EC.1. Deferred Proofs for Tail Asymptotics of SOAP Policies

EC.1.1. Proofs for Heavy-Tailed Job Sizes

LEMMA 7.6. *Suppose Condition 4.5 holds.*

(i) For all $p \geq 0$,

$$\mathbf{E}[X_0[w_x]^{p+1}] \leq \begin{cases} O(1) & \text{if } p < \alpha - 1 \\ O(\log x) & \text{if } p = \alpha - 1 \\ O(x^{\max\{1,\zeta+\theta\}(p-\alpha+1)}) & \text{if } p > \alpha - 1. \end{cases}$$

(ii) For all $p \geq 0$,

$$\sum_{k=1}^{K[w_x]} \mathbf{E}[X_k[w_x]^{p+1}] \leq \begin{cases} O(x^{\theta p+\zeta p-\alpha+1}) & \text{if } \zeta p < \alpha - 1 \\ O(x^{\theta p} \log x^{\eta}) & \text{if } \zeta p = \alpha - 1 \\ O(x^{\theta p+\eta(\zeta p-\alpha+1)}) & \text{if } \zeta p > \alpha - 1. \end{cases}$$

Proof. We first show (i). Because $(x, c_0[w_x])$ is a w_x-interval, Condition 4.5 implies

$$c_0[w_x] - x = O(x^{\zeta+\theta}). \tag{EC.1.1}$$

We compute

$$\begin{aligned} \mathbf{E}[X_0[w_x]^{p+1}] &= \int_0^{c_0[w_x]} (p+1)t^p \overline{F}(t)\,\mathrm{d}t && \text{[by Lemma 6.3]} \\ &\leq \int_0^{O(x^{\max\{1,\zeta+\theta\}})} O(t^{p-\alpha})\,\mathrm{d}t && \text{[by Definition 4.1 and (EC.1.1)]} \\ &= \begin{cases} O(1) & \text{if } p < \alpha - 1 \\ O(\log x) & \text{if } p = \alpha - 1 \\ O(x^{\max\{1,\zeta+\theta\}(p-\alpha+1)}) & \text{if } p > \alpha - 1, \end{cases} \end{aligned}$$

thus proving (i).

We now show (ii), following a similar argument but with a more involved computation. Note that Definitions 6.1 and 6.5 together imply

$$b_k[w_x] \geq x \quad \text{for all } k \geq 1. \tag{EC.1.2}$$

We compute

$$\begin{aligned} \sum_{k=1}^{K[w_x]} \mathbf{E}[X_k[w_x]^{p+1}] &= \sum_{k=1}^{K[w_x]} \int_{b_k[w_x]}^{c_k[w_x]} (p+1)(t - b_k[w_x])^p \overline{F}(t)\,\mathrm{d}t && \text{[by Lemma 6.3]} \\ &\leq \sum_{k=1}^{K[w_x]} (p+1)(c_k[w_x] - b_k[w_x])^p \int_{b_k[w_x]}^{c_k[w_x]} \overline{F}(t)\,\mathrm{d}t && \text{[by (EC.1.2)]} \end{aligned}$$

$$
\begin{aligned}
&\le \sum_{k=1}^{K[w_x]} O(x^{\theta p} \cdot b_k[w_x]^{\zeta p}) \int_{b_k[w_x]}^{c_k[w_x]} O(t^{-\alpha})\,\mathrm{d}t && \text{[by Definition 4.1 and Condition 4.5]}\\
&\le \sum_{k=1}^{K[w_x]} O(x^{\theta p}) \int_{b_k[w_x]}^{c_k[w_x]} O(t^{\zeta p-\alpha})\,\mathrm{d}t\\
&\le O(x^{\theta p}) \int_{x}^{c_{K[w_x]}[w_x]} O(t^{\zeta p-\alpha})\,\mathrm{d}t && \text{[by (EC.1.2)]}\\
&\le O(x^{\theta p}) \int_{x}^{O(x^{\eta})} O(t^{\zeta p-\alpha})\,\mathrm{d}t && \text{[by Condition 4.5]}\\
&= \begin{cases} O(x^{\theta p+\zeta p-\alpha+1}) & \text{if } \zeta p < \alpha-1\\ O(x^{\theta p}\log x^{\eta}) & \text{if } \zeta p = \alpha-1\\ O(x^{\theta p+\eta(\zeta p-\alpha+1)}) & \text{if } \zeta p > \alpha-1, \end{cases}
\end{aligned}
$$

thus proving (ii). □

LEMMA 7.9. *Suppose Condition 4.5 holds, and let* $\kappa = 2\max\{\alpha-1,\theta\}$. *For all* $x \ge 0$ *and* $a \in (y_x, x)$,

$$c_0[w_x(a)-] \ge \Omega\left(\left(\frac{x-a}{x^{\zeta}}\right)^{1/\kappa}\right).$$

Proof. Because $\kappa > \theta \ge 0$, by Condition 4.5 and (EC.1.2), for all $u \ge 0$ and $k \ge 1$,

$$u \ge \Omega\left(\left(\frac{c_k[w_u]-b_k[w_u]}{b_k[w_u]^{\zeta}}\right)^{1/\kappa}\right). \tag{EC.1.3}$$

We now plug in $u = c_0[w_x(a)-]$ and make the following observations.

- By Definition 6.1, we know $u = c_0[w_x(a)-]$ is the earliest age at which a job has rank at least $w_x(a)$, so $w_u = w_x(a)$.
- By Definition 6.5, a job's rank is at most $w_x(a)$ between ages a and x, so there exists $k \ge 1$ such that

  $$b_k[w_x(a)] \le a < x \le c_k[w_x(a)].$$

  In particular, $x > b_k[w_x(a)]$ and $x - a \le c_k[w_x(a)] - b_k[w_x(a)]$.

Applying these observations to (EC.1.3) with $u = c_0[w_x(a)-]$ yields the desired bound. □

EC.1.2. Proofs for Light-Tailed Job Sizes

PROPOSITION 9.2. *Consider an M/G/1 with any nicely light-tailed job size distribution under a SOAP policy. The policy is log-tail-pessimal if* $a^* = x_{\max}$.

Proof. Since no work-conserving policy has response time decay rate lower than a busy period's decay rate $d(B)$ (Mandjes and Nuyens, 2005, Corollary 6), it suffices to show $d(T) \le d(B)$.

Recall that y_x denotes the (first) age of the maximum rank in the interval $[0, x]$. Since $T(x)$ is stochastically increasing in x (Lemma 6.6), it holds that $\mathbf{P}[T(x) > t] \ge \mathbf{P}[T(y_x) > t]$ for all $t \ge 0$. Additionally we have

that $\mathbf{P}[T(y_x) > t] \geq \mathbf{P}[T_{\text{FB}}(y_x) > t]$ for all $t, x \geq 0$, where $T_{\text{FB}}(x)$ is the response time for a job of size x under FB. The reason for this last inequality is that a job of size y_x must wait for all other jobs to receive up to y_x units of service before completing.[1] As a result, (Mandjes and Nuyens, 2005, Proposition 8) implies

$$d(T(x)) \leq d(T(y_x)) \leq d(T_{\text{FB}}(y_x)) = d(B_{y_x}).$$

Additionally, up to its last line the proof of (Mandjes and Nuyens, 2005, Lemma 9) is valid for arbitrary service policies. If $x_0 > 0$ is such that $\mathbf{P}[X \geq x_0] > 0$, we thus find

$$\begin{aligned} d(T) &\leq \mathbf{P}[X \geq x_0]^{-1} \int_{x_0}^{x_{\max}} d(T(x))\, dF(x) \\ &\leq \mathbf{P}[X \geq x_0]^{-1} \int_{x_0}^{x_{\max}} d(B_{y_x})\, dF(x). \end{aligned} \tag{EC.1.4}$$

Our goal is to show $d(T) \leq d(B)$, or equivalently $d(T) < d(B) + \varepsilon$ for all $\varepsilon > 0$. By (EC.1.4), it suffices to show that $\lim_{x \to x_{\max}} d(B_{y_x}) = d(B)$. It is shown in (Mandjes and Nuyens, 2005, Lemma 10) that $\lim_{x \to x_{\max}} d(B_x) = d(B)$, so our task is to show that the limit still holds with y_x instead of x.

Consider arbitrary $\varepsilon > 0$. Because $\lim_{x \to \infty} d(B_x) = d(B)$, there exists $x_0 > 0$ such that $|d(B_x) - d(B)| < \varepsilon$ for all $x > x_0$. Because $a^* = x_{\max}$, there exists $x_1 > x_0$ such that $y_{x_1} = x_1$, and thus $|d(B_{y_{x_1}}) - d(B)| < \varepsilon$. But $d(B_{y_x})$ is decreasing in x, because y_x is increasing in x, and B_x is stochastically increasing in x. We conclude that for all $x > x_1$, we have $|d(B_{y_x}) - d(B)| < \varepsilon$. Our choice of $\varepsilon > 0$ was arbitrary, so $\lim_{x \to x_{\max}} d(B_{y_x}) = d(B)$, as desired. □

LEMMA 9.5. *Let π be a SOAP policy with $0 < a^* < x_{\max}$. We have*

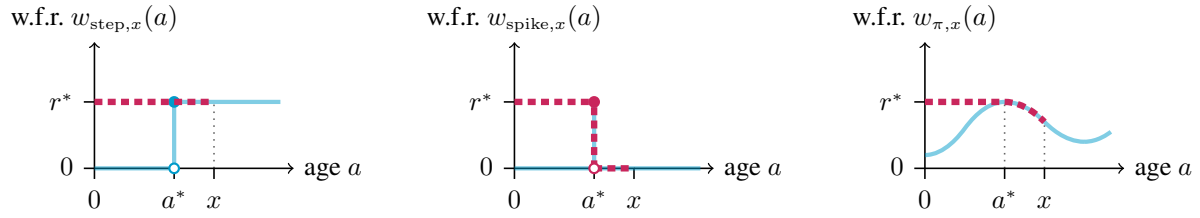
$$d(T_\pi) = d(T_\pi^{(2)}) \in [d(T_{\text{step}}^{(2)}), d(T_{\text{spike}}^{(2)})].$$

Proof. Clearly, T_π is a mixture of $T_\pi^{(1)}$ and $T_\pi^{(2)}$. Lemma 6.6 implies $T_\pi^{(2)} \geq_{\text{st}} T_\pi^{(1)}$, implying $d(T_\pi) = d(T_\pi^{(2)})$. The same reasoning applies to Step and Spike. The lemma thus follows if we can show

$$T_{\text{spike}}^{(2)} \leq_{\text{st}} T_\pi^{(2)} \leq_{\text{st}} T_{\text{step}}^{(2)}. \tag{EC.1.5}$$

The comparison in (EC.1.5) follows from a key fact from the SOAP analysis (Scully and Harchol-Balter, 2018) called the *Pessimism Principle*, which states that the response time of a particular job J is unaffected if, instead of following the usual rank function, job J follows its *worst future rank* function (Definition 6.5). The intuition is that any jobs that will get served ahead of job J in the future may as well be served ahead of it right now.

[1] One can give a more formal proof of the inequality using the SOAP analysis (Scully and Harchol-Balter, 2018).



(a) Worst future rank of Step. **(b)** Worst future rank of Spike. **(c)** Worst future rank of generic policy.

Figure EC.1 Worst future rank functions (Definition 6.5, abbreviated w.f.r., dotted magenta curves) of the policies shown in Figure 9.1, with the original rank functions (translucent cyan curves) for reference. We show the worst future rank functions for a class 2 job of size $x > a^*$.

We illustrate in Figure EC.1 the worst future rank under Step, Spike, and π. Notice that, for any size $x > a^*$, we have

$$a \in [0, a^*] \quad \Rightarrow \quad r^* = w_{\text{spike},x}(a) = w_{\pi,x}(a) = w_{\text{step},x}(a) = r^*,$$
$$a \in (a^*, x) \quad \Rightarrow \quad 0 = w_{\text{spike},x}(a) \le w_{\pi,x}(a) \le w_{\text{step},x}(a) = r^*.$$

The Pessimism Principle says that we can compute a particular job J's response time by imagining that it always has its worst future rank. Increasing a job's rank can only increase its response time, so the above worst future rank comparisons imply that for all $x > a^*$,

$$T_{\text{spike}}(x) \le_{\text{st}} T_\pi(x) \le_{\text{st}} T_{\text{step}}(x).$$

The desired (EC.1.5) follows because class 2 jobs are those of size greater than a^*. □

LEMMA 9.6. *The response time distributions of class 2 jobs under Step and Spike are*

$$T^{(2)}_{\text{step}} =_{\text{st}} B_{a^*}(W) + B_{a^*}(X^{(2)}), \qquad T^{(2)}_{\text{spike}} =_{\text{st}} B_{a^*}(W) + B_{a^*}(a^*) + X^{(2)} - a^*.$$

where $X^{(2)} = (X \mid X > a^*)$ is the size distribution of class 2 jobs, and the random variables in each sum are mutually independent.

Proof. This result follows easily from the SOAP analysis (Scully and Harchol-Balter, 2018). For completeness, we sketch the main ideas of how the SOAP analysis applies to Step and Spike. Consider a class 2 job J.

- Under Step, job J always has worst future rank r^* (Figure EC.1(a)). Job J is thus delayed by any jobs present when it arrives, plus the pre-age-a^* portion of any jobs that arrive while it is in the system.
- Under Spike, job J has worst future rank r^* only until age a^* (Figure EC.1(a)). Job J is thus delayed by any jobs present when it arrives, plus the pre-age-a^* portion of any jobs that arrive before it reaches age a^*. Once job J reaches age a^*, its worst future rank is 0, so no further arrivals delay it.

The reason in both cases for looking at the pre-age-a^* portion of new arrivals is because at age a^*, those new arrivals reach rank r^*, and thus job J has priority over them due to FCFS tiebreaking (Definition 3.1).

The delay due to jobs present when job J arrives corresponds to the W in each formula, and the delay due to new arrivals corresponds to the $B_{a^*}(\cdot)$ uses. The difference between the formulas is due to the fact that under Step, new arrivals delay job J until it completes, whereas under Spike, new arrivals delay job J only if they arrive before it reaches age a^*, with the last $X^{(2)} - a^*$ portion of job J's service occurring uninterrupted. □

LEMMA 9.7. *Consider an M/G/1 with nicely light-tailed job size distribution X, and define*

$$s_0 = \gamma(\sigma^{-1}) = \gamma(\mathcal{L}[X]), \qquad s_2 = \sigma(\gamma(\sigma)) = \arg\min_{s \geq s_0} \sigma^{-1}(s),$$
$$s_1 = \gamma(\mathcal{L}[W]) = \text{least root of } \sigma^{-1}, \qquad s_3 = \gamma(\sigma) = \min_{s \geq s_0} \sigma^{-1}(s).$$

Then, as illustrated in Figure 9.2, the following hold:

(i) *σ^{-1} is convex on (s_0, ∞), decreasing on (s_0, s_2), and increasing on (s_2, ∞);*

(ii) *$s_0 < s_1 < s_2 < s_3 < 0$.*

Analogous statements hold for σ_a for all $a \in (0, x_{\max}]$.

Proof. We prove the statement just for σ, as the argument for σ_a is analogous. The illustration in Figure 9.2 may provide helpful intuition for the arguments that follow.

We begin by observing some general properties of σ^{-1}. Because $\mathcal{L}[X]$ is convex on (s_0, ∞), so is σ^{-1}. This, along with the definition of s_2, implies (a). The slope of σ^{-1} at zero is

$$(\sigma^{-1})'(0) = 1 + \lambda \mathcal{L}[X]'(0) = 1 - \rho \in (0, 1),$$

and by Definition 5.2, we have $\sigma^{-1}(s_0) = \infty > 0$. Additionally, Definition 5.2 implies $s_0 < 0$. This means σ^{-1} is negative on a finite nonempty interval, namely $(s_1, 0)$, and nonnegative outside that interval.

We can now show the inequalities in (b).

- $s_0 < s_1$: Because $|\sigma^{-1}(s_1)| = 0 < \infty$, we have $s_0 = \gamma(\sigma^{-1}) \leq s_1$. But $\sigma^{-1}(s_0) > 0$, so $s_0 \neq s_1$.
- $s_3 < 0$: Because σ^{-1} is negative on some interval, its global minimum is negative.
- $s_1 < s_2$: Because $s_3 = \sigma^{-1}(s_2) < 0$, we must have $s_2 \in (s_1, 0)$.
- $s_2 < s_3$: Because σ^{-1} is convex with $\sigma^{-1}(0) = 0$ and $(\sigma^{-1})'(0) \in (0, 1)$, we have $s_2 < \sigma^{-1}(s_2)$. □

In some of the proofs below, we use the fact that for sums of independent random variables $U, V \geq 0$, (9.1) implies

$$\begin{aligned} d(U+V) &= -\gamma(\mathcal{L}[U+V]) \\ &= -\max\{\gamma(\mathcal{L}[U]), \gamma(\mathcal{L}[V])\} \\ &= \min\{d(U), d(V)\}. \end{aligned} \tag{EC.1.6}$$

This is also shown by Mandjes and Nuyens (2005, Lemma 3) without relying on (9.1).

LEMMA 9.8. *Let π be a SOAP policy with $0 < a^* < x_{\max}$. Then the decay rate of its response time is*

$$d(T_\pi) = -\gamma(\mathcal{L}[W] \circ \sigma_{a^*}).$$

Proof. Combining Lemmas 9.5 and 9.6, we have

$$d(T_\pi) \in \big[d(B_{a^*}(W)), d\big(B_{a^*}(W) + B_{a^*}(X^{(2)})\big)\big].$$

By (9.1) and (9.2), the lower bound is

$$d(B_{a^*}(W)) = -\gamma(\mathcal{L}[W] \circ \sigma_{a^*}).$$

We aim to show that the upper bound matches this. Applying (9.1), (9.2), and (EC.1.6) to the upper bound, we see that it suffices to show

$$\gamma(\mathcal{L}[X^{(2)}] \circ \sigma_{a^*}) \le \gamma(\mathcal{L}[W] \circ \sigma_{a^*}).$$

Lemma EC.1.1, which we state and prove below, implies the above if $\gamma(\mathcal{L}[X^{(2)}]) \le \gamma(\mathcal{L}[W])$, which in turn is implied by Lemma 9.7(ii) and the fact that $\gamma(\mathcal{L}[X^{(2)}]) = \gamma(\mathcal{L}[X])$. □

The following lemma, which is used in the proof above, relates $\gamma(f \circ \sigma)$ to $\gamma(f)$, thus relating the decay rate of a busy period to the decay rate of its initial work.

LEMMA EC.1.1. *Let $f : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ be a function for which $\gamma(f)$ is well defined and finite. Then $\gamma(f \circ \sigma)$ is finite, and*

$$\begin{aligned}\gamma(f \circ \sigma) &= \sigma^{-1}\big(\max\big\{\gamma(f), \sigma(\gamma(\sigma))\big\}\big) \\ &= \begin{cases} \sigma^{-1}(\gamma(f)) & \text{if } \gamma(f) > \sigma(\gamma(\sigma)) \\ \gamma(\sigma) & \text{otherwise.} \end{cases}\end{aligned}$$

In particular, $\gamma(f \circ \sigma)$ is a nondecreasing function of $\gamma(f)$. Analogous statements hold for σ_a for all $a \in [0, x_{\max}]$.

Proof. We prove the statement just for σ, as the proof for σ_a is analogous. There are two reasons $f(\sigma(s))$ can be infinite:

- We can have $\sigma(s)$ infinite, which happens if and only if $s < \gamma(\sigma)$.
- We can have $\sigma(s)$ finite but $f(\sigma(s))$ infinite, which happens if $-\infty < \sigma(s) < \gamma(f)$ and only if $-\infty < \sigma(s) \le \gamma(f)$.

Recalling that $\sigma(\gamma(\sigma))$ is the minimum finite value $\sigma(s)$ can take on (see Figure 9.2), we see that the latter reason can occur for some $s > \gamma(\sigma)$ if and only if $\sigma(\gamma(\sigma)) < \gamma(f)$, implying the desired formula.

The finiteness of $\gamma(f \circ \sigma)$ follows from finiteness of $\sigma^{-1}(\gamma(f))$ when $\gamma(f) > \gamma(\sigma)$, which by Lemma 9.7(ii) includes all cases when $\gamma(f) > \sigma(\gamma(\sigma))$. The monotonicity of $\gamma(f \circ \sigma)$ in $\gamma(f)$ follows Lemma 9.7(i). □

PROPOSITION 9.9. *Consider an M/G/1 with any nicely light-tailed job size distribution under a SOAP policy. The policy is log-tail-intermediate if* $0 < a^* < x_{\max}$.

Proof. The optimal decay rate is that of FCFS. A special case of a result of Stolyar and Ramanan (2001, Theorem 2.2), together with (9.1) and (EC.1.6) implies this is

$$d(T_{\mathrm{FCFS}}) = d(W + X) = d(W) = -\gamma(\mathcal{L}[W]).$$

The pessimal decay rate is that of FB. A result of Mandjes and Nuyens (2005, Theorem 1) states $d(T_{\mathrm{FB}}) = d(B)$. Together with (9.1) and (9.2) and Lemma EC.1.1, this implies

$$\begin{aligned} d(T_{\mathrm{FB}}) = d(B) &= -\gamma(\mathcal{L}[X] \circ \sigma) \\ &= -\gamma(\sigma) = -\gamma(\mathcal{L}[W] \circ \sigma). \end{aligned}$$

Above, we use the fact that $\gamma(\mathcal{L}[X]) < \gamma(\mathcal{L}[W]) < \sigma(\gamma(\sigma))$, as shown in Lemma 9.7(ii), when applying Lemma EC.1.1.

Having computed the optimal and pessimal decay rates in Lemma 9.8, it suffices to show that in the $0 < a^* < x_{\max}$ case, we have

$$\gamma(\mathcal{L}[W]) < \gamma(\mathcal{L}[W] \circ \sigma_{a^*}) < \gamma(\mathcal{L}[W] \circ \sigma),$$

which we may rewrite as

$$\gamma(\mathcal{L}[W] \circ \sigma_0) < \gamma(\mathcal{L}[W] \circ \sigma_{a^*}) < \gamma(\mathcal{L}[W] \circ \sigma_{x_{\max}}).$$

Lemma EC.1.2, which we state and prove below, implies $\gamma(\mathcal{L}[W] \circ \sigma_a)$ is strictly increasing in a. Therefore, the above holds if $0 < a^* < x_{\max}$, as desired. □

It remains only to prove the strict monotonicity of $\gamma(\mathcal{L}[W] \circ \sigma_a)$ in a. We prove a more general statement below.

LEMMA EC.1.2. *Let* $f : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ *be a function for which* $\gamma(f) < 0$*, and let* $0 \le a < b \le x_{\max}$*. Then*

$$\gamma(f \circ \sigma_a) < \gamma(f \circ \sigma_b).$$

Proof. We begin by comparing $\sigma_a^{-1}(s)$ with $\sigma_b^{-1}(S)$ for all $s < 0$, computing[2]

$$\begin{aligned} & a < b \\ \Rightarrow \quad & \min\{X, a\} <_{\mathrm{st}} \min\{X, b\} \\ \Rightarrow \quad & \mathcal{L}[\min\{X, a\}](s) < \mathcal{L}[\min\{X, b\}](s) \\ \Rightarrow \quad & \sigma_a^{-1}(s) < \sigma_b^{-1}(s). \end{aligned} \tag{EC.1.7}$$

[2] Two clarifications about the computation below. First, the notation $U <_{\mathrm{st}} V$ means that $\mathbf{P}[U > t] \le \mathbf{P}[V > t]$ for all $t \in \mathbb{R}$, and the set of $t \in \mathbb{R}$ such that $\mathbf{P}[U > t] < \mathbf{P}[V > t]$ has positive Lebesgue measure. Second, because $a < \infty$, the left-hand sides of the last two steps are always finite for all $s < 0$.

There are two important implications of (EC.1.7). The first implication is that the global minimum of σ_a^{-1} is less than that of σ_b^{-1}. But these global minima are $\gamma(\sigma_a)$ and $\gamma(\sigma_b)$, respectively (see Figure 9.2), so

$$\gamma(\sigma_a) < \gamma(\sigma_b). \tag{EC.1.8}$$

This means $\sigma_a(s)$ is finite whenever $\sigma_b(s)$ is. This contributes to the second implication of (EC.1.7): by Lemma 9.7(i),

$$\sigma_a(s) > \sigma_b(s) \quad \text{for all } s \in [\gamma(\sigma_b), 0). \tag{EC.1.9}$$

Note that $f(\sigma_a(s))$ diverges only if $s \le \gamma(\sigma_a)$ or $\sigma_a(s) \le \gamma(f)$, while $f(\sigma_b(s))$ diverges if $s < \gamma(\sigma_b)$ or $\sigma_b(s) < \gamma(f)$. Therefore, (EC.1.8) and (EC.1.9) together imply that there exists a value of s such that $f(\sigma_b(s))$ diverges while $f(\sigma_a(s))$ does not. □

EC.2. Properties of the Gittins Policy via the "Gittins Game"

The goal of this section is to prove two key remaining properties of the Gittins policy, Theorem 5.10 and Lemma 8.2. To prove both of these properties, we will use a different perspective on the Gittins policy called the "Gittins game" (Scully et al., 2020a). The Gittins game gives an alternative way to define the Gittins rank function. While it is less direct than the definitions we have used so far (Definitions 3.2 and 8.1), the intermediate steps it introduces turn out to be crucial for proving Theorem 5.10 and Lemma 8.2.

Aside from Theorem 5.10 and Lemma 8.2, most of the definitions and results in this section are due to Scully et al. (2020a), who actually study a much more general job model than ours. For simplicity, we restate the key definitions and results in our setting. However, the statements and proofs of Theorem 5.10 and Lemma 8.2 are straightforward to translate to the more general job model of Scully et al. (2020a).

EC.2.1. The Gittins Game

The Gittins game is an optimization problem. Its inputs are a job at some age b and a *penalty* w. During the game, we serve the job for as long as we like. If the job completes, the game ends. At any moment before the job completes, we may choose to give up, in which case we pay the penalty w and the game immediately ends. The goal of the game is to minimize the expected sum of the time spent serving the job plus the penalty paid.

We can think of the Gittins game with penalty w as an optimal stopping problem whose state is the age b of the job. Standard optimal stopping theory (Peskir and Shiryaev, 2006; Shiryaev, 2008) implies that the optimal strategy thus has the following form: serve the job until it reaches some age $c \ge b$, then give up. A possible policy here is never giving up, which is represented by $c = \infty$.

Suppose we start serving a job at age b and stop if it reaches age c. The expected amount of time we spend serving the job is

$$\mathsf{service}(b, c) = \mathbf{E}[\min\{S, c\} \mid S > b] = \int_b^c \frac{\overline{F}(t)}{\overline{F}(b)}\,\mathrm{d}t,$$

and the probability the job finishes before reaching age b is

$$\mathsf{done}(b,c) = \mathbf{P}[S \le c \mid S > b] = 1 - \frac{\overline{F}(c)}{\overline{F}(b)}.$$

We can write the time-per-completion function as $\varphi(b,c) = \mathsf{service}(b,c)/\,\mathsf{done}(b,c)$ (see Definition 8.1).

Suppose we employ the stop-at-age-c policy in the Gittins game starting from age b with penalty w. The expected cost of the Gittins game with this policy is

$$\mathsf{game}(w;b,c) = \mathsf{service}(b,c) + w(1 - \mathsf{done}(b,c))$$

The *optimal* cost of the Gittins game is therefore

$$\mathsf{game}^*(w;b) = \inf_{c \ge b} \mathsf{game}(w;b,c).$$

The lemma below follows immediately from the definition of $\mathsf{game}^*(w;b)$ as an infimum of $\mathsf{game}(w;b,c)$, each of which is a linear function of w (Scully et al., 2020a, Lemmas 5.2 and 5.3).

LEMMA EC.2.1. *For all ages b, the optimal cost $\mathsf{game}^*(w;b)$ is increasing and concave as a function of w. Because giving up immediately is always a possible policy, it is also bounded above by $\mathsf{game}^*(w;b) \le w$.*

EC.2.2. Relating the Gittins Game to the Gittins Rank Function

The Gittins game is intimately connected to the Gittins rank function, and it is this connection that is important for proving Lemma 8.2. The following lemmas state two such connections. They are the same or very similar to many previous results in the literature on Gittins in the M/G/1 (Aalto et al., 2009; Gittins, 1989; Gittins et al., 2011; Scully et al., 2020a, 2018a), but we sketch their proofs for completeness.

LEMMA EC.2.2. *The Gittins rank function can be expressed in terms of the Gittins game as*

$$\begin{aligned} r(a) &= \inf\{w \ge 0 \mid \mathsf{game}^*(w;a) < w\} \\ &= \max\{w \ge 0 \mid \mathsf{game}^*(w;a) = w\}. \end{aligned}$$

Proof. The infimum and maximum are equivalent by Lemma EC.2.1. The infimum is equal to the rank $r(a) = \inf_{c>a} \varphi(a,c)$ because, by the fact that we can write $\mathsf{game}(w;b,c)$ as

$$\mathsf{game}(w;b,c) = w - (w - \varphi(b,c))\,\mathsf{done}(b,c), \tag{EC.2.1}$$

we have $\mathsf{game}(w;b,c) < w$ if and only if $\varphi(b,c) < w$.[3] □

[3] Recall that $\mathsf{done}(b,c) \in [0,1]$ and that if $\mathsf{done}(b,c) = 0$, then $\varphi(b,c) = \infty$ (Definition 8.1).

LEMMA EC.2.3. *In the Gittins game with penalty w with the job currently at age a, it is optimal to continue serving the job if and only if $r(a) \leq w$,*[4] *and it is optimal to give up if and only if $r(a) \geq w$.*

Proof. Giving up incurs cost w, so by the maximum in Lemma EC.2.2, it is optimal to give up if and only if $r(a) \geq w$. This means it is optimal to continue serving the job if $r(a) < w$. The fact that serving is optimal in the $r(a) = w$ edge case follows from the fact that if $\varphi(a, c) = w$ for some $c > a$,[5] then by (EC.2.1), we have $\mathsf{game}(w; a, c) = w$. □

We are now ready to prove Lemma EC.2.3, which we restate below. Recall that a w-interval is one in which the Gittins rank is bounded above by w. The key to the proof is that Lemma EC.2.3 relates w-intervals to optimal play the Gittins game.

LEMMA 8.2. *Under Gittins, for any right-maximal w-interval (b, c), we have $\varphi(b, c) \leq w$.*

Proof. Consider playing the Gittins game starting from age b. By Lemma EC.2.3, giving up if the job reaches age c is an optimal policy. Specifically, because (b, c) is a w-interval, it is optimal to continue serving the job until at least age c, and because the w-interval is right-maximal, it is optimal to give up if the job reaches age c (which never happens if $c = x_{\max}$). This means $\mathsf{game}^*(w; b) = \mathsf{game}(w; b, c)$. Combining Lemma EC.2.1 and (EC.2.1) implies $\varphi(b, c) \leq w$. □

We note that Lemma 8.2 is similar, but not identical, to properties of Gittins in the M/G/1 studied by Aalto et al. (2009, 2011). Related properties have also been shown for versions of Gittins in discrete-time settings (Dumitriu et al., 2003; Gittins, 1989; Gittins et al., 2011).

EC.2.3. Relating the Gittins Game to Mean Response Time

It remains only to prove Theorem 5.10, which bounds the mean response time of q-approximate Gittins policies. To do so, we use a result of Scully et al. (2020a) that relates the Gittins game to a system's mean response time.

DEFINITION EC.2.4. Let $r : [0, x_{\max}) \to \mathbb{R}$ be the rank function of some SOAP policy, and let $w \in \mathbb{R}$.

(i) The (r, w)-*relevant work of a job* is the amount of service the job requires to either complete or reach rank at least w according to r, meaning reaching an age a satisfying $r(a) \geq w$.

(ii) The (r, w)-*relevant work of the system* is the total (r, w)-relevant work of all jobs present. We denote the steady-state distribution of (r, w)-relevant work under policy π by $W_\pi(r, w)$. Note that r need not be the rank function of policy π.

[4] Strictly speaking, it is optimal to continue serving the job if and only if the rank is upper bounded in a "forward neighborhood" of a, meaning there exists $\varepsilon > 0$ such that for all $\delta \in [0, \varepsilon)$, we have $r(a + \delta) \leq w$. For non-pathological job size distributions, this holds in the $r(a) < w$ case (Aalto et al., 2011), so it only needs to be checked when $r(a) = w$.

[5] The $c > a$ restriction is why we need the rank to be bounded not just at a but in a forward neighborhood of a.

The (r_{Gtn}, w)-relevant work of a job is related to the Gittins game via Lemma EC.2.3: it is the amount of time we would serve the job when optimally playing the Gittins game with penalty w. It turns out that mean (r_{Gtn}, w)-relevant work directly translates into mean response time.

LEMMA EC.2.5 (Scully et al. (2020a, Theorem 6.3)). *Under any nonclairvoyant scheduling policy π, the mean response time can be written in terms of (r_{Gtn}, w)-relevant work as*

$$\mathbf{E}[T_\pi] = \frac{1}{\lambda} \int_0^\infty \frac{\mathbf{E}[W_\pi(r_{\mathrm{Gtn}}, w)]}{w^2} \, \mathrm{d}w.$$

With Lemma EC.2.5 in hand, the proof of Theorem 5.10, restated below, reduces to bounding the mean amount of (r_{Gtn}, w)-relevant work under q-approximate Gittins policies.

THEOREM 5.10. *Consider an M/G/1 with any job size distribution. For any $q \geq 1$ and any q-approximate Gittins policy π,*[6]

$$\mathbf{E}[T_\pi] \leq q\mathbf{E}[T_{\mathrm{Gtn}}].$$

Proof. Recall from Definition 5.9 that we may assume $r_\pi(a)/r_{\mathrm{Gtn}}(a) \in [1, q]$ for all ages a without loss of generality. We will prove

$$\mathbf{E}[W_\pi(r_{\mathrm{Gtn}}, w)] \leq \mathbf{E}[W_\pi(r_\pi, qw)] \leq \mathbf{E}[W_{\mathrm{Gtn}}(r_{\mathrm{Gtn}}, w)], \tag{EC.2.2}$$

from which the theorem follows by the computation below:

$$\begin{aligned}
\mathbf{E}[T_\pi] &= \frac{1}{\lambda} \int_0^\infty \frac{\mathbf{E}[W_\pi(r_{\mathrm{Gtn}}, w)]}{w^2} \, \mathrm{d}w && \text{[by Lemma EC.2.5]} \\
&\leq \frac{1}{\lambda} \int_0^\infty \frac{\mathbf{E}[W_{\mathrm{Gtn}}(r_{\mathrm{Gtn}}, qw)]}{w^2} \, \mathrm{d}w && \text{[by (EC.2.2)]} \\
&= \frac{1}{\lambda} \int_0^\infty \frac{\mathbf{E}[W_{\mathrm{Gtn}}(r_{\mathrm{Gtn}}, w')]}{(w'/q)^2} \, \mathrm{d}(w'/q) && \text{[by substituting } w' = qw\text{]} \\
&= q\mathbf{E}[T_{\mathrm{Gtn}}]. && \text{[by Lemma EC.2.5]}
\end{aligned}$$

To show the left-hand inequality of (EC.2.2), it suffices to show that an arbitrary job's (r_{Gtn}, w)-relevant work is upper bounded by its (r_π, qw)-relevant work (Definition EC.2.4). This is indeed the case: $r_{\mathrm{Gtn}}(a) \leq w$ implies $r_\pi(a) \leq q r_{\mathrm{Gtn}}(a) \leq qw$, so the job will reach rank w under Gittins after at most as much service as it needs to reach rank qw under π.

To show the right-hand inequality of (EC.2.2) we use a property of SOAP policies due to Scully and Harchol-Balter (2018, proof of Lemma 5.2). The property implies that for any rank w and SOAP policy π, we can express $\mathbf{E}[W_\pi(r_\pi, w)]$ in terms of just the job size distribution X, arrival rate λ, and the set of ages $A_\pi[w] = \{a \in [0, x_{\max}) \mid r_\pi(a) < w\}$.[7] In particular, for any fixed job size distribution, arrival rate, and

[6] This is a special case of a more general result (Scully, 2022, Chapter 16), which appeared while this work was in revision.

[7] Scully and Harchol-Balter (2018) actually focus on $\mathbf{E}[W_\pi(r_\pi, w+)] = \lim_{w' \downarrow w} \mathbf{E}[W_\pi(r_\pi, w')]$ as opposed to $\mathbf{E}[W_\pi(r_\pi, w)]$, but the same reasoning applies to $\mathbf{E}[W_\pi(r_\pi, w+)]$.

rank w, $\mathbf{E}[W_\pi(r_\pi, w)]$ is a nondecreasing function of $A_\pi[w]$, where we order sets by the usual subset partial ordering. We have $r_\pi(a) \geq r_{\mathrm{Gtn}}(a)$, which means $A_\pi[w] \subseteq A_{\mathrm{Gtn}}[w]$, which implies the right-hand inequality of (EC.2.2), as desired. □

We note that one can use the techniques of Scully et al. (2018b) to generalize the statement and proof of Theorem 5.10 beyond SOAP policies. It turns out that Theorem 5.10 still holds even if we allow q-approximate Gittins policies to *adversarially* assign ranks to jobs, provided that the assigned ranks are still within a factor-q window around the rank Gittins would assign.

EC.3. Relationship Between Decay Rate and Laplace-Stieltjes Transform

The goal of this appendix is to justify our computation of decay rates (Definition 5.1) by means of Laplace-Stieltjes transform convergence (Section 9.3). Our specific goal is to justify our use of (9.1), which states $d(V) = -\gamma(\mathcal{L}[V])$. As a reminder,

$$d(V) = \lim_{t\to\infty} \frac{-\log \mathbf{P}[V > t]}{t},$$
$$\gamma(f) = \inf\{s \in \mathbb{R} \mid |f(s)| < \infty\}.$$

EC.3.1. Sufficient Condition for Computing Decay Rates

Our main tool for translating between $d(V)$ and $\gamma(\mathcal{L}[V])$ is a result of Mimica (2016), restated as Lemma EC.3.2 below, which gives a sufficient condition for $d(V) = -\gamma(\mathcal{L}[V])$. The result rests on the following definition.

DEFINITION EC.3.1. We say a function $f : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ is *regularly varying from the right at s^* with negative index*, or simply "regularly varying at s^*", if there exists $\alpha > 0$ such that for all $c > 0$,

$$\lim_{s\downarrow 0} \frac{f(s^* + cs)}{f(s^* + s)} = c^{-\alpha}.$$

In particular, f having a pole of finite order at s^* suffices.

It turns out being regularly varying at the singularity is the condition we need to express decay rate in terms of Laplace-Stieltjes transform convergence.

LEMMA EC.3.2 (special case of Mimica (2016, Corollary 1.3)). *Let V be a non-negative random variable with $\gamma(\mathcal{L}[V]) > -\infty$. If either $\mathcal{L}[V]$ or $\mathcal{L}[V]'$ is regularly varying at $\gamma(\mathcal{L}[V])$, then*

$$d(V) = -\gamma(\mathcal{L}[V]).$$

EC.3.2. Showing the Sufficient Condition for Computing Decay Rates Holds

It remains to show that the precondition of Lemma EC.3.2 holds whenever we apply (9.1) in Section 9.3. It turns out that all of the Laplace-Stieltjes transforms to which we apply (9.1) have a common form, so we will show that Lemma EC.3.2 applies to all functions of that form. To describe the form, we need the following definition.

DEFINITION EC.3.3. Consider an M/G/1 with arrival rate λ, job size distribution X, and load $\rho = \lambda \mathbf{E}[X]$.

(i) We define the function

$$\sigma_X^{-1}(s) = s - \lambda(1 - \mathcal{L}[X](s)).$$

Note that $\sigma_X^{-1}(s) = \infty$ if and only if $\mathcal{L}[X](s) = \infty$.

(ii) We define σ_X to be the the inverse of σ_X^{-1}, choosing the branch that passes through the origin. That is, for $s \geq \inf_r \sigma_X^{-1}(r)$, we define $\sigma_X(s)$ to be the greatest real solution to

$$\sigma_X(s) = s + \lambda(1 - \mathcal{L}[X](\sigma_X(s))).$$

If $s < \inf_r \sigma_X^{-1}(r)$, then no such solution exists, so we define $\sigma_X(s) = -\infty$.

(iii) We define the *work-in-system transform*

$$\mathcal{L}[W_X](s) = \frac{s(1-\rho)}{\sigma_X^{-1}(s)}.$$

Note that all of the above definitions depend on both λ and X, However, because the following discussion considers a fixed arrival rate λ and varies only the job size distribution X, we keep λ implicit to reduce clutter. Additionally, we assume in all uses of the above definitions that $\rho < 1$.

One may recognize the functions defined in Definition EC.3.3 as core to the theory of the M/G/1 with job size distribution X (Harchol-Balter, 2013).

- The work-in-system transform is, as suggested by its name, the Laplace-Stieltjes transform of the equilibrium distribution W_X of the total workload in the M/G/1.
- The function σ_X is related to busy periods in the M/G/1. Specifically, the length of a busy period started by initial workload V has Laplace-Stieltjes transform $\mathcal{L}[V](\sigma_X(s))$.

It turns out that throughout Section 9.3, all of the Laplace-Stieltjes transforms to which we apply (9.1) are of the form $\mathcal{L}[W_X]$ or $\mathcal{L}[W_X] \circ \sigma_Y$, the latter meaning $s \mapsto \mathcal{L}[W_X](\sigma_Y(s))$, for nicely light-tailed job size distributions X and Y (Definition 5.2). Specifically, X is the system's job size distribution, and Y is either X or a truncation $\min\{X, a^*\}$. Therefore, to justify the uses of (9.1) using Lemma EC.3.2, it suffices to prove Propositions EC.3.4 and EC.3.5 below.[8]

[8] While Definition EC.3.3 assumes a single arrival rate λ, Proposition EC.3.5 easily generalizes to the case where $\mathcal{L}[W_X]$ and σ_Y are defined using different arrival rates.

PROPOSITION EC.3.4. *For any nicely light-tailed job size distribution X,*

(i) $\gamma(\mathcal{L}[W_X]) \in (-\infty, 0)$*; and*

(ii) $\mathcal{L}[W_X]$ *has a first-order pole at* $\gamma(\mathcal{L}[W_X])$*, so it is regularly varying at* $\gamma(\mathcal{L}[W_X])$*.*

PROPOSITION EC.3.5. *For any nicely light-tailed job size distributions X and Y,*

(i) $\gamma(\mathcal{L}[W_X] \circ \sigma_Y) \in (-\infty, 0)$*, and*

(ii) either $\mathcal{L}[W_X] \circ \sigma_Y$ *or* $(\mathcal{L}[W_X] \circ \sigma_Y)'$ *is regularly varying at* $\gamma(\mathcal{L}[W_X] \circ \sigma_Y)$*.*

Our approach is as follows. We first prove Proposition EC.3.4. We then prove a lemma characterizing σ_X, which we use in conjunction with Proposition EC.3.4 to prove Proposition EC.3.5

Proof of Proposition EC.3.4. Recall from Definition EC.3.3 that $\mathcal{L}[W_X](s) = s(1-\rho)/\sigma_X^{-1}(s)$, so we focus on σ_X^{-1}. Because $\mathcal{L}[X]$ is a mixture of exponentials, σ_X^{-1} is convex, so it has at most two real roots. It is well-known that under the assumption on X made in Definition 5.2, σ_X^{-1} has a first-order root at 0 and a negative first-order root (Abate and Whitt, 1997; Mandjes and Nuyens, 2005), the latter of which is $\gamma(\mathcal{L}[W_X])$, but we give a brief proof for completeness. One can compute $\sigma_X^{-1}(0) = 0$ and $(\sigma_X^{-1})'(0) = 1-\rho$, so σ_X^{-1} has a first-order root at 0. Definition 5.2 implies $\mathcal{L}[X](\gamma(\mathcal{L}[X])) = \infty$, so $\sigma_X^{-1}(\gamma(\mathcal{L}[X])) = \infty$. This means σ_X^{-1} has another first-order root in $(\gamma(\mathcal{L}[X]), 0)$. □

LEMMA EC.3.6. *For any nicely light-tailed job size distribution X,*

(i) $\gamma(\sigma_X) \in (-\infty, 0)$*;*

(ii) $\sigma_X(\gamma(\sigma_X)) \in (-\infty, 0)$*; and*

(iii) there exist $C_0, C_1 > 0$ *such that in the* $s \downarrow 0$ *limit,*

$$\sigma_X(\gamma(\sigma_X) + s) = \sigma_X(\gamma(\sigma_X)) + C_0\sqrt{s} \pm \Theta(s),$$
$$\sigma_X'(\gamma(\sigma_X) + s) = \frac{C_1}{\sqrt{s}} \pm \Theta(1),$$

so σ_X' *is regularly varying at* $\gamma(\sigma_X)$*.*

Proof. As in the proof of Proposition EC.3.4, we again use the fact that σ_X^{-1} is convex, has roots at a negative number and at zero, and is negative between its roots. Specifically, this fact implies that σ_X^{-1} has a finite negative global minimum. By Definition EC.3.3, this minimum is $\gamma(\sigma_X)$, and the value at which the minimum is attained is $\sigma_X(\gamma(\sigma_X))$ proving (i) and (ii).

It remains only to prove (iii). The fact that Laplace-Stieltjes transforms are analytic in the interior of their domains of convergence implies that σ_X^{-1} can be written as a Taylor series about $\gamma(\sigma_X)$ whose first nonzero coefficient is quadratic, i.e. for some constant $K > 0$,

$$\sigma_X^{-1}(s) = Ks^2 \pm \Theta(s^3).$$

An extension of the Lagrange inversion theorem (DLMF, §1.10(vii)) implies that the inverse of σ_X^{-1}, namely σ_X, may thus be written in the desired form. The desired form for σ_X', which completes (iii), then follows from

$$(\sigma_X^{-1})'(s) = 2Ks \pm \Theta(s^2),$$
$$\sigma_X'(s) = \frac{1}{(\sigma^{-1})'(\sigma_X(s))}.$$ □

Proof of Proposition EC.3.5. There are three cases to consider:

- $\gamma(\mathcal{L}[W_X]) > \sigma_Y(\gamma(\sigma_Y))$,
- $\gamma(\mathcal{L}[W_X]) < \sigma_Y(\gamma(\sigma_Y))$, and
- $\gamma(\mathcal{L}[W_X]) = \sigma_Y(\gamma(\sigma_Y))$.

For an intuitive grasp of these cases, it is helpful to imagine decreasing s starting at $s = 0$, tracking the behavior of $\mathcal{L}[W_X](\sigma_Y(s))$ as s decreases.

If $\gamma(\mathcal{L}[W_X]) > \sigma_Y(\gamma(\sigma_Y))$, then at some point before $s = s^*$ reaches $\gamma(\sigma_Y)$, meaning for some $s^* \in (-\gamma(\sigma_Y), 0)$, we have $\gamma(\mathcal{L}[W_X]) = \sigma_Y(s^*)$. This means $\gamma(\mathcal{L}[W_X] \circ \sigma_Y) = s^*$. The Lagrange inversion theorem (DLMF, §1.10(vii)) and the fact that $s > \gamma(\sigma_Y)$ imply that σ_Y can be linearly approximated near s^*, so the result follows from Proposition EC.3.4.

If $\gamma(\mathcal{L}[W_X]) < \sigma_Y(\gamma(\sigma_Y))$, then in contrast to the previous case, s reaches $\gamma(\sigma_Y)$, the last point at which $\sigma_Y(s)$ is finite, before $\sigma(s)$ reaches the pole of $\mathcal{L}[W_x]$. This means $\gamma(\mathcal{L}[W_X] \circ \sigma_Y) = \gamma(\sigma_Y)$ Similarly to the previous case, we can linearly approximate $\mathcal{L}[W_x]$ near $\gamma(\sigma_Y)$, so the result follows from Lemma EC.3.6.

If $\gamma(\mathcal{L}[W_X]) = \sigma_Y(\gamma(\sigma_Y))$, then roughly speaking, both of the previous cases' events happen simultaneously: just as s reaches $\gamma(\sigma_Y)$, the last point at which $\sigma_Y(s)$ is finite, $\sigma_Y(s)$ reaches the pole of $\mathcal{L}[W_X]$. Combining Proposition EC.3.4 and Lemma EC.3.6 implies that in the $s \downarrow \gamma(\sigma_Y)$ limit, we can approximate $\mathcal{L}[W_X](\sigma_Y(s))$ as

$$\mathcal{L}[W_X](\sigma_Y(\gamma(s)) = \frac{K_0}{\sigma_Y(\gamma(s))} \pm \Theta(1) = \frac{K_1}{\sqrt{s - \gamma(\sigma_Y)}} \pm \Theta(1)$$

for some constants $K_0, K_1 > 0$, from which the result follows. □

EC.3.3. Expanding the Definition of Nicely Light-Tailed Job Size Distributions

The class of light-tailed distributions we consider in Definition 5.2, namely what Abate and Whitt (1997) call "Class I" distributions, is well behaved enough for Propositions EC.3.4 and EC.3.5 to hold. More generally, our results apply to any job size distribution with positive decay rate for which one can show Propositions EC.3.4 and EC.3.5. In particular, this includes many distributions that Abate and Whitt (1997) call "Class II". These are job size distributions X such that $\mathcal{L}[X](\gamma(\mathcal{L}[X])) < \infty$.

In order to prove Propositions EC.3.4 and EC.3.5 for Class II job size distributions, one would need to assume a regularity condition. We believe it would suffice to assume that $\mathcal{L}[X]'$ is regularly varying at $\gamma(\mathcal{L}[X])$. The main change to the proofs would be additional casework. For example, it may be that $\mathcal{L}[W_X]$ still has a first-order pole, or it may be that it diverges without a pole because $\mathcal{L}[X]$ does. See Abate and Whitt (1997) and references therein for additional discussion.

More generally, it likely suffices to assume that some higher-order derivative $\mathcal{L}[X]^{(n)}$ is regularly varying at $\gamma(\mathcal{L}[X])$, as the result of Mimica (2016, Corollary 1.3) we use applies to higher derivatives as well. Other results of Mimica (2016) may allow one to relax the assumption even further.