Simple Bounds on SMART Scheduling

Adam Wierman [1]   Mor Harchol-Balter [2]

November 2003

CMU-CS-03-199

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We define the class of SMART scheduling policies. These are policies that bias towards jobs with short remaining service times, jobs with small original sizes, or both, with the motivation of minimizing mean response time and/or mean slowdown. Examples of SMART policies include PSJF, SRPT, and hybrid policies such as RS (which biases according to the product of the response time and size of a job).

For many policies in the SMART class, the mean response time and mean slowdown are not known or have complex representations involving multiple nested integrals, making evaluation difficult. In this work, we prove three main results. First, for all policies in the SMART class, we prove simple upper and lower bounds on mean response time. In particular, we focus on the SRPT and PSJF policies and prove even tighter bounds in these cases. Second, we show that all policies in the SMART class, surprisingly, have very similar mean response times. Third, we show that the response times of SMART policies are largely invariant to the variability of the job size distribution.

[1]Carnegie Mellon University, Computer Science Department. Email: acw@cs.cmu.edu.
[2]Carnegie Mellon University, Computer Science Department. Email: harchol@cs.cmu.edu.

This work was supported by NSF ITR Grant 99-167 ANI-0081396, a grant from EMC^2 Corporation, and an NSF Graduate Research Fellowship.

Keywords: Scheduling, queueing, SRPT, shortest remaining processing time, PSJF, preemptive shortest job first, M/GI/1, response time, SMART.

1 Introduction

It is well-known that policies that bias towards small job sizes[1] or jobs with small remaining service times perform well with respect to mean response time and mean slowdown. This idea has been fundamental in many system implementations including, for example, the case of Web servers, where it has been shown that by giving priority to requests for small files, a Web server can significantly reduce mean response time and mean slowdown [4, 9]. The heuristic has also been applied to other application areas; for example, scheduling in supercomputing centers. Here too it is desirable to get small jobs out quickly to improve the overall mean response time.

Two specific examples of policies that employ this powerful heuristic are the Shortest-Remaining-Processing-Time (SRPT) policy, which preemptively runs the job with shortest remaining processing requirement and has been proven to be optimal with respect to mean response time [18]; and the Preemptive-Shortest-Job-First (PSJF) policy, which is easier to implement and preemptively runs the job with shortest original size.

While formulas are known for the mean response time under both SRPT and PSJF, these formulas are complex, involving multiple nested integrals. The formulas can be evaluated numerically, but the numerical calculations are quite time-consuming – in many situations simulating the policy is faster than evaluating the formulas numerically in Mathematica – and are numerically imprecise at high loads. No *simple* closed form formula is known for either of these policies. The complexity of these formulas also make understanding how far the mean response time of a novel policy is from optimal quite difficult. Furthermore, one can imagine many other scheduling policies that are hybrids of the SRPT and PSJF policies for which response time has never been analyzed.

In the current work, we define the SMART policies: a classification of all scheduling policies that "do the smart thing," i.e. follow the heuristic of biasing towards jobs that are originally short or have small remaining service requirements (see Definition 3.1). We then validate the heuristic of "biasing towards small job sizes" by deriving simple bounds on the mean response time of any policy in the SMART class, as well as tighter bounds on two important policies in the class: PSJF and SRPT. Our bounds illustrate that *all* the policies in the SMART class have surprisingly similar mean response times; and since our bounds are close, they allow us to predict this mean response time quite accurately. Our bounds also show the effect of the variability of the service distribution on the overall mean response time. Surprisingly, the mean response time is largely invariant to the variability of the service distribution, provided that the service distribution has at least the variability of an exponential distribution. This is contrary to intuition in the literature that suggests that the mean response time of SRPT significantly improves under highly variable service distributions. Most importantly however, these bounds are simple functions of the system load (see Theorem 5.1) and thus provide accurate, back-of-the-envelope calculations that can be used to understand the mean response times of these policies. In particular, we prove a simple lower bound on the optimal mean response time that is tight for highly variable service distributions. This lower bound provides a benchmark for describing the mean response times of other scheduling policies. Prior to this result, it has been difficult to assess the optimality of the mean response times of scheduling policies in a queueing setting. But, the simplicity of the lower bound in Theorem 5.1 facilitates such comparisons.

Throughout the paper we will consider only an M/GI/1 system with a differentiable service distribution having finite mean and finite variance. We let $T(x)$ be the steady-state response time for a job of size x, where the response time is the time from when a job enters the system until it completes service. Let $\rho < 1$ be the system load. That is $\rho \stackrel{\text{def}}{=} \lambda E[X]$, where λ is the arrival rate of the system and X is a random variable distributed according to the service (job size) distribution $F(x)$ having density function $f(x)$ defined for all $x \geq 0$. The expected response time for a job of size x under scheduling policy P is $E[T(x)]^P$, and the expected overall response time under scheduling policy P is $E[T]^P = \int_0^\infty E[T(x)]^P f(x)dx$.

[1] The "size" of a job is its service requirement. A small job is one with small (original) service requirement.

2 Background

There have been countless papers written on the analysis and implementation of individual scheduling policies. The "smarter" policies, such as SRPT dominate this literature [5, 13, 14, 19, 20]. Many individual "smart" policies have been analyzed for mean response time; two particularly important examples are SRPT and PSJF .

Before introducing the known results about PSJF and SRPT , it is important to point out that although formulas have been derived for the mean performance of both SRPT and PSJF, these formulas are not closed form. For many service distributions, because of the $1/(1-\rho(x))$ terms, these formulas must be evaluated numerically. Further, the complicated nature of these formulas hide any information about how properties of the service distribution effect the mean response time.

Under the SRPT policy, at every moment of time, the server is processing the job with the shortest remaining processing time. The SRPT policy is well-known to minimize overall mean response time [18]. The mean response time for a job of size x is as follows [19]:

$$E[T(x)]^{SRPT} = E[R(x)]^{SRPT} + E[W(x)]^{SRPT}$$

where $E[R(x)]^P$ (a.k.a the expected residence time for a job of size x under policy P) is the time for a job of size x to complete once it begins execution, and $E[W(x)]^P$ (a.k.a the expected waiting time for a job of size x under policy P) is the time between when a job of size x arrives and when it begins to receive service.

$$\begin{aligned} E[R(x)]^{SRPT} &= \int_0^x \frac{dt}{1-\rho(t)} \\ E[W(x)]^{SRPT} &= \frac{\lambda m_2(x) + \lambda x^2 \overline{F}(x)}{2(1-\rho(x))^2} = \frac{\lambda \int_0^x t\overline{F}(t)dt}{(1-\rho(x))^2} \end{aligned}$$

where $\rho(x) \stackrel{\text{def}}{=} \lambda \int_0^x tf(t)dt$ and $m_i(x) \stackrel{\text{def}}{=} \int_0^x t^i f(t)dt$. We will further use the notation

$$E[R]^P \stackrel{\text{def}}{=} \int_0^\infty E[R(x)]^P f(x)dx$$
$$E[W]^P \stackrel{\text{def}}{=} \int_0^\infty E[W(x)]^P f(x)dx$$

Under the PSJF policy, at every moment of time, the server is processing the job with the shortest initial size (service requirement). The mean response time for a job of size x is [11]:

$$E[T(x)]^{PSJF} = E[R(x)]^{PSJF} + E[W(x)]^{PSJF}$$
$$E[R(x)]^{PSJF} = \frac{x}{1-\rho(x)}$$
$$E[W(x)]^{PSJF} = \frac{\lambda m_2(x)}{2(1-\rho(x))^2}$$

Not only have countless papers been written analyzing individual scheduling policies; many others have been written comparing the response times of pairs of policies. Mean response time comparisons for SRPT and PS are made in [1, 8]; the mean response times for FB and PS are compared in [7, 21], and all three policies are compared in [17].

Recently however, there has been a trend in scheduling research towards grouping policies and proving results about policies with certain characteristics or structure. For example, the recent work of Borst, Boxma and Nunez groups policies with respect to their tail behavior [3, 16]. These authors have discovered that the tail of response time under SRPT, FB, and PS is the same as the tail of the service time distribution; however all non-preemptive policies, such as FCFS, have response time distributions with tails equivalent to the integrated service distribution. Another example of a classification of scheduling policies is with respect to their "fairness" properties [10, 22]. All this work has had a large impact on the implementation of scheduling policies. Across domains, scheduling policies that bias towards small job sizes are beginning to be adopted [4, 7, 9, 17]. This paper continues the trend towards classifying scheduling policies by defining a particular class of scheduling policies that all have similar, near optimal mean response time; thus placing important, additional structure on the vast domain of scheduling policies.

3 Defining the SMART class

We define the SMART class of scheduling policies as follows:

Definition 3.1 *A work conserving policy $P \in$ SMART if (i) a job of remaining size greater than x can never have priority over a job of original size x, and (ii) a job being run at the server can only be preempted by new arrivals.*

This definition has been crafted to mimic the heuristic of biasing towards jobs that are (originally) short or have small remaining service requirements. *Thus, we are able to analytically evaluate the validity of this heuristic by bounding the performance of all policies that obey it.* The heart of the SMART definition is in the first part which says that the job being run must have remaining size smaller than the original size of all jobs in the system. In particular, this implies that if $P \in$ SMART, P will never work on a new arrival of size greater than x while a previous arrival of original size x remains in the system. The second part of the definition intuitively says that the relative priority of two jobs does not change over time; thus if job a that is running currently has priority over job b, then job b will never preempt job a.

The class of SMART policies is very broad. Consider the following example of two jobs a and b with original size 10 and 8 respectively, where a arrives at time 0 and b arrives at time 3. At time 3, the remaining sizes of a and b are 7 and 8 respectively. A policy which at time 3 chooses to prioritize in favor of job a (e.g. SRPT) satisfies the definition for being in SMART. Likewise, a policy which at time 3 chooses to prioritize in favor of job b (e.g. PSJF) also satisfies the SMART definition. Furthermore, a policy which at time 3 probabilistically chooses between jobs a and b is likewise SMART. Although this example only includes two jobs, it illustrates the variety of choices possible under SMART policies — choices may be made probabilistically, based on only original size, based on only remaining size, or based on some combination of the original and remaining size.

We complete this section by giving more specific examples of policies included and not included in SMART. Observe that the class of SMART policies does not include non-preemptive policies, not even Shortest-Job-First (SJF). However, as noted above, the SMART class does include the SRPT and PSJF policies. Further, it is easy to prove that the SMART class includes the RS policy, which assigns to each job the product of its remaining size and its original size and then gives highest priority to the job with lowest product. The motivation for the RS policy is improving mean slowdown, where a job's slowdown is defined as its response time divided by its original size. By incorporating size into the priority scheme, the RS policy is able to improve mean slowdown over SRPT in many cases. In simulations, we have found that under distributions having variability greater than that of the exponential, RS can achieve up to a 10% relative improvement in mean slowdown. Furthermore, the SMART class includes many generalizations of the RS policy that can (under some distributions) lead to further improvements in mean slowdown. Specifically, SMART includes all policies of the form $R^i S^j$, where $i, j > 0$ and a job is assigned the product of its remaining size raised to the ith power and its original size raised to the jth power (where again the job with highest priority is the one with lowest product). The SMART class also includes a range of policies having more complicated priority schemes; see Definition 3.2.

Definition 3.2 *A policy $P \in$ SMART$*$ if P at any given time schedules the job with the highest priority and gives each job of size s and remaining size r a priority $p(s, r)$ such that for $s_1 < s_2$ and $r_1 < r_2$, $p(s_1, r_1) > p(s_2, r_2)$ and $p(s_1, r_1) > p(s_1, r_2)$.*

We will next prove that SMART$* \subsetneq$ SMART.

Theorem 3.1 SMART$* \subsetneq$ SMART

Proof : Suppose policy $P \in$ SMART$*$. We will first show that Definition 3.1 is satisfied by P. Notice that part (ii) of the definition is trivially satisfied. To see that part (i) is satisfied, let s_1 and r_1 be the initial size and current remaining size of a tagged job in the queue. Suppose s_2 and r_2 correspond to the the initial size and current remaining size of another job in the queue such that $r_2 > s_1$. It follows that $s_2 > s_1$, and further that $r_2 > r_1$. Thus, $p(s_2, r_2) < p(s_1, r_1)$, so job 2 will not be served.

Finally, notice that SMART is strictly larger than SMART$*$. We can see this by giving an example of a policy in SMART that is not in SMART$*$. Define P to be the policy that for each busy period uses priority function $p_1(s, r)$ with probability q and priority function $p_2(s, r)$ with probability $1-q$ where both p_1 and p_2 are in SMART$*$. Then, $P \in$ SMART but $P \notin$ SMART$*$. ■

4 Bounding the per-size response time under SMART policies

In this section, we present an upper bound on the mean response time for a job of size x under policies in SMART. The purpose of this bound is solely in its use towards deriving an upper bound on the overall mean response time, $E[T]$, under SMART policies in Section 5, although the proof technique is elegant in its own right.

Theorem 4.1 *The mean response time for a job of size x under any policy $P \in$ SMART satisfies:*

$$E[T(x)]^P \leq \frac{x}{1-\rho(x)} + \frac{\lambda m_2(x) + \lambda x^2 \overline{F}(x)}{2(1-\rho(x))^2}$$

Proof : We will break up the mean response time for a job of size x into the sum of the waiting time $W(x)^P$ and the residence time $R(x)^P$, defined in Section 2.

We first notice that the residence time under any SMART policy is upper bounded by:

$$E[R(x)]^P \leq \frac{x}{1-\rho(x)}$$

This bound follows from the fact that no arrival of size greater than x will be worked on while a job of original size x is in the system. Thus, the response time for such a job of size x is bounded by the length of a busy period made up of only jobs with sizes smaller than x.

It now remains to bound the waiting time for a job of size x, $W(x)$, under any SMART policy P. Consider an M/GI/1 queue with scheduling policy P. Let V be the work in the system as seen by an arrival of size x, having higher priority than x under policy P. Observe that

$$E[W(x)]^P \leq \frac{E[V]^P}{1-\rho(x)}$$

This follows from the fact that no arrival of size greater than x will be worked on while our job of size x is in the system. Thus $W(x)$ is bounded by a busy period started by V work including only arriving jobs of size x or smaller.

To analyze V, we consider a "transformed" system, which perfectly mimics the original system, running the same jobs at the same times, however where jobs with remaining size greater than x are simply non-existent in the transformed system. To be precise, there are two types of arrivals into the transformed system: type 1 arrivals occur when jobs of original size greater than x in the original system have been worked on to the point where their remaining size is now exactly x (call this time t). We restrict the type 1 arrivals further to include just those jobs whose priority at time t would have exceeded that of our arrival of size x. Type 2 arrivals occur when jobs arrive into the original system with size less than x.

We make three claims about type 1 jobs arriving into the transformed system:

1. The type 1 arrivals enter the transformed system at the server.
2. The type 1 arrivals occur only when the transformed system is idle of jobs of type 2.
3. There is only one job of type 1 in the transformed system at a time.

The first point is obvious. The second point follows from the fact that when the type 1 arrival enters the transformed system, it has highest priority at that moment, and therefore there cannot be any job of original size less than x in the system (by the definition of SMART). To argue the third point, consider a job j which becomes a type 1 arrival into the transformed system at time t. Clearly, j has the highest priority of those jobs currently in the system at time t, and thus it will, by part (ii) of the SMART definition, forever continue to have priority over those jobs that were in the system at time t. Furthermore, consider any new arrival, j' into the system of size greater than x that arrives while job j is in the transformed system. We claim that j' has lower priority than j and thus will never become a type 1 job while j is in the system. To see that j' has lower priority than j, observe that (a) at time t job j had higher priority than our arrival of size x, by definition of a type 1 arrival, and (b) an arrival of size x has priority over job j' by definition of SMART, since the size of j' exceeds x. Thus, by transitivity, j' has lower priority than j, and, by part (ii) of the SMART definition, will continue to.

Recall that our goal is the work in the transformed system. Since the transformed system is work-conserving, the work is equal to that in a further-transformed system, where we now change the service policy in the transformed system so that it is non-preemptive, specifically, a job in service is never interrupted (in particular a type 1 job will never be sent to the queue), and all type 2 jobs are served in FCFS order. Aside from the scheduling policy, the further-transformed system is identical to the transformed system.

Now observe that the work in the further-transformed system is identical to the waiting time (delay) experienced by a type 2 arrival into the further-transformed system. Thus, we have equated the work in the further-transformed system with the delay experienced by a type 2 (Poisson) arrival into a single-server system consisting of a queue made up of all Poisson arrivals of size less than x and a server which may be busy with jobs of type 1 or 2. That is, the distribution of jobs at the server in our further-transformed system is $X_x = \min(x, X)$, and the load at the server is $\rho_x \stackrel{\text{def}}{=} \lambda E[X_x]$. Letting $\widetilde{N_Q}$ be the number of jobs in the queue of the further-transformed system, $\widetilde{V}$ be the work in the further-transformed system, and noting that the mean excess of X_x is $E[X_x^2]/(2E[X_x])$, we have:

$$\begin{aligned}
E[V] &\leq E[\text{work in transformed system}] \\
&= E[\widetilde{V}] \\
&= \rho_x \frac{E[X_x^2]}{2E[X_x]} + E[\widetilde{N_Q}] \int_0^x t \frac{f(t)}{F(x)} dt \\
&= \frac{\lambda E[X_x^2]}{2} + \lambda F(x) E[\widetilde{V}] \int_0^x t \frac{f(t)}{F(x)} dt \\
&= \frac{\lambda E[X_x^2]}{2(1-\rho(x))} \\
&= \frac{\lambda \int_0^x t^2 f(t) dt + \lambda x^2 \overline{F}(x)}{2(1-\rho(x))}
\end{aligned}$$

which completes the proof. ■

Notice that the upper bound in Theorem 4.1 is tight, since one can define a policy P where for an arrival of size x, all jobs with remaining size less than x have priority over the arrival; and further all arriving jobs of size less than x have priority over the arrival.

5 Bounding mean response time under SMART policies

In this section we derive bounds on the overall mean response time of policies in SMART. To do this, it will be helpful to start by deriving bounds on the PSJF policy, then use those bounds to derive bounds on the SRPT policy, and finally use those bounds to bound the entire SMART class. It is important to notice that all these bounds are very simple. They do not involve nested integrals; yet we will see in Section 6 that they are nevertheless accurate.

In order to better understand the results in this section, all of our bounds will be stated in terms of the mean response time of Processor-sharing (PS), a very common scheduling policy that serves as a convenient benchmark for mean response time. Under the PS policy, at any point in time, the service rate is shared evenly among all jobs in the system. Recall that the overall mean response time under PS is [11]:

$$E[T]^{PS} = \frac{E[X]}{1-\rho}$$

The main results in this section are stated in the following theorem. Recall that

$$C^2[X] \stackrel{\text{def}}{=} \frac{E[X^2]}{E[X]^2} - 1$$

Theorem 5.1 *Let $f(x)$ be decreasing. Then*

$$\begin{aligned}
E[T]^{PSJF} &\geq -\left(\frac{1-\rho}{\rho}\right) \log(1-\rho) E[T]^{PS} \\
E[T]^{PSJF} &\leq \left(\frac{1}{3} - \frac{2}{3}\left(\frac{1-\rho}{\rho}\right) \log(1-\rho)\right) E[T]^{PS} \\
E[T]^{SRPT} &\geq -\left(\frac{1-\rho}{\rho}\right) \log(1-\rho) E[T]^{PS} \\
E[T]^{SRPT} &\leq \left(\frac{2}{3} - \frac{\rho}{3} - \frac{1}{3}\left(\frac{1-\rho}{\rho}\right) \log(1-\rho)\right) E[T]^{PS} \\
E[T]^{SMART} &\geq -\left(\frac{1-\rho}{\rho}\right) \log(1-\rho) E[T]^{PS} \\
E[T]^{SMART} &\leq \left(-\frac{1}{6} + \frac{\rho(1-\rho)}{4}\left(2 + C^2[X]\right) \right. \\
&\qquad \left. - \frac{7}{6}\left(\frac{1-\rho}{\rho}\right) \log(1-\rho)\right) E[T]^{PS}
\end{aligned}$$

The above bounds are tighter than those previously known relating mean response time under SRPT and PS, [1, 8].

In interpreting the above theorem, it is useful to consider that the lower bound shown in all cases above is equal to the mean residence time under the PSJF policy. This will be proven in Lemma 5.1, which shows that:

$$E[R]^{PSJF} = -\left(\frac{1-\rho}{\rho}\right) \log(1-\rho) E[T]^{PS}$$

An important point to notice is that the bounds for SRPT and PSJF are independent of the variability of the service distribution. Although, as discussed in Section 2, there are known formulas for the mean response times of SRPT and PSJF, the complicated nature of these formulas hid this fact from prior research. The simplicity of the bounds in 5.1 illuminate this practical property. We will see later that these bounds are in fact *tight* in the sense that there are distributions with low variability for which the upper bounds are exact and there are distributions with high variability for which the lower bounds are exact.

A second important point about Theorem 5.1 is that it provides a lower bound on the mean response time of the optimal scheduling policy, SRPT. Despite the fact that there is a known formula for the mean performance of SRPT, researchers have been forced to resort to computational techniques when comparing the performance of new scheduling policies to that of SRPT. The lower bound in 5.1 provides a *simple* benchmark that can be used to understand how far the mean response times of other scheduling policies are from optimal.

The results of Theorem 5.1 are presented in greater generality in Theorems 5.3, 5.4, 5.6, 5.7, and 5.9 in this section, where they are stated in terms of a parameter K. This K parameter is a constant such that $\lambda m_2(x) \leq K x \rho(x)$, which serves to bound the $\lambda m_2(x)$ term that arises in Theorem 4.1. Theorem 5.2 shows that the constant K may be set at $\frac{2}{3}$ when the service distribution is decreasing, as has been done in Theorem 5.1. But in more generality, it defines K in a way that is highly tied to the tail properties of $f(x)$. Note that $K \leq 1$ under all service distributions.

Theorem 5.2 *Let i be a positive integer. Define j such that $x^j f(x)$ is decreasing and $j < i+1$. Then,*

$$m_{i+1}(x) \le \left(\frac{i-j+1}{i-j+2}\right) x m_i(x)$$

Corollary 5.1 *Let X be such that $f(x)$ is decreasing. Then,*

$$\lambda m_2(x) \le \frac{2}{3} x\rho(x)$$

We defer the proof of Theorem 5.2 to Section 5.4 and we will first use this bound on $\lambda m_2(x)$ to bound the performance of PSJF, SRPT, and all SMART policies. In reading this section, note that Appendix A contains a list of integrals that are useful in these calculations and that Appendix B contains some crucial technical lemmata.

5.1 Bounding mean response time under PSJF

In this section, we derive bounds on the overall mean response time under PSJF, $E[T]^{PSJF}$. To accomplish this, we will first bound the residence time, $E[R]^{PSJF}$, and then the waiting time, $E[W]^{PSJF}$, under PSJF. Both of these preliminary bounds will be useful in later sections as well. In all of the following proofs, observe that $\frac{d}{dx}\rho(x) = \lambda x f(x)$.

Lemma 5.1

$$E[R]^{PSJF} = -\frac{1}{\lambda}\log(1-\rho)$$

Proof : Follows immediately from the fact that $E[R]^{PSJF} = \int_0^\infty \frac{xf(x)}{1-\rho(x)}dx$ and $\frac{d}{dx}\rho(x) = \lambda x f(x)$. ■

We now move to bounding the waiting time under PSJF.

Lemma 5.2 *Let K satisfy $\lambda m_2(x) \le K x\rho(x)$. Then*

$$E[W]^{PSJF} \le \frac{K}{2\lambda}\left(\frac{\rho}{1-\rho} + \log(1-\rho)\right)$$

Proof : Using Lemma A.3, we have:

$$\begin{aligned} E[W]^{PSJF} &= \int_0^\infty \frac{\lambda m_2(x)}{2(1-\rho(x))^2} f(x)dx \\ &\le \frac{K}{2\lambda}\int_0^\infty \frac{\lambda x f(x)\rho(x)}{(1-\rho(x))^2}dx \\ &= \frac{K}{2\lambda}\left(\frac{\rho}{1-\rho} + \log(1-\rho)\right) \end{aligned}$$

■

Lemma 5.3

$$E[W]^{PSJF} \ge \frac{\lambda}{4} E[\min(X_1,X_2)^2]$$

where X_1 and X_2 are independent random variables from the service distribution on an M/GI/1.

Proof : Recall that the p.d.f. of $\min(X_1,X_2)$ is $f_{min}(x) = 2f(x)\overline{F}(x)$. Thus

$$\begin{aligned} E[W]^{PSJF} &= \int_0^\infty \frac{\lambda\int_0^x t^2 f(t)dt}{2(1-\rho(x))^2} f(x)dx \\ &\ge \frac{\lambda}{2}\int_0^\infty f(x)\int_0^x t^2 f(t)dtdx \\ &= \frac{\lambda}{4}\int_0^\infty 2t^2 f(t)\overline{F}(t)dt \\ &= \frac{\lambda}{4}E[\min(X_1,X_2)^2] \end{aligned}$$

■

Using our bounds on the waiting time under PSJF, we can now derive bounds on the overall mean response time under PSJF.

Theorem 5.3 *Let K satisfy $\lambda m_2(x) \le K x\rho(x)$. Then*

$$E[T]^{PSJF} \le \left(\frac{K}{2} + \left(\frac{K}{2}-1\right)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho)\right) E[T]^{PS}$$

Proof : Using Lemma 5.1 and Lemma 5.2, we have:

$$\begin{aligned} E[T]^{PSJF} &= \int_0^\infty \left(\frac{x}{1-\rho(x)} + \frac{\lambda m_2(x)}{2(1-\rho(x))^2}\right) f(x)dx \\ &= \frac{1}{\lambda}\int_0^\infty \left(\frac{\lambda x f(x)}{1-\rho(x)} + \frac{\lambda x f(x)\rho(x)}{2(1-\rho(x))^2}\right)dx \\ &= -\frac{1}{\lambda}\log(1-\rho) + \frac{K}{2\lambda}\left(\frac{\rho}{1-\rho} + \log(1-\rho)\right) \\ &= \left(\frac{K}{2} + \left(\frac{K}{2}-1\right)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho)\right) E[T]^{PS} \end{aligned}$$

■

Theorem 5.4

$$\begin{aligned} E[T]^{PSJF} \ge \Bigg(&\frac{\lambda E[\min(X_1,X_2)^2]}{4E[X]}(1-\rho) \\ &-\frac{1-\rho}{\rho}\log(1-\rho)\Bigg) E[T]^{PS} \end{aligned}$$

Proof : Using Lemma 5.1 and Lemma 5.3, we have:

$$\begin{aligned} E[T]^{PSJF} &= \int_0^\infty \left(\frac{\lambda\int_0^x t^2 f(t)dt}{2(1-\rho(x))^2} + \frac{x}{1-\rho(x)}\right) f(x)dx \\ &\ge \frac{\lambda}{4}E[\min(X_1,X_2)^2] - \frac{1}{\lambda}\log(1-\rho) \\ &= \Bigg(\frac{\lambda E[\min(X_1,X_2)^2]}{4E[X]}(1-\rho) \\ &\quad -\frac{1-\rho}{\rho}\log(1-\rho)\Bigg) E[T]^{PS} \end{aligned}$$

■

5.2 Bounding mean response time under SRPT

Using the results from the previous section and the technical lemmata in Appendix B, we can now derive bounds on the overall mean response time under SRPT. Our goal in this section is to bound $E[T]^{SRPT}$. To do this, we first bound the residence time, $E[R]^{SRPT}$.

Lemma 5.4

$$E[R]^{SRPT} \geq E[X] + \frac{\rho^2}{2\lambda} - \frac{\lambda}{2}E[\min(X_1, X_2)^2]$$

where X_1 and X_2 are independent random variables from the service distribution on an M/GI/1.

Proof : Recall that the p.d.f. of $\min(X_1, X_2)$ is $f_{min}(x) = 2f(x)\overline{F}(x)$. Thus

$$\begin{aligned}
E[R]^{SRPT} &= \int_0^\infty f(x) \int_0^x \frac{dt}{1-\rho(t)} dx \\
&= \int_0^\infty f(x) \left(x + \int_0^x \frac{\rho(t)}{1-\rho(t)} dt \right) dx \\
&\geq \int_0^\infty f(x) \left(x + \int_0^x \rho(t) dt \right) dx \\
&= E[X] + \int_0^\infty f(x) \left(x\rho(x) - \lambda m_2(x) \right) dx \\
&= E[X] + \frac{1}{\lambda} \int_0^\infty \rho'(x)\rho(x) dx \\
&\quad -\lambda \int_0^\infty t^2 f(t)\overline{F}(t) dt \\
&= E[X] + \frac{\rho^2}{2\lambda} - \frac{\lambda}{2}E[\min(X_1, X_2)^2]
\end{aligned}$$

■

Interestingly, we can exactly characterize the improvement SRPT makes over PSJF. Define

$$E[W_2] \stackrel{\text{def}}{=} \int_0^\infty \frac{\lambda x^2 f(x)\overline{F}(x)}{2(1-\rho(x))^2} dx$$

Although we cannot evaluate $E[W_2]$ exactly, we can show that the mean response time of PSJF is exactly $E[W_2]$ away from optimal.

Theorem 5.5

$$E[T]^{SRPT} = E[T]^{PSJF} - E[W_2]$$

Proof : Using Lemma B.1, we have:

$$\begin{aligned}
E[T]^{SRPT} &= E[R]^{SRPT} + E[W]^{PSJF} + E[W_2] \\
&= \frac{1}{2}E[R]^{PSJF} + \frac{1}{2}E[R]^{SRPT} + E[W]^{PSJF} \\
&= E[T]^{PSJF} - \frac{1}{2}E[R]^{PSJF} + \frac{1}{2}E[R]^{SRPT} \\
&= E[T]^{PSJF} - E[W_2]
\end{aligned}$$

■

We are now ready to bound the overall mean response time of SRPT.

Theorem 5.6 *Let K satisfy $\lambda m_2(x) \leq Kx\rho(x)$. Then*

$$E[T]^{SRPT} \leq \left(K - \frac{K\rho}{2} + (K-1)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho) \right) E[T]^{PS}$$

Proof : Using Lemma B.4, we have:

$$\begin{aligned}
E[T]^{SRPT} &= E[W]^{SRPT} + E[R]^{SRPT} \\
&= -\frac{1}{2\lambda}\log(1-\rho) - \frac{1}{2}E[R]^{SRPT} \\
&\quad +E[W]^{PSJF} + E[R]^{SRPT} \\
&\leq -\frac{1}{2\lambda}\log(1-\rho) \\
&\quad +\frac{1}{2\lambda}\left(\frac{K\rho^2}{(1-\rho)} + 2K\rho + (2K-1)\log(1-\rho) \right) \\
&= \frac{K-1}{\lambda}\log(1-\rho) + \frac{K\rho}{2}E[T]^{PS} + KE[X] \\
&= \left(K - \frac{K\rho}{2} + (K-1)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho) \right) E[T]^{PS}
\end{aligned}$$

■

Theorem 5.7

$$E[T]^{SRPT} \geq -\left(\frac{1-\rho}{\rho}\right)\log(1-\rho)E[T]^{PS}$$

Proof : Using Lemma B.5, we have:

$$\begin{aligned}
E[T]^{SRPT} &= E[W]^{SRPT} + E[R]^{SRPT} \\
&= -\frac{1}{2\lambda}\log(1-\rho) - \frac{1}{2}E[R]^{SRPT} \\
&\quad +E[W]^{PSJF} + E[R]^{SRPT} \\
&\geq -\frac{1}{2\lambda}\log(1-\rho) - \frac{1}{2\lambda}\log(1-\rho) \\
&= -\left(\frac{1-\rho}{\rho}\log(1-\rho)\right) E[T]^{PS}
\end{aligned}$$

■

An interesting observation about Theorem 5.7 is that the lower bound we have proven is exactly the mean residence time under PSJF, that is, we have shown that $E[T]^{SRPT} \geq E[R]^{PSJF}$. Further, Theorem 5.7 is perhaps the most important result of this section because it provides a *simple* lower bound on the optimal mean response time. Thus, it provides a simple benchmark that can be used in evaluating the mean response times of other scheduling policies.

5.3 Bounding the mean response time under all SMART policies

In this section, we derive an upper bound on the overall mean response time under any policy in the SMART class. Note that the lower bound on SRPT serves as a lower bound on the mean response time of any policy in the SMART class since SRPT is known to be optimal with respect to overall mean response time.

To derive an upper bound on the response time of SMART policies, we start by integrating the expression for $E[T(x)]$ from Theorem 4.1. The result is shown in Theorem 5.9. Before we present this result, we make another interesting observation: the mean response time of any SMART policy is at most $2E[W_2]$ away from optimal, where (by Theorem 5.5) we can think of $E[W_2]$ as being the difference in mean mean response time between SRPT and PSJF. Another way to think about the $E[W_2]$ is stated in Lemma B.1: $2E[W_2] = E[R]^{PSJF} - E[R]^{SRPT}$.

Theorem 5.8

$$\begin{aligned} E[T]^{SMART} &\le E[T]^{PSJF} + E[W_2] \\ &= E[T]^{SRPT} + 2E[W_2] \end{aligned}$$

Proof: Proof follows immediately by comparing the result in Theorem 4.1 with the the formulas on PSJF given in Section 2, and using the result of Theorem 5.5. ■

We are now ready to upper bound the mean response time of policies in SMART. In this proof we again make use of the technical lemmata in Appendix B.

Theorem 5.9 *Let K satisfy $\lambda m_2(x) \le Kx\rho(x)$. Then*

$$\begin{aligned} E[T]^{SMART} \le & \left(\frac{\rho}{4} + \frac{K-1}{2} + \frac{\rho^2}{4} + (1-\rho)\frac{\lambda E[\min(X_1,X_2)^2]}{4E[X]} \right. \\ & \left. + \left(\frac{K-3}{2}\right)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho)\right) E[T]^{PS} \end{aligned}$$

Proof: Using Theorem 5.3, Lemma B.1, and Lemma 5.4, we have:

$$\begin{aligned} E[T]^{SMART} &\le E[T]^{PSJF} + E[W_2] \\ &= \frac{K-3}{2\lambda}\log(1-\rho) + \frac{K}{2}E[T]^{PS} \\ &\quad -\frac{1}{2}E[R]^{SRPT} \\ &\le \frac{K-3}{2\lambda}\log(1-\rho) + \frac{K}{2}E[T]^{PS} \\ &\quad -\frac{1}{2}\left(E[X] + \frac{\rho^2}{2\lambda} - \frac{\lambda}{2}E[\min(X_1,X_2)^2]\right) \\ &= E[T]^{PS}\left(\left(\frac{K-3}{2}\right)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho) + \frac{K}{2}\right. \\ &\quad \left. -\frac{1}{2}(1-\rho)\left(1 + \frac{\rho}{2} - \frac{\lambda E[\min(X_1,X_2)^2]}{2E[X]}\right)\right) \\ &= \left(\frac{\rho}{4} + \frac{K-1}{2} + \frac{\rho^2}{4} + (1-\rho)\frac{\lambda E[\min(X_1,X_2)^2]}{4E[X]}\right. \\ &\quad \left. + \left(\frac{K-3}{2}\right)\left(\frac{1-\rho}{\rho}\right)\log(1-\rho)\right)E[T]^{PS} \end{aligned}$$

■

Theorem 5.9 and Theorem 5.7 together provide upper and lower bounds on the mean response time of any SMART policy. In the next section we will see that these bounds are very close together; thus any SMART policy is guaranteed near optimal mean response time. One important consequence of these bounds is that there are now simple benchmarks that provide upper and lower bounds on the mean response times of "smart" scheduling policies, which facilitates the evaluations of policies that are not "smart" but still claim to provide good mean response time.

5.4 A proof of Theorem 5.2

The upper bounds for all SMART policies are expressed in terms of a constant K, which is the smallest constant satisfying: $\lambda m_2(x) \le Kx\rho(x)$, where $m_i(x) = \int_0^x t^i f(t)dt$. In this section we derive this constant K.

Proof: (of Theorem 5.2)

First, we observe the following equality:

$$\begin{aligned} \int_{t=0}^x m_i(t)dt &= \int_{t=0}^x \int_{s=0}^t s^i f(s)dsdt \\ &= \int_{s=0}^x s^i f(s) \int_{t=s}^x dtds \\ &= \int_{s=0}^x (x-s)s^i f(s)ds \\ &= xm_i(x) - m_{i+1}(x) \end{aligned} \tag{1}$$

We will now use this relation to bound $m_{i+1}(x)$ in terms of $m_i(x)$ by first bounding $\int_0^x m_i(t)dt$. Remember, by assumption we know that $s^j f(s)$ is decreasing for some j such that $j < i+1$.

$$\begin{aligned} \int_{t=0}^x m_i(t)dt &= \int_{t=0}^x \int_{s=0}^t s^i f(s)dsdt \\ &\ge \int_{t=0}^x t^j f(t) \int_{s=0}^t s^{i-j}dsdt \\ &= \frac{1}{i-j+1}\int_{t=0}^x t^j f(t)t^{i-j+1}dt \\ &= \frac{1}{i-j+1}m_{i+1}(x) \end{aligned} \tag{2}$$

In this chain of equalities, the inequality follows directly from the assumption that $s^j f(s)$ is decreasing.

Finally, combining Equation 1 and Equation 2, we can complete the proof.

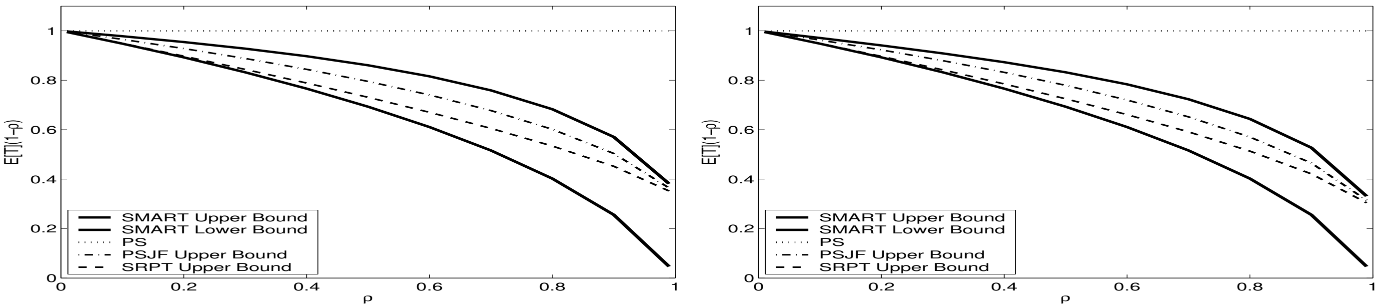
$$\begin{aligned} xm_i(x) - m_{i+1}(x) &\ge \frac{1}{i-j+1}m_{i+1}(x) \\ \left(\frac{i-j+1}{i-j+2}\right)xm_i(x) &\ge m_{i+1}(x) \end{aligned}$$

■

A few comments are in order about this theorem. First, notice that in this work we only apply the lemma in the case where $i = 1$, but the more general form is useful for investigating higher moments. Second, notice that because K is defined in terms of j, where j is such that $x^j f(x)$ is decreasing in x, K is related to the variability of the service distribution. Third, notice that for any service distribution, $K \le 1$.

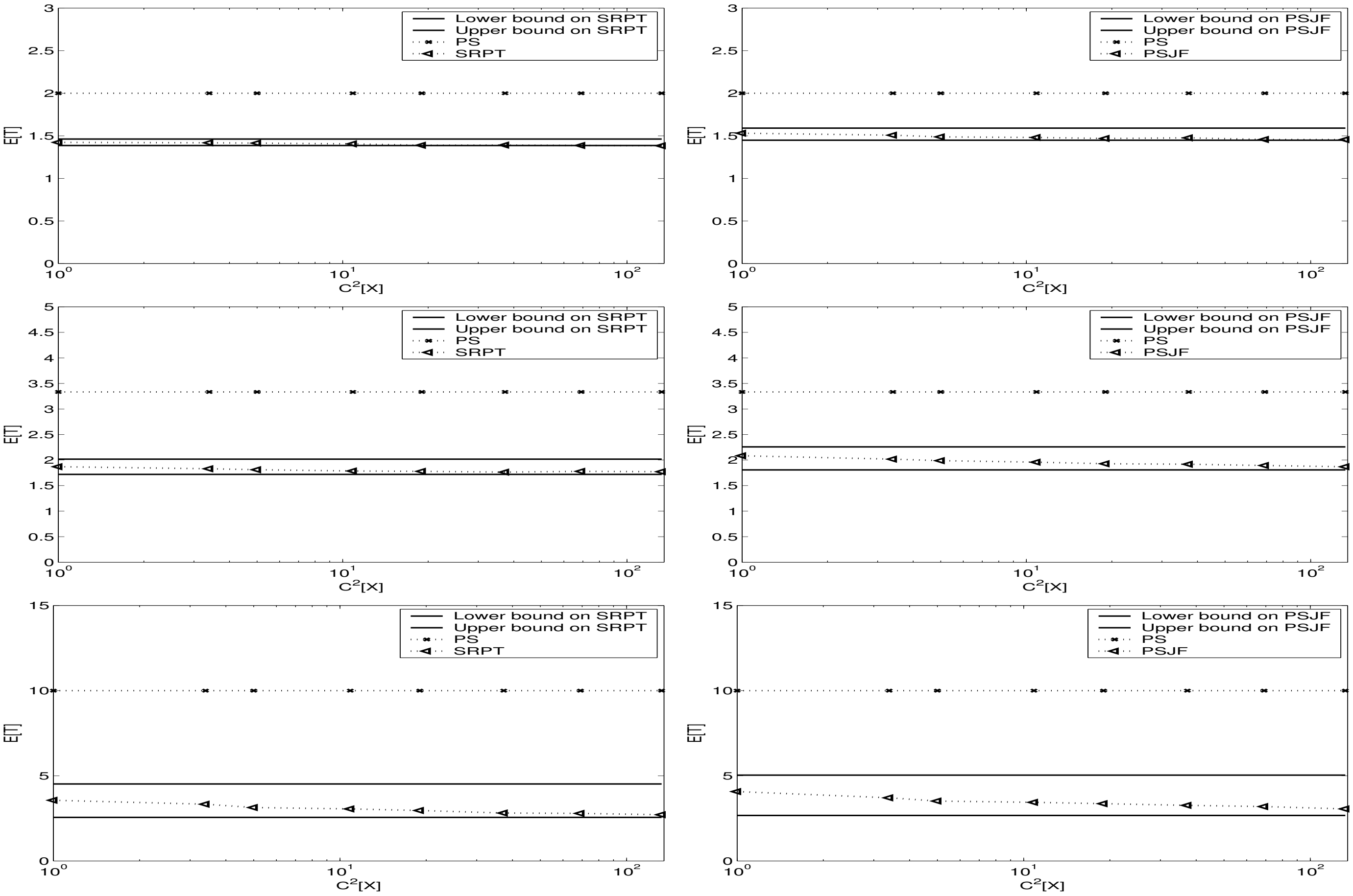
6 Evaluating the bounds

In order to better understand the bounds derived in the previous section, we investigate how the bounds perform for specific service distributions.



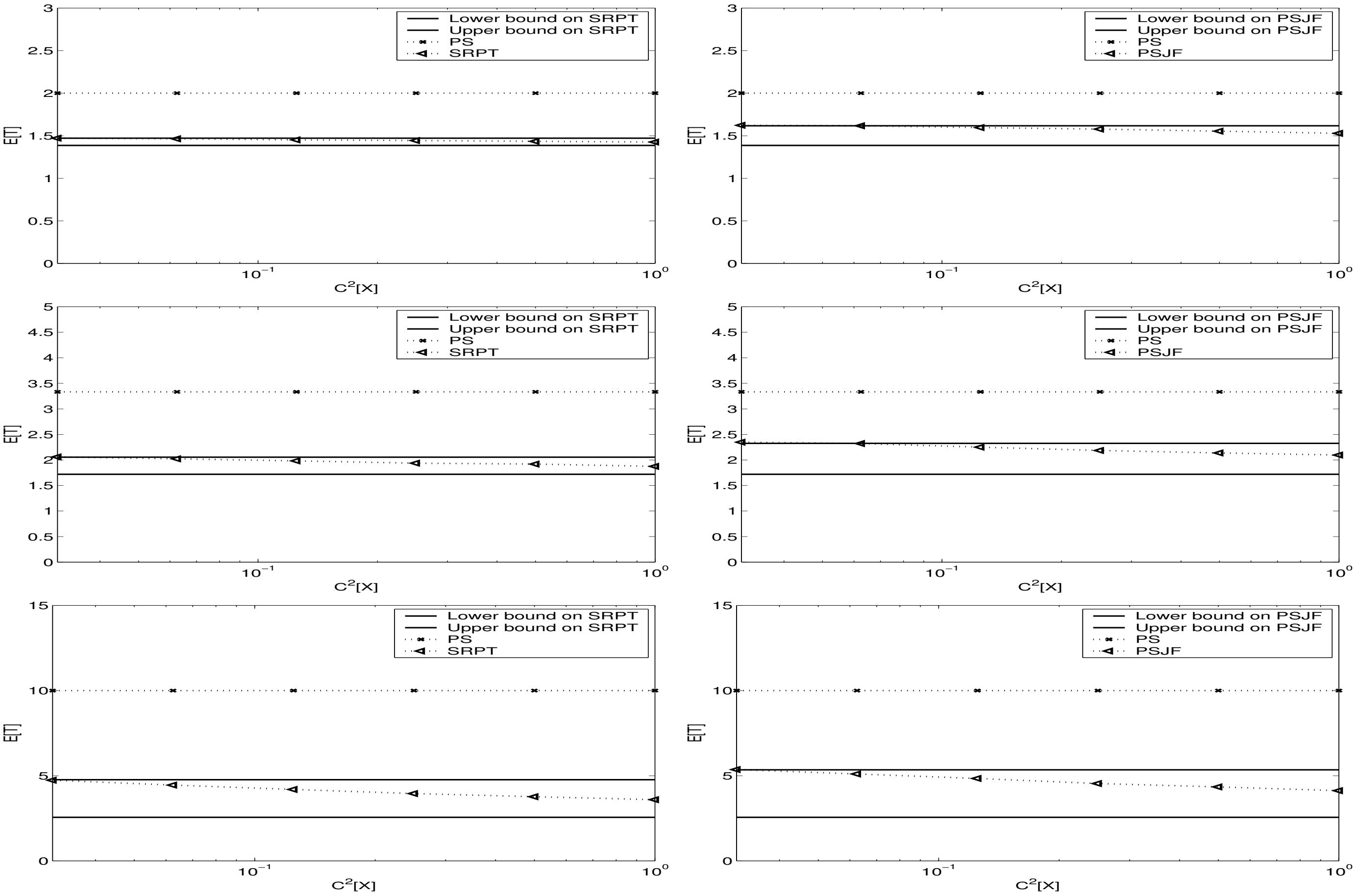
(a) $C^2[X] = 1$ (b) $C^2[X] = 10.865$

Figure 1: These plots show our analytic upper and lower bounds on the mean response time of SMART policies (shown in solid lines). The metric shown, $E[T](1-\rho)$, depicts the improvement made by SMART policies over PS. Between the solid lines are dashed lines showing our tighter bounds for PSJF and SRPT. The service distribution in these plots is Weibull with mean 1 and (a) $C^2[X] = 1$, (b) $C^2[X] = 10.865$, respectively.



(a) SRPT (b) PSJF

Figure 2: These plots show a comparison of the bounds proven for (a) SRPT and (b) PSJF with simulation results. The service distribution in these plots is an Weibull with mean 1, and varying coefficient of variation. System loads are 0.5, 0.7, and 0.9 in the first, second, and third rows respectively. These plots illustrate that the lower bounds on both PSJF and SRPT are tight as the variability of the service distribution increases. Surprisingly, they also show that the mean response times under both SRPT and PSJF are nearly independent of the service distribution's variability, once the service distribution has at least the variability of an exponential.



(a) SRPT (b) PSJF

Figure 3: These plots show a comparison of our analytic bounds proven for (a) SRPT and (b) PSJF with exact results. The service distribution in these plots is an Erlang with mean 1, and varying coefficient of variation. The system loads are 0.5, 0.7, and 0.9 in the first, second, and third rows respectively. These plots illustrate that the upper bounds on both PSJF and SRPT are tight as the variability of the service distribution decreases.

The Weibull and Erlang distributions are convenient ways to evaluate the effects of variability in the service distribution because they allow a wide range of variability and tail behavior. Investigating the effect of the weight of the tail of the service distribution is important in light of many recent measurements that have observed job size distributions that are well-modeled by heavy tailed distributions such as the Weibull distribution [2, 6, 12, 15].

The goal in investigating how the bounds perform under these service distributions is twofold. Our first goal is to illustrate the similar mean response time attained by all policies in SMART, and in particular PSJF and SRPT. It is well known that SRPT is optimal, but it is quite surprising to the authors of this paper how close to optimal the mean response time of PSJF is — and further, how close to optimal the mean response time of any SMART policy is.

Second, our bounds on the mean response time of PSJF and SRPT are independent of the variability of the service distribution. Thus, it is difficult to tell how tight they are without investigating the mean response time of these two policies under a wide range of service distributions. This section will illustrate that the bounds are tight in the sense that there are low variability service distributions under which the mean response time of these two policies match our upper bounds, and high variability service distributions under which the mean response times of these two policies match our lower bounds. Thus, no bounds independent of the variability of the service distribution can improve significantly on the bounds presented in this work.

6.1 The Weibull distribution

We will first investigate the Weibull distribution. The Weibull distribution is defined by:

$$\begin{aligned} f(x;b,c) &= \frac{cx^{c-1}}{b^c} e^{-\left(\frac{x}{b}\right)^c} \\ \overline{F}(x;b,c) &= e^{-\left(\frac{x}{b}\right)^c} \end{aligned}$$

Notice that $Wei(b, c = 1) \sim Exp(1/b)$. We will be concerned with the case where $c \leq 1$, which corresponds to the case where the distribution is at least as variable as an exponential. Note also that for $c \leq 1$ the Weibull distribution has a decreasing failure rate. To get a feeling for the variability of this distribution notice that for $c = 1/l$ where l is limited to positive integer values, we have that $C^2[X] = \binom{2l}{l} - 1$. Thus, as c decreases the distribution becomes more variable very quickly. Typical observed values for the variability parameter, c, range between 1/3 and 2/3 which correspond to $C^2[X]$ values in the range of 3 to 19.

First, in Figure 1, the bounds on SRPT, PSJF, and SMART are pictured as a function of ρ both in the case of a service distribution with low variability and high variability. These plots illustrate the huge performance gains (a factor of 2 – 3 under high load) made by SRPT and PSJF over PS. We also see that any policy in SMART will have a huge performance gain over PS – also a factor of 2 – 3 under high load. Further, the mean response time of any of the SMART policies cannot differ too much from the mean response time of the optimal policy, SRPT. Thus, by simply following the "smart" rule of not allowing a job with remaining time greater than x to run when a job of original size x is in the system, a policy is guaranteed to achieve near-optimal mean response time.

Second, in Figure 2, the bounds derived for SRPT and PSJF are compared with the exact mean response time of these policies under a Weibull service distribution. It is important to point out that the "exact results" for the points in these plots are often obtained via simulation, and then spot-checked via analysis. This is because simulations, despite being slow, are still orders of magnitude faster than Mathematica on evaluating the expressions for the exact mean response time. Thus, the methodology used in creating all the plots in this paper was to pick a mesh of points on the plot and calculate the exact mean response time of these points. Then, using these points to judge the accuracy of simulations, determine how many iterations of simulations are necessary to attain the desired accuracy, and fill in the plot using simulated values. *The fact that simulations are used to generate these plots underscores the importance of the results in this paper, which provide simple, back-of-the-envelope calculations for the mean response time.*

Throughout the plots in Figure 2, the mean of the service distribution is fixed at 1, and $C^2[X]$ is allowed to vary. The values of the variability parameter range between $c = 1$ and $c = 2/9$, which corresponds to a range of $C^2[X]$ from 1 to more than 100. Thus, the plots show the effect variability has on the mean response time of SRPT and PSJF.

Note that the lower bound becomes extremely accurate when the service distribution has high variability, but that the upper bound is loose throughout these plots. The reason the upper bound appears loose in this figure is that we keep the parameter $c \leq 1$, so the Weibull cannot have $C^2[X] < 1$. Thus, since the upper bound applies for all distributions, it is tight for distributions with much lower $C^2[X]$. We will see this when we look at Erlang distributions in the next section.

An important point that Figure 2 illustrates is the surprisingly small effect of variability on the overall mean response time. The fact that PS is insensitive to variability in the service distribution is usually thought of as a very special property. However, these plots illustrate that both SRPT and PSJF are almost insensitive to the variability of the service distribution once the $C^2[X] > 1$. This is in contrast to the common intuition that as the variability of the service distribution increases there will be a larger separation between the large and small job sizes and thus SRPT will perform significantly better.

6.2 The Erlang distribution

When looking at the Weibull distribution in the previous section, we were able to illustrate that our lower bounds are tight as the variability of the service distribution increases. Our goal in this section is to show that our upper bounds are tight as the variability decreases. Thus, we investigate how our bounds perform under the Erlang service distribution. Recall that the $Erl(n, \mu)$ distribution is the sum of n exponential distributions each having rate μ.

The key differences between the Erlang and Weibull distributions are (1) the Erlang distribution is limited to having $C^2[X] \leq 1$ and (2) under the Erlang distribution as n grows smaller and smaller j are necessary in order guarantee that $x^j f(x)$ is decreasing. This second point tells us that in order to bound all Erlang distributions we must weaken the bounds by setting $K = 1$ 5.4.

In Figure 3, the bounds derived for SRPT and PSJF are compared with the exact values for these policies under an Erlang service distribution. We follow the same methodology for generating these plots as described in the previous section. Thus, these plots represent a mixture of simulated and exact values, where the accuracy of the simulations is held in check using exact calculations.

Throughout these plots, the mean of the service distribution is fixed at 1, and $C^2[X]$ is allowed to vary. The plots show the affect of a wide range of variability on the mean response times of SRPT and PSJF.

The important difference between these plots and the plots in Figure 2 is that the Erlang distribution can have $C^2[X]$ far below 1. This allows us to see that for distributions with low variability the upper bound is quite accurate. Thus, our bounds give an excellent characterization of the mean response times of SRPT and PSJF over distributions with widely ranging $C^2[X]$, and are as tight as possible without including the variability of the service distribution.

7 Conclusion

The heuristic of "biasing towards small job sizes" is commonly accepted as a way of providing good mean response times. However, some practical roadblocks remain.

First, the mean response time for policies that bias towards small jobs is often not known; and even in the cases where the policy has been analyzed, the resulting formula is typically complex, involving multiple nested integrals. Consequently, evaluating the mean response times of such policies via lengthy simulation is actually faster than evaluating the known complex analytical expressions using Mathematica. This evokes the question of whether there exists a simpler, quicker way to estimate mean response time for these policies.

Second, there is the question of how such policies that bias towards small jobs compare to each other with respect to mean response time. There are many possible variants of such policies, each with their own benefits and weaknesses. Some, like PSJF, are relatively easy to implement, because priority is never updated. Others, like SRPT, are more complex to implement because they require updating priorities as jobs run, but have superior fairness properties. Yet others, like RSimprove mean slowdown. However, when choosing among these policies, it is not clear how much one sacrifices with respect to mean response time in order to attain these other benefits. The little work that exists on comparing mean response time among policies compares specific, individual policies and leads to bounds that are not as tight as the ones provided in this work.

This paper fills both gaps above. We begin by formalizing the heuristic of biasing towards short jobs by defining the SMART class, which is very broadly defined to include all policies that "do the smart thing," i.e. bias towards jobs that are originally short or have small remaining service requirements (see Definition 3.1). We then evaluate the validity of this heuristic by proving *simple* upper and lower bounds on the mean response of any SMART policy. Surprisingly, these upper and lower bounds are reasonably close, leading us to conclude that, although the SMART class includes many different policies, all SMART policies are quite similar with respect to mean response time. In fact, all are far superior to PS, and most importantly, all have quite close to the optimal mean response time. This result theoretically validates the heuristic of "biasing towards small job sizes" that many system designers apply. We then go on to prove even tighter bounds on two particular SMART policies: SRPT and PSJF . The bounds proven are far tighter than anything previously known for these policies, and allow us to "quickly and simply" predict mean response time for these policies as a function of the workload.

An unanticipated discovery of this work is the invariance of SMART policies to the variability of the job size distribution (particularly for $C^2 > 1$). It is well-known that the mean response time of PS is independent of the service distribution's variability, but the fact that mean response time for policies like SRPT and PSJF is nearly independent of the service distribution's variability is counter the folklore of the community. This unanticipated discovery illustrates the importance of providing simple bounds on the known, but complex, formulas for mean response time under SRPT and PSJF.

There are some long term impacts of our results on future scheduling research. First, our results show that understanding the mean response time of a SMART policy in the case of an M/M/1 queue may suffice to reasonably predict its mean response time for an M/GI/1 queue. And finally, the simple bounds on mean response time for SMART policies provide a benchmark for showing that a policy P is "good" even if its particular definition precludes it from belonging to the SMART class. More strongly, the very simple lower bound proven on SRPT's mean response time, facilitates comparison with any new policy P in order to assess P's optimality or lack thereof. Prior to this work, in order to assess the mean response time of a new policy P it was necessary to compare the the mean response time to a very complicated expression that is even time consuming to evaluate numerically. Now it is possible to do so by simply comparing the mean response time of P to the lower bound $-\frac{1}{\lambda}\log(1-\rho)$, which is a much simpler calculation and is tight for highly variable distributions.

References

[1] N. Bansal and M. Harchol-Balter. Analysis of SRPT scheduling: Investigating unfairness. In *Proceedings of ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2001.

[2] P. Barford and M. Crovella. Generating representative web workloads for network and server performance evaluation. In *Proceedings of ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 1998.

[3] S. Borst, O. Boxma, and R. N. Queija. Heavy tails: the effect of the service discipline. In *Computer Performance Evaluation - Modelling Techniques and Tools (TOOLS)*, pages 1–30, 2002.

[4] L. Cherkasova. Scheduling strategies to improve response time for web applications. In *High-performance computing and networking: international conference and exhibition*, pages 305–314, 1998.

[5] R. W. Conway, W. L. Maxwell, and L. W. Miller. *Theory of Scheduling*. Addison-Wesley Publishing Company, 1967.

[6] A. B. Downey. Evidence for long-tailed distributions in the internet. In *Proceedings of ACM SIGCOMM Internet Measurment Workshop*, 2001.

[7] H. Feng and V. Misra. Mixed scheduling disciplines for network flows (the optimality of FBPS). In *Workshop on MAthematical performance Modeling and Analysis (MAMA 2003)*, 2003.

[8] M. Gong and C. Williamson. Quantifying the properties of SRPT scheduling. In *IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2003.

[9] M. Harchol-Balter, B. Schroeder, N. Bansal, and M. Agrawal. Implementation of SRPT scheduling in web servers. *ACM Transactions on Computer Systems*, 21(2), May 2003.

[10] M. Harchol-Balter, K. Sigman, and A. Wierman. Asymptotic convergence of scheduling policies with respect to slowdown. *Performance Evaluation*, 49(1-4):241–256, 2002.

[11] L. Kleinrock. *Queueing Systems*, volume II. Computer Applications. John Wiley & Sons, 1976.

[12] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. In *Proceedings of SIGCOMM '93*, pages 183–193, September 1993.

[13] T. O'Donovan. Direct solutions of M/G/1 priority queueing models. *Revue Francaise d'Automatique Informatique Recherche Operationnelle*, 10:107–111, 1976.

[14] A. Pechinkin, A. Solovyev, and S. Yashkov. A system with servicing discipline whereby the order of remaining length is serviced first. *Tekhnicheskaya Kibernetika*, 17:51–59, 1979.

[15] D. L. Peterson. Data center I/O patterns and power laws. In *CMG Proceedings*, December 1996.
[16] R. N. Queija. Queues with equally heavy sojourn time and service requirement distributions. *Ann. Oper. Res*, 113:101–117, 2002.
[17] I. Rai, G. Urvoy-Keller, and E. Biersack. Analysis of LAS scheduling for job size distributions with high variance. In *Proceedings of ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2003.
[18] L. E. Schrage. A proof of the optimality of the shortest remaining processing time discipline. *Operations Research*, 16:678–690, 1968.
[19] L. E. Schrage and L. W. Miller. The queue M/G/1 with the shortest remaining processing time discipline. *Operations Research*, 14:670–684, 1966.
[20] D. Smith. A new proof of the optimality of the shortest remaining processing time discipline. *Operations Research*, 26:197–199, 1976.
[21] A. Wierman, N. Bansal, and M. Harchol-Balter. A note comparing response times in the M/GI/1/FB and M/GI/1/PS queues. *Operations Research Letters*, 32:73–76, 2003.
[22] A. Wierman and M. Harchol-Balter. Classifying scheduling policies with respect to unfairness in an M/GI/1. In *Proceedings of ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2003.

A Useful Integrals

This section contains integrals that are useful in the calculations of Section 5.

Lemma A.1

$$\int_0^x \rho(t)dt = \lambda \int_0^x (x-t)tf(t)dt = x\rho(x) - \lambda m_2(x)$$

Lemma A.2

$$E[R]^{SRPT} = \int_0^\infty f(x) \int_0^x \frac{dt}{1-\rho(t)} = \int_0^\infty \frac{\overline{F}(x)}{1-\rho(x)}dx$$

Lemma A.3

$$\int_0^\infty \frac{\lambda x f(x)\rho(x)}{(1-\rho(x))^2}dx = \frac{\rho}{1-\rho} + \log(1-\rho)$$

Proof :

$$\begin{aligned}\int_0^\infty \frac{\rho'(x)\rho(x)}{(1-\rho(x))^2}dx &= \left.\frac{\rho(x)}{1-\rho(x)}\right|_0^\infty - \int_0^\infty \frac{\rho'(x)}{1-\rho(x)}dx \\ &= \frac{\rho}{1-\rho} + \log(1-\rho)\end{aligned}$$

■

Lemma A.4

$$\int_0^\infty \frac{\lambda x f(x)\rho(x)}{1-\rho(x)}dx = -\log(1-\rho) - \rho$$

Proof :

$$\begin{aligned}\int_0^\infty \frac{\rho'(x)\rho(x)}{1-\rho(x)}dx &= -\rho(x)\log(1-\rho(x))|_0^\infty \\ &\quad - \int_0^\infty -\rho'(x)\log(1-\rho(x))dx \\ &= -\rho\log(1-\rho) - (1-\rho)\log(1-\rho) - \rho \\ &= -\log(1-\rho) - \rho\end{aligned}$$

■

Lemma A.5

$$\int_0^\infty \frac{\lambda x f(x)\rho(x)^2}{(1-\rho(x))^2}dx = \frac{\rho^2}{1-\rho} + 2\log(1-\rho) + 2\rho$$

Proof :

$$\begin{aligned}\int_0^\infty \frac{\rho'(x)\rho(x)^2}{(1-\rho(x))^2}dx &= \left.\frac{\rho(x)^2}{1-\rho(x)}\right|_0^\infty - \int_0^\infty \frac{2\rho'(x)\rho(x)}{1-\rho(x)}dx \\ &= \frac{\rho^2}{1-\rho} + 2\log(1-\rho) + 2\rho\end{aligned}$$

■

B Some technical lemmata

In performing the analyses of SRPT and SMART, we need a few technical lemmata. These lemmata relate the waiting time and residence times under PSJF, SRPT, and our upper bound on SMART policies. Define

$$E[W_2] \stackrel{\text{def}}{=} \int_0^\infty \frac{\lambda x^2 f(x)\overline{F}(x)}{2(1-\rho(x))^2}dx$$

Lemma B.1

$$2E[W_2] = E[R]^{PSJF} - E[R]^{SRPT}$$

Proof : Using Lemmas 5.1 and A.2, we have:

$$\begin{aligned}2E[W_2] &= \int_0^\infty \frac{\lambda x^2 f(x)\overline{F}(x)}{(1-\rho(x))^2}dx \\ &= \int_0^\infty f(t)\int_0^t \frac{x\rho'(x)}{(1-\rho(x))^2}dxdt \\ &= \int_0^\infty f(t)\left(\left.\frac{x}{1-\rho(x)}\right|_0^t - \int_0^t \frac{1}{1-\rho(x)}dx\right)dt \\ &= \frac{1}{\lambda}\int_0^\infty \frac{\rho'(t)}{1-\rho(t)} - \int_0^\infty f(t)\int_0^t \frac{1}{1-\rho(x)}dxdt \\ &= -\frac{1}{\lambda}\log(1-\rho) - \int_0^\infty \frac{\overline{F}(x)}{1-\rho(x)}dx \\ &= E[R]^{PSJF} - E[R]^{SRPT}\end{aligned}$$

■

Lemma B.2

$$
\begin{aligned}
& E[R(x)]^{SRPT} + 2E[W(x)]^{PSJF} \\
& \leq E[R(x)]^{PSJF} + \frac{\lambda m_2(x)\rho(x)}{(1-\rho(x))^2}
\end{aligned}
$$

Proof : Using Lemma A.1, we have:

$$
\begin{aligned}
& E[R(x)]^{SRPT} + 2E[W(x)]^{PSJF} \\
= \; & \int_0^x \frac{dt}{1-\rho(t)} + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & \frac{x}{1-\rho(x)} - \int_0^x \frac{\rho(x)-\rho(t)}{(1-\rho(x))(1-\rho(t))} dt + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
\leq \; & \frac{x}{1-\rho(x)} - \int_0^x \frac{\rho(x)-\rho(t)}{(1-\rho(x))} dt + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & \frac{x}{1-\rho(x)} - \frac{x\rho(x) - x\rho(x) + \lambda m_2(x)}{(1-\rho(x))} + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & E[R(x)]^{PSJF} + \frac{\lambda m_2(x)\rho(x)}{(1-\rho(x))^2}
\end{aligned}
$$

■

Lemma B.3

$$
E[R(x)]^{SRPT} + 2E[W(x)]^{PSJF} \geq E[R(x)]^{PSJF}
$$

Proof : Using Lemma A.1, we have:

$$
\begin{aligned}
& E[R(x)]^{SRPT} + 2E[W(x)]^{PSJF} \\
= \; & \int_0^x \frac{dt}{1-\rho(t)} + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & \frac{x}{1-\rho(x)} - \int_0^x \frac{\rho(x)-\rho(t)}{(1-\rho(x))(1-\rho(t))} dt + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
\geq \; & \frac{x}{1-\rho(x)} - \int_0^x \frac{\rho(x)-\rho(t)}{(1-\rho(x))^2} dt + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & \frac{x}{1-\rho(x)} - \frac{x\rho(x) - x\rho(x) + \lambda m_2(x)}{(1-\rho(x))^2} + \frac{\lambda m_2(x)}{(1-\rho(x))^2} \\
= \; & E[R(x)]^{PSJF}
\end{aligned}
$$

■

Lemma B.4 *Let K satisfy $\lambda m_2(x) \leq Kx\rho(x)$.*

$$
\begin{aligned}
E[R]^{SRPT} \; & + \; 2E[W]^{PSJF} \\
& \leq \frac{1}{\lambda}\left(\frac{K\rho^2}{1-\rho} + 2K\rho + \frac{2K-1}{\lambda}\log(1-\rho)\right)
\end{aligned}
$$

Proof : Using Lemma B.2 and Lemma A.5, we have:

$$
\begin{aligned}
E[R]^{SRPT} \; + & \; \int_0^\infty \frac{\lambda m_2(x)}{(1-\rho(x))^2} f(x)dx \\
= & \; \int_0^\infty \left(\frac{x}{1-\rho(x)} + \frac{\lambda m_2(x)\rho(x)}{(1-\rho(x))^2}\right) f(x)dx \\
\leq & \; -\frac{1}{\lambda}\log(1-\rho) + \frac{K}{\lambda}\int_0^\infty \frac{\lambda x f(x)\rho(x)^2}{(1-\rho(x))^2} dx \\
= & \; -\frac{1}{\lambda}\log(1-\rho) + \frac{K}{\lambda}\left(\frac{\rho^2}{1-\rho} + 2\log(1-\rho) + 2\rho\right) \\
= & \; \frac{1}{\lambda}\left(\frac{K\rho^2}{1-\rho} + 2K\rho + \frac{2K-1}{\lambda}\log(1-\rho)\right)
\end{aligned}
$$

■

Lemma B.5

$$
E[R]^{SRPT} + 2E[W]^{PSJF} \geq E[R]^{PSJF}
$$

Proof : Using Lemma B.3, we have:

$$
\begin{aligned}
E[R]^{SRPT} + 2E[W]^{PSJF} \; & \geq \; \int_0^\infty E[R(x)]^{PSJF} f(x)dx \\
& = \; E[R]^{PSJF}
\end{aligned}
$$

■