $\alpha<1/2$ by assumption), we can choose $\varepsilon>0$ sufficiently small such that $((1-\varepsilon)/2)(1-\xi)^\alpha+(\xi-\varepsilon)^\alpha>1$, and, thus we have $f_2 = o(\mathbb{P}[X>\nu x])$ as $x\to\infty$.

Finally, for some $\Delta > 0$, we estimate

$$\begin{aligned} f_3 &\le Cx \sum_{i=1}^{\lceil \xi\nu/\Delta \rceil} \int_{(i-1)\Delta x}^{i\Delta x} e^{-Q(\nu x - u)}\, d\mathbb{P}[X \le u] \\ &\le Cx \sum_{i=1}^{\lceil \xi\nu/\Delta \rceil} e^{-Q(\nu x - i\Delta x) - Q((i-1)\Delta x)} \end{aligned}$$

and from Lemma 1(iv) it follows that, if Δ is chosen to be sufficiently small, then each summand in the preceding sum is $o(\mathbb{P}[X > \nu x])$ as $x \to \infty$; therefore, $f_3 = o(\mathbb{P}[X > \nu x])$ as $x \to \infty$.

Replacing the preceding bounds on f_1, f_2, and f_3 in (9) yields the proof. □

Appendix.

Definition 2. A nonnegative random variable X is called *subexponential*, $X \in \mathcal{S}$, if

$$\lim_{x\to\infty} \frac{\mathbb{P}[X_1 + X_2 > x]}{\mathbb{P}[X > x]} = 2,$$

where X_1 and X_2 are independent copies of X.

Definition 3. A nonnegative random variable X *belongs to class* $\mathcal{S}^*$, $X \in \mathcal{S}^*$, if X has finite expectation and

$$\lim_{x\to\infty} \int_0^x \frac{\mathbb{P}[X > x - y]}{\mathbb{P}[X > x]} \mathbb{P}[X > y]\, dy = 2\mathbb{E}X.$$

Acknowledgment. The authors would like to thank the anonymous reviewers for comments and suggestions. This work is supported by the NSF PECASE Award No. 9875156.

References

Abate, J., W. Whitt. 1997. Asymptotics for M/G/1 low-priority waiting-time tail probabilities. *Queueing Systems Theory Appl.* **25**(1–4) 173–223.

Asmussen, S. 1998. Subexponential asymptotics for stochastic processes: Extremal behavior, stationary distributions and first passage probabilities. *Ann. Appl. Probab.* **8**(2) 354–374.

Asmussen, S., V. Kalashnikov, D. Konstantinides, C. Klüppelberg, G. Tsitiashvili. 2002. A local limit theorem for random walk maxima with heavy tails. *Statist. Probab. Lett.* **56**(4) 399–404.

Asmussen, S., C. Klüppelberg, K. Sigman. 1999. Sampling at subexponential times, with queueing applications. *Stochastic Process. Appl.* **79** 265–286.

Asmussen, S., J. Teugels. 1996. Convergence rates for M/G/1 queues and ruin problems with heavy tails. *J. Appl. Probab.* **33** 1181–1190.

Bingham, N. H., C. M. Goldie, J. L. Teugels. 1987. *Regular Variation*. Cambridge University Press, Cambridge, MA.

Borst, S., O. Boxma, P. Jelenković. 2000. Coupled processors with regularly varying service times. *Proc. IEEE Infocom*. Tel Aviv, Israel.

Borst, S., O. Boxma, P. Jelenković. 2003. Reduced-load equivalence and induced burstiness in GPS queues with long-tailed traffic flows. *Queueing Systems Theory Appl.* **43**(4) 273–306.

Boxma, O., V. Dumas. 1998. The busy period in the fluid queue. *Performance Evaluation Rev.* **26** 100–110.

Cohen, J. W. 1982. *The Single Server Queue*. North-Holland, Amsterdam, The Netherlands.

de Meyer, A., 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.

Foss, S., D. Korshunov. 2000. Sampling at random time with a heavy-tailed distribution. *Markov Process Related Fields* **6** 543–568.

Heyde, C. C. 1967. A contribution to the theory of large deviations for sums of independent random variables. *Z. Wahr. verw. Gebiete* **7** 303–308.

Jelenković, P., A. Lazar. 1999. Asymptotic results for multiplexing subexponential on-off processes. *Adv. Appl. Probab.* **31**(2) 394–421.

Jelenković, P., P. Momčilović. 2003. Large deviation analysis of subexponential waiting times in a processor sharing queue. *Math. Oper. Res.* **28**(3) 587–608.

Jelenković, P., P. Momčilović, B. Zwart. Reduced load equivalence under subexponentiality. *Queueing Systems Theory Appl.* Forthcoming.

Klüppelberg, C., T. Mikosch. 1997. Large deviations of heavy-tailed random sums with applications in insurance and finance. *J. Appl. Probab.* **34** 293–308.

Nagaev, A. V. 1969. Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I, II. *Theory Probab. Appl.* **14** 51–64, 193–208.

Nagaev, A. V. 1977. On a property of sums of independent random variables. *Theory Probab. Appl.* **22**(2) 326–338.

Nagaev, S. V. 1979. Large deviations of sums of independent random variables. *Ann. Probab.* **7**(5) 745–789.

Zwart, B. 2001. Tail asymptotics for the busy period in the GI/G/1 queue. *Math. Oper. Res.* **26**(3) 485–493.