Formalizing SMART Scheduling*

Adam Wierman
School of Computer Science
Carnegie Mellon University
acw@cs.cmu.edu

Mor Harchol-Balter
School of Computer Science
Carnegie Mellon University
harchol@cs.cmu.edu

1. INTRODUCTION

It is well-known that policies which bias towards small job sizes 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 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 [1]. The heuristic has also been applied to other application areas; for example, scheduling in supercomputing centers.

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 [3]; 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 [4] and PSJF[2], 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 imprecise at high loads. No *simple* closed form formula is known for either of these policies. 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.[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. Most importantly however, these bounds are simple functions of the system load and thus provide accurate, back-of-the-envelope calculations that can be used to understand the mean response times of these policies.

Throughout the paper we will consider 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 $x \geq 0$. Let $\overline{F}(x) \stackrel{\text{def}}{=} 1 - F(x)$. 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 $E[T]^P = \int_0^\infty E[T(x)]^P f(x)dx$. Further, define $m_i(x) \stackrel{\text{def}}{=} \int_0^x t^i f(t)dt$ and $\rho(x) = \lambda m_1(x)$.

2. DEFINING THE SMART CLASS

We define the **SMART** class of scheduling policies as follows:

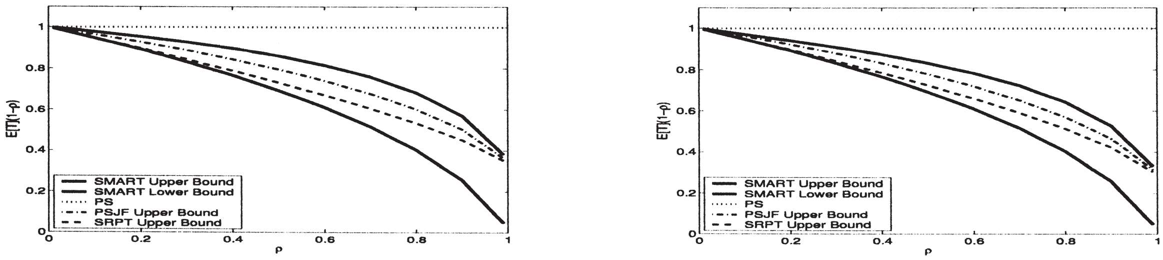
DEFINITION 2.1. *A work conserving policy $P \in$* **SMART** *if (i) a job of remaining size greater than x can never have priority over a job of original size x, and (ii) a job being run at the server can only be preempted by new arrivals.*

This definition has been crafted to mimic the heuristic of biasing towards jobs that are (originally) short or have small remaining service requirements. 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 **SMART** policies do not second guess themselves; thus if job a that is running currently has priority over job b, then job b will not preempt job a.

The class of **SMART** policies is very broad. Although the class does not include non-preemptive policies, not even Shortest-Job-First, 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

* This work was supported by NSF Career Grant CCR-0133077, NSF ITR Grant 99-167 ANI-0081396, Cisco Systems, Spinnaker Networks via Pittsburgh Digital Greenhouse Grant 01-1, and an NSF Graduate Research Fellowship.

[1] We thank Hanoch Levy for his observation at the MAMA workshop that **SMART** can be viewed as an acronym for "SMAll Response Times."



(a) $C^2 = 1$ (b) $C^2 = 10.865$

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) $C^2 = 1$, (b) $C^2 = 10.865$, respectively (where $C^2 = E[X^2]/E[X]^2 - 1$).

can achieve up to a 10% relative improvement in mean slowdown over SRPT. 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^iS^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).

3. BOUNDING THE RESPONSE TIME UNDER SMART POLICIES

We now present both an upper bound on the mean response time for a job of size x, $E[T(x)]$, under policies in SMART and bounds on the overall mean response time, $E[T]$, of policies in SMART.

THEOREM 3.1. *The mean response time for a job of size x under any policy $P \in$ SMART satisfies:*

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

The proof of this theorem combines the tagged job technique with busy period analyses and conservation techniques [5].

We now present bounds on the overall mean response time of policies in SMART. It is important to notice that all these bounds are very simple: they do not involve nested integrals; yet they are nevertheless accurate. All of the bounds are 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. The overall mean response time under PS is $E[T]^{PS} = \frac{E[X]}{1-\rho}$ [2]. Further, define C to be the coefficient of variation of the service distribution.

THEOREM 3.2. *Let $f(x)$ be decreasing*[2]. *Define* $h(\rho) = \left(\frac{1-\rho}{\rho}\right)\log(1-\rho)$. *Then*

$$-h(\rho)E[T]^{PS} \leq E[T]^{SRPT} \leq \left(\frac{2}{3} - \frac{\rho}{3} - \frac{1}{3}h(\rho)\right)E[T]^{PS}$$

$$E[T]^{PSJF} \leq \left(\frac{1}{3} - \frac{2}{3}h(\rho)\right)E[T]^{PS}$$

$$E[T]^{SMART} \leq \left(-\frac{1}{6} + \frac{\rho(1-\rho)}{4}(2+C^2) - \frac{7}{6}h(\rho)\right)E[T]^{PS}$$

Surprisingly, these upper and lower bounds are quite close, leading us to conclude that, although the SMART class includes many different policies, all SMART policies are similar with respect to mean response time. In fact, all are far superior to PS, and most importantly, all have mean response time quite close to the optimal (see Figure 1). This result theoretically validates the heuristic of "biasing towards small job sizes" that system designers apply.

The bounds show that SRPT and PSJF are nearly insensitive to the variability of the service distribution. Although, 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 Theorem 3.2 illuminate this practical property. Further, 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 (see [5]).

Another important point about Theorem 3.2 is that 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. 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 on SRPT of $-\frac{1}{\lambda}\log(1-\rho)$, which is a much simpler calculation and is tight for highly variable distributions.

[2] These results are generalized beyond $f(x)$ decreasing in [5], where they are stated in terms of a parameter K. This K is a service-distribution dependent parameter such that $\lambda m_2(x) \leq Kx\rho(x)$, which serves to bound the $\lambda m_2(x)$ term that arises in Theorem 3.1.

4. REFERENCES

[1] 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.

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

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

[4] 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.

[5] A. Wierman and M. Harchol-Balter. Bounding the performance of SMART scheduling policies. Technical Report CMU-CS-03-199, Carnegie Mellon University, 2003.