Nearly Insensitive Bounds on SMART Scheduling[*]

Adam Wierman
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
acw@cs.cmu.edu

Mor Harchol-Balter
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
harchol@cs.cmu.edu

Takayuki Osogami
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
osogami@cs.cmu.edu

ABSTRACT

We define the class of SMART scheduling policies. These are policies that bias towards jobs with small 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 remaining size and the original 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. 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 insensitive to the variability of the job size distribution. In particular, we focus on the SRPT and PSJF policies and prove insensitive bounds in these cases.

Categories and Subject Descriptors

F.2.2 [**Nonnumerical Algorithms and Problems**]: Sequencing and Scheduling; G.3 [**Probability and Statistics**]: Queueing Theory; C.4 [**Performance of Systems**]: Performance Attributes

General Terms

Performance, Algorithms

Keywords

Scheduling; response time; SMART; M/G/1; PS; processor sharing; SRPT; shortest remaining processing time; PSJF; preemptive shortest job first

1. INTRODUCTION

It is well known that policies that bias towards small job sizes (service requirements) 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 [9, 18]. This heuristic has also been applied to other application areas; such as, 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 [19]; 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 makes it difficult to understand how far the mean response time of a novel policy is from optimal. 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* policies in the SMART class have near optimal mean response times. In fact all SMART policies have mean response time within a factor of 2 of optimal across all loads and all service distributions. Further, 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 insensitive to the variability of the service distribution, provided that the service distribution has

[*]Supported by NSF Career Grant CCR-0133077, NSF Theory CCR-0311383, NSF ITR CCR-0313148, IBM Corporation via Pittsburgh Digital Greenhouse Grant 2003, and a NSF Graduate Research Fellowship.

SIGMETRICS'05, June 6–10, 2005, Banff, Alberta, Canada.
Copyright 2005 ACM 1-59593-022-1/05/0006 ...$5.00.

at least the variability of an exponential distribution. This has escaped prior investigation due to the complexity of the known representations of mean response time and is contrary to intuition in the literature, which suggests that the mean response time of SRPT significantly improves under highly variable service distributions.

Throughout the paper we will consider only a preempt-resume M/GI/1 system with a differentiable service distribution having finite variance. We focus on work conserving scheduling policies. 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. Define the slowdown for a job of size x, $S(x) \overset{\text{def}}{=} T(x)/x$. Let $\rho < 1$ be the system load. That is $\rho \overset{\text{def}}{=} \lambda E[X]$, 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 $x \geq 0$. Let $\overline{F}(x) \overset{\text{def}}{=} 1 - F(x)$. 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 $E[T]^P = \int_0^\infty E[T(x)]^P f(x)dx$. Define $m_i(x) \overset{\text{def}}{=} \int_0^x t^i f(t)dt$ and $\widetilde{m}_i(x) \overset{\text{def}}{=} i\int_0^x t^{i-1}\overline{F}(t)dt$. Notice that equivalently $m_i(x)/F(x) = E[X^i|X \leq x]$ and $\widetilde{m}_i(x)$ is the ith moment of $X_x \overset{\text{def}}{=} \min(X,x)$. Further define $\rho(x) = \lambda m_1(x)$ and $\widetilde{\rho}(x) = \lambda\widetilde{m}_1(x)$. Finally, define $C^2[X] \overset{\text{def}}{=} E[X^2]/E[X]^2 - 1$ to be the squared coefficient of variation of X.

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, 14, 15, 20, 21]. 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 these formulas must be evaluated numerically. Further, the complicated nature of these formulas hide any information about how properties of the service distribution affect the mean response time.

Under the SRPT policy, the server is processing the job with the shortest remaining processing time at every moment of time. The SRPT policy is well known to minimize overall mean response time [19]. The mean response time for a job of size x is as follows [20]:

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

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.

$$\begin{aligned} E[R(x)]^{SRPT} &= \int_0^x \frac{dt}{1-\rho(t)} \\ E[W(x)]^{SRPT} &= \frac{\lambda m_2(x) + \lambda x^2 \overline{F}(x)}{2(1-\rho(x))^2} = \frac{\lambda \widetilde{m}_2(x)}{2(1-\rho(x))^2} \end{aligned}$$

We will further use the notation $E[R]^P \overset{\text{def}}{=} \int_0^\infty E[R(x)]^P f(x)dx$ and $E[W]^P \overset{\text{def}}{=} \int_0^\infty E[W(x)]^P f(x)dx$.

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

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

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 [2, 8]; the mean response times for FB and PS are compared in [7, 22], 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-Queija groups policies with respect to their tail behavior [4, 13]. 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, 23].

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 [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 additional structure on the vast domain of scheduling policies.

3. DEFINING THE SMART CLASS

We will need the following notation throughout. Jobs will typically be denoted by a, b, or c. Job a will have remaining size r_a, original size s_a, and arrival time t_a. The original sizes, remaining sizes, and arrival times of b and c are defined similarly.

Throughout this paper, we define *job a to have priority over job b* if job b can never run while job a is in the system.

We now define SMART as follows.

DEFINITION 3.1. *Every work conserving policy $P \in$ SMART must obey the following properties.*

Bias Property: *If $r_b > s_a$, then job a has priority over job b.*

Consistency Property: *If job a ever receives service while job b is in the system, thereafter job a has priority over job b.*

Transitivity Property: *If an arriving job b preempts job c; thereafter, until job c receives service, every arrival, a, with size $s_a < s_b$ is given priority over job c.*[1]

This definition has been crafted to mimic the heuristic of biasing towards jobs that are (originally) short or have small remaining service requirements. Each of the Properties that make up the definition formalizes a notion of "smart" scheduling. The Bias Property guarantees that the job being run at the server will 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*

[1] Note that every such job a would have had priority over job b at time t due to the Bias Property since $r_a = s_a < s_b = r_b(t)$, where $r_b(t)$ is the remaining size of b at time t.

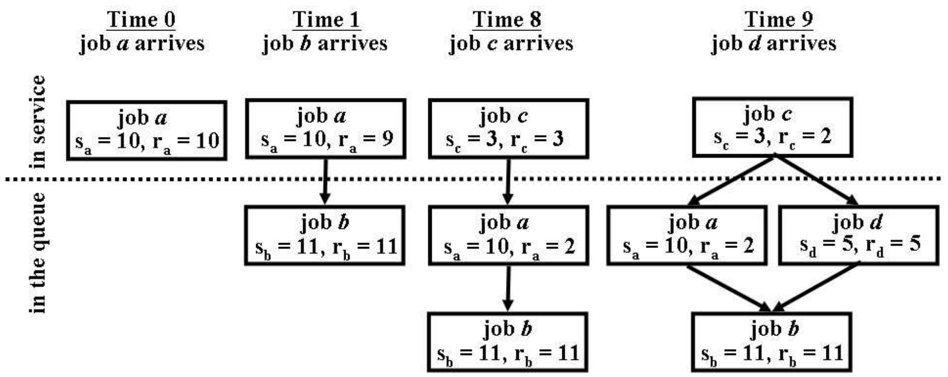


Figure 1: This example illustrates that the SMART definition only enforces a partial ordering on the priorities of jobs in the system. Thus a SMART policy may change its priority rule over time, e.g. from PSJF to SRPT at time 9 in the example. In the diagram, an arrow from a to b indicates that a has priority over b. Up until time 9, jobs have been scheduled according to PSJF . However, after time 9, if PSJF scheduling is continued, job d will receive service before job a, and if SRPT is used instead, job a will receive service before job d. Both of these choices are possible regardless of the priority rule used up to time 9.

arrival of size greater than x while a previous arrival of original size x remains in the system. The Consistency Property prevents time-sharing by guaranteeing that after job a is chosen to run ahead of b, job b will never run ahead of job a. The Transitivity Property guarantees that SMART policies do not second guess themselves: if an arrival a is determined to be "better" than job b, future arrivals that are smaller than a must also be considered "better" than b until b receives service.

The first thing to notice about the class of SMART policies is that many common policies are SMART. The SMART class includes 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. Likewise, the SMART class includes many generalizations of these policies. 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. We now introduce the SMART$*$ classification, a subset of SMART, in order to illustrate the breadth of the static priority schemes that are allowed in the SMART class. Note that SMART$*$ includes all common SMART policies (e.g. SRPT, PSJF, and RS).

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

THEOREM 3.1. SMART$*$ $\subsetneq$ SMART

PROOF. Suppose policy $P \in$ SMART$*$. To see that the Bias Property is maintained, let s_1 and r_1 be the original size and current remaining size of a tagged job in the queue. Suppose s_2 and r_2 correspond to the the original size and current remaining size of another job in the queue such that $r_2 > s_1$. It follows that $s_2 > r_2 > s_1 > r_1$. Thus, $p(s_2, r_2) < p(s_1, r_1)$, so job 2 will not be served.

To see that SMART$*$ policies obey the Consistency Property observe that $p(s, r_1) > p(s, r_2)$ for $r_1 < r_2$ under all SMART$*$ policies. Thus, while serving, a job can only increase its priority, which is already the highest in the system.

To see that SMART$*$ policies obey the Transitivity Property, assume that an arrival with size s_2 preempts a job in service with size s_1 and remaining size r_1. Thus $p(s_1, r_1) < p(s_2, s_2)$. Under any SMART$*$ policy, a future arrival of size $s_3 < s_2$, must have $p(s_3, s_3) > p(s_2, s_2) > p(s_1, r_1)$, which completes the argument.

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$*$. One such example is a policy P that simply alternates the priority function across busy periods, i.e. uses priority function $p_1(s, r)$ for odd numbered busy periods and priority function $p_2(s, r)$ for even numbered busy periods where $p_1 \neq p_2$ are both in SMART$*$. □

Beyond the static priority policies of SMART$*$, SMART policies can also change how they make decisions based on system state information, randomization, or even arbitrarily as long as the Bias, Consistency, and Transitivity Properties are maintained. These generalizations are possible because the SMART definition does not force a total ordering on the priorities of jobs in the system. Instead, only a *partial ordering* is forced, and thus SMART policies can, for instance, change how the policy makes decisions at arrival and departure instants. See Figure 1 for an example. Traditional analysis of scheduling policies assumes that policies obey one fixed rule. In analyzing SMART policies, we are analyzing policies that may change their prioritization rule over time.

The power of the SMART classification is that we can show that all SMART policies have near optimal $E[T]$. Think of SMART policies as policies that provide "SMAll Response Times"[2] by "doing the smart thing." However, SMART policies can differ significantly in their performance on other metrics. For instance, by incorporating the original size and the remaining size into the priority scheme, the RS policy and its variations are able to improve mean slowdown over SRPT when the service distribution is highly variable. Thus, the SMART class provides a starting point for picking a scheduling policy when an application wants to optimize for both $E[T]$ and some other metric of interest.

Despite its breadth, many policies are excluded from SMART. The

[2] We thank Hanoch Levy for his suggestion of this acronym.

class of SMART policies does not include any non-preemptive policies, not even Shortest-Job-First (SJF); nor does it include any age based policies, not even Least-Attained-Service (LAS). This restrictiveness is necessary in order to show that SMART policies provide near optimal $E[T]$ across all service distributions and all loads. For example, though LAS can provide near optimal $E[T]$ under service distributions having decreasing failure rates, when the service distribution has an increasing failure rate LAS is far from optimal.

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 $E[T]$ under SMART policies in Section 5.

Define V_x^{SRPT} to be the steady state work in the system with remaining size less than x under SRPT. Further, define $B_x(y)$ to be the length of a busy period started by a job of size y and made up of only arrivals having size less than x.

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

$$E[T(x)]^P \leq \frac{x}{1-\rho(x)} + \frac{\lambda \widetilde{m}_2(x)}{2(1-\rho(x))^2}$$

Further,

$$T(x)^P \leq_{st} B_x\left(x + V_x^{SRPT}\right)$$

Observe that the upper bound on $E[T(x)]^P$ for $P \in$ SMART in Theorem 4.1 is a combination of the residence time of PSJF, $x/(1-\rho(x))$, and the waiting time of SRPT. Intuitively, this is not surprising. PSJF maximizes residence time among SMART policies because it allows the greatest number of arrivals to preempt service. SRPT maximizes waiting time among SMART policies because it allows the greatest amount of work already in the system to finish before an arriving job. This observation illustrates the tightness of the upper bound and the proof of the theorem formalizes of these ideas. Note that, though the following proof of Theorem 4.1 for SMART is quite involved, a simpler proof is possible if Theorem 4.1 is proven instead only for $P \in$ SMART$*$. The fixed priority structure used in SMART$*$ policies significantly simplifies the proof.

PROOF. We break up the mean response time for a tagged job j_x of size x arriving to the steady state system at time t_{j_x} into: (i) V_x^P, the portion of the work in the system when j_x arrives that will complete under P before j_x completes, (ii) x work made up by j_x, and (iii) the work done by P on jobs that arrive after j_x arrives.

Notice that the Bias Property guarantees that (iii) includes, at most, all arriving jobs of size less than x. Thus, we can stochastically upper bound $T(x)^P$ with the length of a busy period started by $x + V_x^P$ work and made up of only arrivals having size $< x$:

$$T(x)^P \leq_{st} B_x(x + V_x^P)$$

for $P \in$ SMART. In expectation, this gives:

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

It remains to bound V_x^P. We will show that $V_x^P \leq_{st} V_x^{SRPT}$ for any $P \in$ SMART. Noting that [20]:

$$E[V_x^{SRPT}] = \frac{\lambda \widetilde{m}_2(x)}{2(1-\rho(x))}$$

this will complete the proof.

In the remainder of the proof, in order to analyze V_x^P, we track "contributing" work. At time t_{j_x}, the "contributing" work will be equal to V_x^P.

We define "Small Contributors" as all jobs of original size $< x$. For SMART policies, all Small Contributors in the system at time t_{j_x} serve ahead of j_x and thus add their remaining size at time t_{j_x} to the response time of job j_x. We say a Small Contributor is "contributing" the whole time that it is in the system and its "contribution" at any time is its remaining size. Thus, at time t_{j_x} every Small Contributor in the system is "contributing" the amount of work it adds to the response time of j_x.

We define "Large Jobs" as all jobs of original size $\geq x$. For all SMART policies, at most *one* Large Job, c, in the system at time t_{j_x} can add to the response time of job j_x; call job c a "Large Contributor." The uniqueness of c is proven in Lemma 4.1. We say that Large Job c becomes a Large Contributor when r_c becomes x. The amount job c adds to the response time of j_x is the remaining size of c at time t_{j_x}, which can be at most x. We consider c to be "contributing" r_c at all times when $r_c \leq x$. Thus, at time t_{j_x}, c is "contributing" the amount it adds to the response time of j_x.

We now limit our discussion to times $t \in [t_0, t_{j_x}]$ where t_0 is the last moment before j_x arrives that no job is "contributing." So, at t_0 either a Large Job becomes a Large Contributor, a Small Contributor arrives, or j_x arrives ($t_0 = t_{j_x}$). Further, for $t \in (t_0, t_{j_x})$, there is always either a Large or Small Contributor in the system. We refer to t_0 as the beginning of the "contribution period" into which j_x arrives.

We define $V_x^P(t)$ as the total work being contributed by Small and Large Contributors in the system at time t under P, where, as usual, the definition of Contributors is relative to job j_x arriving at time t_{j_x}. It is important to point out that $V_x^P(t_{j_x}) = V_x^P$, i.e. the work contributing when j_x arrives is exactly the work that will serve ahead of j_x.

There are three types of periods into which j_x can arrive:

Type (a) A period idle of contributing jobs (i.e. $t_{j_x} = t_0$). Thus, job j_x sees $V_x^P(t_0) = 0$ for all $P \in$ SMART.

Type (b) A contribution period started by a Small Contributor b arriving and contributing $s_b < x$. Thus, $V_x^P(t_0) = s_b$ under all $P \in$ SMART.

Type (c) A contribution period started by a Large Job c becoming a Large Contributor and contributing x, i.e. r_c becomes x at time t_0. Thus, $V_x^P(t_0) = x$ under all $P \in$ SMART.

Let p_a^P, p_b^P, and p_c^P be the time-average probability of j_x arriving into a contribution period of type (a), (b), and (c) respectively under policy $P \in$ SMART. Recall that j_x is a Poisson arrival, so PASTA applies. Notice that these are the only legal possibilities for what can occur at time t_0 and that there is zero probability of more than one event happening.

Claim (1) $p_a^P \geq p_a^{SRPT}$, $p_b^P \geq p_b^{SRPT}$, and thus $p_c^P \leq p_c^{SRPT}$.

CLAIM (1): We divide the proof of claim (1) into two parts.

Part (a): We will first show that p_a^P is minimized under SRPT. Under SRPT, the system is idle of Small and Large Contributors exactly when there are no jobs in the system having remaining size $< x$. Using PASTA and the fact that j_x is a Poisson arrival, this gives that $p_a^{SRPT} = 1 - \widetilde{\rho}(x)$, i.e. the time-average idle time in a system having arrival rate λ and job sizes $X_x = \min(x, X)$. All $P \in$ SMART are also guaranteed to be idle of Small and Large Contributors when there are no jobs in the system with remaining size $< x$; however they may also be idle of Contributors when there *exist* jobs in the system with remaining size $< x$ if these jobs will not receive priority over j_x when j_x arrives. Thus, $p_a^P \geq p_a^{SRPT}$.

Part (b): We now prove that $p_b^P \geq p_b^{SRPT}$. A type (b) period is started when a Small Contributor arrives into a system idle of contributors. Small Contributors arrive independently of P according to a Poisson process with rate $\lambda F(x)$. Thus, $p_b^P \geq p_b^{SRPT}$ because SRPT is the least likely $P \in$ SMART to be idle of contributing jobs (from part (a)). It follows that $p_c^P \leq p_c^{SRPT}$ since $p_a^P \geq p_a^{SRPT}$ and $p_b^P \geq p_b^{SRPT}$. We can also see that $p_c^P \leq p_c^{SRPT}$ directly by noting that *all* Large Jobs can become Large Contributors and thus start type (c) periods under SRPT. We are now finished with the proof of claim (1).

Consider what j_x sees when it arrives into the system. With probability $p_a^P \geq p_a^{SRPT}$, j_x sees a type (a) period, and with probability $p_b^P + p_c^P = 1 - p_a^P \leq 1 - p_a^{SRPT} = p_b^{SRPT} + p_c^{SRPT}$, j_x sees a contribution period. Thus, in proving $V_x^P \leq_{st} V_x^{SRPT}$ it suffices analyze the $V_x^P(t_{j_x})$ in a contribution period, i.e. given j_x arrives into a type (a) or (b) period.

We will complete the proof of the theorem by showing that

Claim (2) $V_x^P(t_0) \leq_{st} V_x^{SRPT}(t_0)$, i.e. the initial jump of the contribution period is smaller under P than under SRPT.

Claim (3) For $t \in (t_0, t_{j_x})$, $V_x^P(t)$ is always reduced at the full service rate and increases only at the Poisson arrivals of Small Contributors under all $P \in$ SMART.

Claim (4) $V_x^P(t_{j_x}) \leq_{st} V_x^{SRPT}(t_{j_x})$ for Poisson arrival j_x during a contribution period.

CLAIM (2): Note that the initial contribution in a type (b) period is at most the initial contribution in a type (c) period. The claim then follows because $p_b^P \geq p_b^{SRPT}$ and $p_c^P \leq p_c^{SRPT}$.

CLAIM (3): To prove claim (3), notice that, under all $P \in$ SMART, Large Jobs that are not Large Contributors cannot receive service given $t \in (t_0, t_{j_x})$ (Lemma 4.1). Thus, all $P \in$ SMART reduce $V_x^P(t)$ at the maximal rate for all t, i.e. the full service rate is devoted to contributing jobs. Further, under all $P \in$ SMART, arriving Large Jobs cannot become Large Contributors after time t_0 (Lemma 4.1). Thus, the only arrivals that affect $V_x^P(t)$ are Small Contributors, which arrive according to a Poisson process of rate $\lambda F(x)$ under all $P \in$ SMART, including SRPT.

CLAIM (4): To prove claim (4) we will analyze the contributing work that j_x sees upon arrival into a contribution period under $P \in$ SMART and SRPT . Note that j_x arriving into a contribution period under P sees $V_x^P | (V_x^P > 0)$ contributing work. By claim (2), $V_x^P(t_0) \leq_{st} V_x^{SRPT}(t_0)$. Thus, there is some random time $t^* >$ t_0 when $V_x^P(t_0) \stackrel{d}{=} V_x^{SRPT}(t^*)$ for the first time. If $t_{j_x} \geq t^* \geq t_0$ under SRPT then $V_x^P(t_{j_x}) \stackrel{\text{def}}{=} V_x^{SRPT}(t_{j_x})$ (by claim (3) and the definition of t^*). If $t_0 < t_{j_x} < t^*$, then j_x sees a stochastically larger amount of contributing work (by the definition of t^*). So, $V_x^P(t_{j_x}) \leq_{st} V_x^{SRPT}(t_{j_x})$. □

We now prove the Lemma used in the proof of Theorem 4.1.

LEMMA 4.1. *There is at most one Large Contributor in the system at any time, where a Large Contributor is defined with respect to job j_x. Further, no Large Jobs that are not Large Contributors can receive service while a Large or Small Contributor is in the system.*

PROOF. Suppose b becomes a Large Contributor at time t_1 and is the only Large Contributor in the system at t_1. We will show that no other Large Jobs can become Large Contributors while b is in the system.

Note that a Large Job must be receiving service when it becomes a Large Contributor, and thus a Large Job can only become a Large Contributor when the system is idle of Small Contributors due to the Bias Property.

We first show that a Large Job $c \neq b$, in the system at time t_1, cannot become a Large Contributor. Note that c is, by definition, not a Large Contributor at t_1, and thus must receive service in order to become a Large Contributor. Further, c is in the queue at t_1 and b is at the server. So c can never receive service while b is in the system because of the Consistency Property.

To complete the proof, we will show that a Large Job c that arrives after t_1 cannot become a Large Contributor. Again, c must receive service before time t_{j_x} in order to become a Large Contributor. Further, c must be in the system at time t_{j_x} to be a Large Contributor. However, upon arrival $s_c = r_c > x$, so if job c runs ahead of job b, the Consistency Property gives job c priority over job b. Further, since c is in the system at time t_{j_x}, b cannot receive service until then, and thus the Transitivity Property will give j_x priority over b when j_x arrives. This contradicts the fact that b is a Large Contributor. Thus c can never run ahead of b, and c can never become a Large Contributor. □

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.

We derive two types of bounds. The first type illustrates that all SMART policies are near optimal in a very strong sense: they all have $E[T]$ within a factor of 2 of optimal.

THEOREM 5.1. *For $P \in$ SMART:*

$$\begin{aligned} E[T]^{SRPT} &\leq E[T]^P \leq 2E[T]^{SRPT} \\ E[T]^{SRPT} &\leq E[T]^{PSJF} \leq \frac{3}{2}E[T]^{SRPT} \end{aligned}$$

We prove these bounds in Section 5.3. These bounds serve to validate the heuristic of "biasing towards small job sizes," but they do not provide any simpler representation of $E[T]$ under SMART policies. The second type of result in this section provides computationally simple bounds on $E[T]$ that are insensitive to the variability of the service distribution. The bounds do not involve nested integrals; yet we will see in Section 6 that they are nevertheless accurate. All of these 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]: $E[T]^{PS} = \frac{E[X]}{1-\rho}$. Recall that $C^2[X]$ is the square coefficient of variation of X.

THEOREM 5.2. *Let $f(x)$ be decreasing and define*

$$h(\rho) = \left(\frac{1-\rho}{\rho}\right) \log\left(\frac{1}{1-\rho}\right)$$

Then for $P \in$ SMART:

$$\begin{aligned} h(\rho)E[T]^{PS} &\le E[T]^{SRPT} \le \left(\frac{2}{3}-\frac{\rho}{3}+\frac{1}{3}h(\rho)\right)E[T]^{PS}\\ E[T]^{PSJF} &\le \left(\frac{1}{3}+\frac{2}{3}h(\rho)\right)E[T]^{PS}\\ E[T]^{P} &\le \left(-\frac{1}{6}+\frac{\rho(1-\rho)}{4}\left(2+C^2[X]\right)+\frac{7}{6}h(\rho)\right)E[T]^{PS}\end{aligned}$$

The above bounds are tighter than those previously known relating mean response time under SRPT and PS [2, 8].

An important point to notice is that the bounds for SRPT and PSJF are insensitive to 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 5.2 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 5.2 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 5.2 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 5.2 are presented in greater generality in Theorems 5.4, 5.5, 5.7, 5.8, and 5.9 in this section, where they are stated in terms of a parameter K. This K parameter is a constant such that $\lambda m_2(x) \le Kx\rho(x)$, which serves to bound the $\lambda m_2(x)$ term that arises in Theorem 4.1. Theorem 5.3 shows that the constant K may be set at $\frac{2}{3}$ when the service distribution is decreasing, as has been done in Theorem 5.2. But in more generality, it defines K in a way that is highly tied to the tail properties of $f(x)$. Note that $K \le 1$ under all service distributions.

THEOREM 5.3. *Let i be a positive integer. Define j such that $x^j f(x)$ is decreasing and $j < i+1$. Then,*

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

We defer the proof of Theorem 5.3 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 lemmas.

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 calculate the residence time, $E[R]^{PSJF}$, and then bound the waiting time, $E[W]^{PSJF}$. 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 $\lambda m_2(x) \le Kx\rho(x)$. 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:

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

□

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

$$\begin{aligned} E[W]^{PSJF} &\ge \frac{\lambda}{2}\int_0^\infty f(x)\int_0^x t^2 f(t)dtdx\\ &= \frac{\lambda}{4}\int_0^\infty 2t^2 f(t)\overline{F}(t)dt\end{aligned}$$

□

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

THEOREM 5.4. *Let K satisfy $\lambda m_2(x) \le Kx\rho(x)$. 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. The result follows from Lemmas 5.1 and 5.2. □

THEOREM 5.5.

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

PROOF. The result follows from Lemmas 5.1 and 5.3. □

5.2 Bounding mean response time under SRPT

Using the results from the previous section and the technical lemmas in Appendix B, we can now derive bounds on $E[T]^{SRPT}$. Similar bounds have been derived in the case of the M/M/1/SRPT queue with the focus of understanding the performance of SRPT as $\rho \to 1$ [1]. Our goal is to obtain bounds on the M/GI/1/SRPT queue that are tight across ρ and G. To do this, we first bound the mean residence time, $E[R]^{SRPT}$.

LEMMA 5.4.

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

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

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

□

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 5.6.

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

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 $E[T]^{SRPT}$.

THEOREM 5.7. *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 Lemmas B.1 and B.4, we have:

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

□

THEOREM 5.8.

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

PROOF. Using Lemma B.5, we have:

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

□

An interesting observation about Theorem 5.8 is that the lower bound we have proven is exactly the mean residence time under PSJF. Further, Theorem 5.8 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 5.6) 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 $E[W_2]$ is stated in Lemma B.1: $2E[W_2] = E[R]^{PSJF} - E[R]^{SRPT}$.

LEMMA 5.5. *For $P \in$ SMART:*

$$E[T]^P \leq E[T]^{SRPT} + 2E[W_2]$$

Using the previous lemma, we can prove Theorem 5.1.

PROOF. (of Theorem 5.1)

We will prove the first statement only, since the second statement follows using the same technique.

It is clear that $E[T]^{SRPT} \leq E[T]^P$ because SRPT is optimal with respect to mean response time. Thus we need only show the upper bound. Using Lemmas 5.5 and B.5, we have

$$\begin{aligned}
E[T]^P &\leq E[T]^{SRPT} + 2E[W_2] \\
&= E[T]^{SRPT}\left(1 + 2\frac{\frac{1}{2}E[W_2] + \frac{1}{2}E[W_2]}{E[T]^{SRPT}}\right) \\
&\leq E[T]^{SRPT}\left(1 + \frac{E[W]^{PSJF} + E[W_2]}{E[T]^{SRPT}}\right) \\
&\leq 2E[T]^{SRPT}
\end{aligned}$$

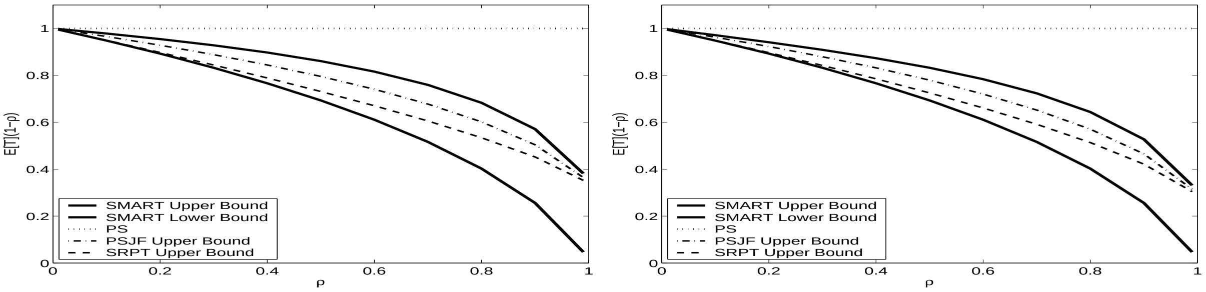
□

We are now ready to upper bound the mean response time of policies in SMART.

THEOREM 5.9. *Let K satisfy $\lambda m_2(x) \leq Kx\rho(x)$. Then for $P \in$ SMART:*

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

PROOF. Using Theorem 5.4, Lemma B.1, and Lemma 5.4, we



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

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

Figure 2: 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.

have:

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

□

Theorems 5.9 and 5.8 give simple benchmarks that provide upper and lower bounds on the mean response times of “smart” scheduling policies. These bounds will hopefully facilitate the evaluation of policies that are not `SMART` but still claim to provide good mean response time.

5.4 A proof of Theorem 5.3

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

PROOF. (of Theorem 5.3)
First, we observe the following equality:

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

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 $j < i+1$.

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

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.

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

□

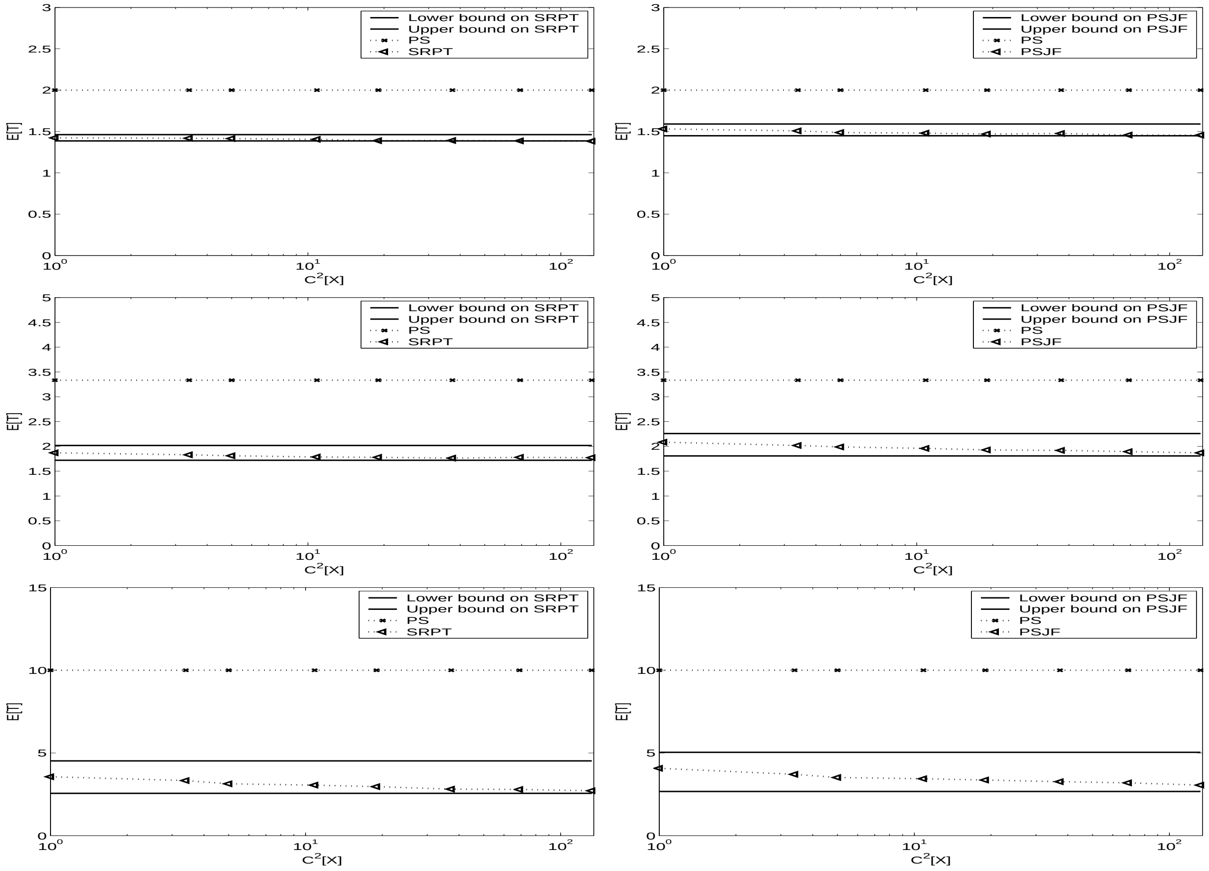
A few comments are in order about this theorem. First, notice that in this work we only apply the lemma in the case where $i = 1$, but the more general form is useful for investigating higher moments. Second, notice that because K is defined in terms of j, where j is such that $x^j f(x)$ is decreasing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

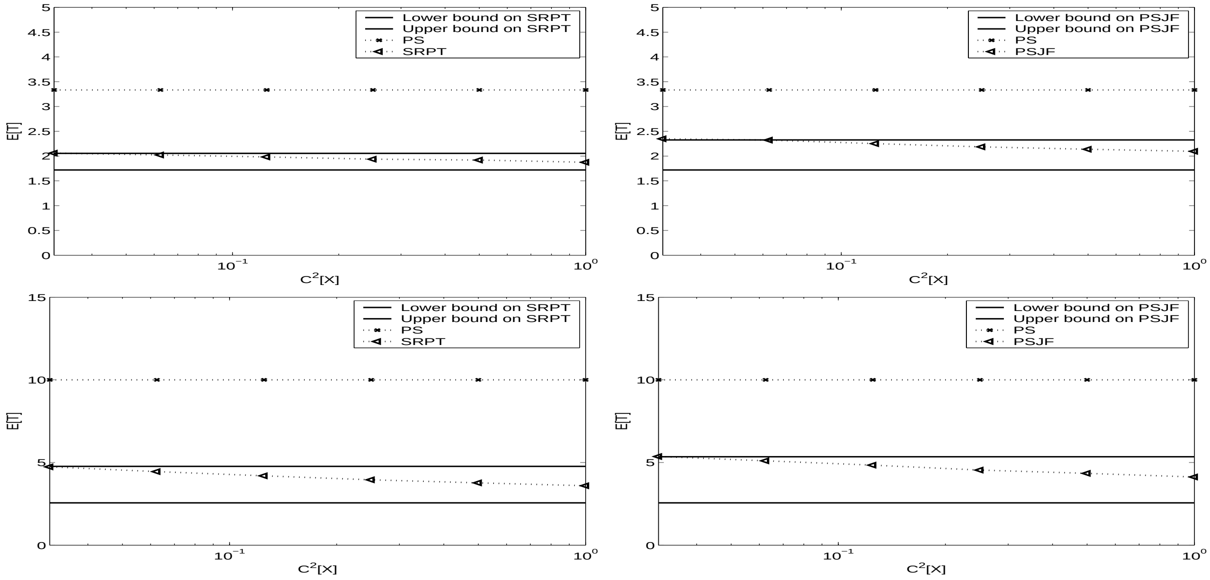
The Weibull and Erlang distributions are convenient ways to evaluate the effects of variability in the service distribution because they allow a wide range of variability and tail behavior. Investigating the effect of the weight of the tail of the service distribution is important in light of many recent measurements that have observed job size distributions that are well-modeled by heavy tailed distributions such as the Weibull distribution [3, 6, 12, 16].

The goal in investigating how the bounds perform under these service distributions is twofold. Our first goal is to illustrate the



(a) SRPT (b) PSJF

Figure 3: 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 a 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.



(a) SRPT (b) PSJF

Figure 4: 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 0.7 and 0.9 in the first and second rows respectively. These plots illustrate that the upper bounds on both PSJF and SRPT are tight.

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 insensitive to 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 insensitive bounds 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 $\overline{F}(x;b,c) = e^{-\left(\frac{x}{b}\right)^c}$.

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. 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.

First, in Figure 2, the bounds in Theorem 5.2 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 SMART policy 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.

Second, in Figure 3, the bounds for SRPT and PSJF in Theorem 5.2 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 in 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 3, 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 illustrate this in the next section.

An important point that Figure 3 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 $C^2[X] > 1$ under moderate ρ. 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. The $Erl(n,\mu)$ distribution is the convolution 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, smaller j are necessary in order to guarantee that $x^j f(x)$ is decreasing. This second point means that, for Erlang distributions, we must weaken the bounds by setting $K = 1$ (as discussed in Section 5.4).

In Figure 4, 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. Throughout these plots, the mean of the service distribution is fixed at 1, and $C^2[X]$ is allowed to vary.

Figure 4 illustrates that the upper bounds on SRPT and PSJF are tight for distributions with low variability.

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.

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 RS improve 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). Interestingly, SMART policies do not necessarily obey a static priority rule, but may also switch between different priority rules (e.g. changing between SRPT and PSJF over time). We prove *simple* upper and lower bounds on the mean response of any SMART policy. These bounds show that all SMART policies have $E[T]$ within a factor of 2 of optimal. 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 .

An unanticipated discovery of this work is the near insensitivity of SMART policies to the variability of the job size distribution (particularly when $C^2[X] > 1$). It is well known that the mean response time of PS is insensitive to the service distribution's variability, but the fact that mean response time for policies like SRPT and PSJF is nearly insensitive of the service distribution's variability is counter the folklore of the community.

Beyond the definition of the SMART class, we believe some of the observations in this work can impact 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. Second, 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.

This work is only the first step towards characterizing SMART policies. Many interesting questions remain. Can the definition of SMART be loosened to include other policies with $E[T]^P \leq 2E[T]^{SRPT}$? If only a limited class of service distributions are considered (e.g. distributions with decreasing failure rates), how can SMART be extended? Are SMART policies near optimal for measures other than $E[T]$?

8. REFERENCES

[1] N. Bansal. On the average sojourn time under M/M/1/SRPT. *Oper. Res. Letters*, 22(2):195–200, 2005.

[2] N. Bansal and M. Harchol-Balter. Analysis of SRPT scheduling: Investigating unfairness. In *Proceedings of ACM Sigmetrics*, 2001.

[3] P. Barford and M. Crovella. Generating representative web workloads for network and server performance evaluation. In *Proceedings of ACM Sigmetrics*, 1998.

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

[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.

[8] M. Gong and C. Williamson. Quantifying the properties of SRPT scheduling. In *IEEE/ACM Symposium on Mod., Anal., and Sim. of Comp. and Telecomm. Sys. (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] R. Nunez-Queija. Queues with equally heavy sojourn time and service requirement distributions. *Ann. Oper. Res*, 113:101–117, 2002.

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

[15] 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.

[16] D. L. Peterson. Data center I/O patterns and power laws. In *CMG Proceedings*, December 1996.

[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*, 2003.

[18] M. Rawat and A. Kshemkalyani. SWIFT: Scheduling in web servers for fast response time. In *Symp. on Network Computing and App.*, 2003.

[19] L. E. Schrage. A proof of the optimality of the shortest remaining processing time discipline. *Operations Research*, 16:678–690, 1968.

[20] 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.

[21] D. Smith. A new proof of the optimality of the shortest remaining processing time discipline. *Operations Research*, 26:197–199, 1976.

[22] 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.

[23] A. Wierman and M. Harchol-Balter. Classifying scheduling policies with respect to unfairness in an M/GI/1. In *Proceedings of ACM Sigmetrics*, 2003.

APPENDIX

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.

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

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.

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

□

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.

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

□

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:

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

□

LEMMA B.2.

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

PROOF. Using Lemma A.1, we have:

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

□

LEMMA B.3.

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

PROOF. Using Lemma A.1, we have:

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

□

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

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

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

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

□

LEMMA B.5.

$$\begin{aligned}E[R]^{SRPT} + 2E[W]^{PSJF} &\ge E[R]^{PSJF} \quad \textit{and thus} \\ E[W]^{PSJF} &\ge E[W_2]\end{aligned}$$

PROOF. Using Lemma B.3, we have:

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

Further, combining the above with Lemma B.1, we have:

$$E[W]^{PSJF} \ge \frac{1}{2}\left(E[R]^{PSJF} - E[R]^{SRPT}\right) = E[W_2]$$

□