Available online at www.sciencedirect.com
SCIENCE DIRECT
Operations Research Letters 33 (2005) 519 – 524
Operations Research Letters
www.elsevier.com/locate/orl

M/G/1/MLPS compared to M/G/1/PS

Samuli Aalto[a,*], Urtzi Ayesta[b], Eeva Nyberg-Oksanen[a,1]

[a] *Networking Laboratory, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Helsinki, Finland*
[b] *France Telecom R & D and INRIA Sophia Antipolis, France*

Received 21 June 2004; accepted 20 September 2004

Abstract

Multilevel processor sharing scheduling disciplines have recently been resurrected in papers that focus on the differentiation between short and long TCP flows in the Internet. We prove that, for M/G/1 queues, such disciplines are better than the processor sharing discipline with respect to the mean delay whenever the hazard rate of the service time distribution is decreasing.

Keywords: Scheduling; Multilevel processor sharing; Mean delay; Unfinished truncated work; M/G/1

1. Introduction

We consider multilevel processor sharing (MLPS) scheduling disciplines in the context of M/G/1 queues. MLPS disciplines were introduced by Kleinrock in the early 1970s, see [9]. An MLPS discipline π is defined by a finite set of thresholds $a_1 < \cdots < a_N$ defining $N+1$ levels, $N \geqslant 0$. A job belongs to level n if its attained service is at least a_{n-1} but less than a_n, where $a_0 = 0$ and $a_{N+1} = \infty$. Between these levels, a strict priority discipline is applied with the lowest level having the highest priority. Thus, those jobs with attained service less than a_1 are served first. Within each level n, an internal discipline π_n is applied. We let the internal disciplines vary in the set $\{\mathrm{FB}, \mathrm{PS}\}$, where FB refers to the foreground–background discipline that gives priority to the job with the least-attained service and PS to the processor sharing discipline that shares the service capacity evenly among all jobs.

The MLPS disciplines form a subset of a larger family of scheduling disciplines that are based on the attained service of jobs. Yashkov has proven that FB minimizes the mean delay among such disciplines whenever the service time distribution is of type decreasing hazard rate (DHR), see [15]. Righter and Shanthikumar [12] proved that, under the DHR condition, FB minimizes the queue length even stochastically. Righter et al. [13] showed that FB minimizes the mean delay whenever the service time distribution is of type increasing mean residual life (IMRL), which is a weaker condition than DHR. Recently, Wierman

* Corresponding author.
E-mail address: samuli.aalto@hut.fi (S. Aalto).
[1] Partly supported by the Academy of Finland, and the Nokia and TES foundations.

0167-6377/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2004.09.009

et al. [14] proved that FB is better than PS with respect to the mean delay whenever the service time distribution is of type DHR, and vice versa if the service time distribution is of type increasing hazard rate (IHR). A fundamental fact behind these results is the following extremal property of FB regarding unfinished truncated work U_x, which refers to the sum of remaining truncated service times of the jobs in the system: FB minimizes U_x at every moment in each sample path for all truncation thresholds x, independent of the service time distribution type [1, Proposition 5].

The PS discipline has been proposed as an appropriate model for the bandwidth sharing among TCP flows in a bottleneck router [3,8,11]. On the other hand, MLPS disciplines have recently been resurrected in some papers that focus on the differentiation between short and long TCP flows in the Internet [2,6,7]. Flow sizes in the Internet have been modelled by, e.g., Pareto and hyperexponential distributions [4,5]. The latter type satisfies the DHR condition, while the Pareto distribution defined in [4] by

$$P\{X \leqslant x\} = 1 - \left(\frac{k}{x}\right)^{\alpha}, \quad x \geqslant k,$$

has a decreasing hazard rate only from the lower limit k on. However, an arbitrarily small k can be chosen while keeping fixed the shape parameter α that controls the rate at which the tail disappears. Another possibility is to define the Pareto distribution slightly differently by

$$P\{X \leqslant x\} = 1 - \left(\frac{1}{1+cx}\right)^{\alpha}, \quad x \geqslant 0,$$

as done, e.g., in [5,10]. This distribution type satisfies the DHR condition.

In [1], we proved that the MLPS disciplines with just two levels are better than PS with respect to the mean delay whenever the hazard rate of the service time distribution is decreasing, and vice versa if the hazard rate is increasing and bounded. In this paper we show that these results are valid for *any* MLPS discipline.

The paper is organized as follows. The notation and the essential existing results concerning the comparison of MLPS disciplines are given in Section 2, while the new results are developed in Section 3. Section 4 concludes the paper.

2. Notation and existing results

We denote by MLPS the family of MLPS disciplines π for which $\pi_n \in \{\mathrm{FB}, \mathrm{PS}\}$ for all n. Among the disciplines $\{\mathrm{FB}, \mathrm{PS}\}$, we define the following order relation:

$$\mathrm{FB} \preceq \mathrm{FB}, \quad \mathrm{FB} \preceq \mathrm{PS}, \quad \mathrm{PS} \not\preceq \mathrm{FB}, \quad \mathrm{PS} \preceq \mathrm{PS}.$$

Furthermore, we denote by $(N+1)$PS the family of MLPS disciplines with $N+1$ levels (and N thresholds) that use PS as the internal scheduling discipline within all the levels. Thus, 1PS refers to the PS discipline alone, 2PS to the PS + PS disciplines, 3PS to the PS + PS + PS disciplines, etc. Finally, we denote by TLPS the family of MLPS disciplines that have just two levels, i.e., $N = 1$.

In this section we present the results concerning the comparison of MLPS disciplines from our previous work [1] that form a basis for the new results to be presented in the following section. The results are grouped into two subsections: the first reviews existing sample path results and the second existing mean value results.

2.1. Sample path results

Consider a single-server queueing system starting empty at time $t=0$ and obeying a scheduling discipline $\pi \in$ MLPS. We assume that the jobs arrive one at a time. Let A_i denote the arrival time of job i, S_i its service time, and $X_i^\pi(t)$ its attained service at time t. Let $\mathscr{A}(t)$ denote the set of jobs arrived until time t,

$$\mathscr{A}(t) = \{i : A_i \leqslant t\},$$

$\mathscr{N}^\pi(t)$ the set of jobs in the system at time t,

$$\mathscr{N}^\pi(t) = \{i \in \mathscr{A}(t) : X_i^\pi(t) < S_i\},$$

and $N^\pi(t) = |\mathscr{N}^\pi(t)|$. Furthermore, for all $x \geqslant 0$, let $\mathscr{N}_x^\pi(t)$ denote the set of jobs whose attained service is less than x,

$$\mathscr{N}_x^\pi(t) = \{i \in \mathscr{A}(t) : X_i^\pi(t) < \min\{S_i, x\}\},$$

$N_x^\pi(t) = |\mathscr{N}_x^\pi(t)|$, and $U_x^\pi(t)$ the unfinished truncated work with truncation threshold x at time t,

$$U_x^\pi(t) = \sum_{i \in \mathscr{N}_x^\pi(t)} (\min\{S_i, x\} - X_i^\pi(t)).$$

Proposition 1 (*Aalto et al.* [1, Proposition 8]). *Let $\pi, \pi' \in$ MLPS with the same thresholds $\{a_1, \ldots, a_N\}$ such that $\pi_n \preceq \pi'_n$ for all $n \in \{1, \ldots, N+1\}$. Then $U_x^{\pi}(t) \leqslant U_x^{\pi'}(t)$ for all $x \geqslant 0$ and $t \geqslant 0$.*

Proposition 2. *Let $\pi \in$ MLPS with thresholds $\{a_1, \ldots, a_N\}$ and $\pi' \in (N+1)$PS with the same thresholds $\{a_1, \ldots, a_N\}$. Then $U_x^{\pi}(t) \leqslant U_x^{\pi'}(t)$ for all $x \geqslant 0$ and $t \geqslant 0$.*

Proof. This follows immediately from Proposition 1 since $\pi_n \preceq \mathrm{PS} = \pi'_n$ for all n. □

2.2. Mean value results

Consider an M/G/1 queue obeying a scheduling discipline $\pi \in$ MLPS. Let λ denote the arrival rate and S the service time of a job. We assume that $E[S] < \infty$ and that the system is stable, i.e., $\rho = \lambda E[S] < 1$. Furthermore, we assume that the service time distribution is continuous with the corresponding density function denoted by $f(x)$. Let $F(x) = \int_0^x f(y)\,\mathrm{d}y$ and $\overline{F}(x) = 1 - F(x)$. The corresponding hazard rate function is denoted by $h(x) = f(x)/\overline{F}(x)$.

Let U_x^{π} denote the unfinished truncated work with truncation threshold x and $T^{\pi}(y)$ the delay of a job with service time of y time units. By [9, Eq. (4.60)],

$$\overline{U}_x^{\pi} = \lambda \int_0^x \overline{F}(y)\overline{T}^{\pi}(y)\,\mathrm{d}y, \tag{1}$$

where $\overline{U}_x^{\pi} = E[U_x^{\pi}]$ and $\overline{T}^{\pi}(y) = E[T^{\pi}(y)]$. Let then T^{π} denote the delay of any job. As explained in [1], it follows from (1) that

$$\overline{T}^{\pi} = \frac{1}{\lambda} \int_0^{\infty} (\overline{U}_x^{\pi})' h(x)\,\mathrm{d}x,$$

where $\overline{T}^{\pi} = E[T^{\pi}]$ and $(\overline{U}_x^{\pi})' = (\mathrm{d}/\mathrm{d}x)E[U_x^{\pi}]$.

In fact, these results, as well as the following one, are valid for all scheduling disciplines that are based on the attained service of jobs. However, for the purposes of this paper, it is sufficient to consider the family of MLPS disciplines.

Proposition 3 (*Aalto et al.* [1, Propositions 1 and 2]). *Let $\pi, \pi' \in$ MLPS such that $\overline{U}_x^{\pi} \leqslant \overline{U}_x^{\pi'}$ for all $x \geqslant 0$.*

(i) *If the hazard rate $h(x)$ is decreasing, then $\overline{T}^{\pi} \leqslant \overline{T}^{\pi'}$.*
(ii) *If the hazard rate $h(x)$ is increasing and bounded, then $\overline{T}^{\pi} \geqslant \overline{T}^{\pi'}$.*

Proposition 4 (*Aalto et al.* [1, Proposition 4]). *Let $\pi \in$ 2PS. Then $\overline{U}_x^{\pi} \leqslant \overline{U}_x^{\mathrm{PS}}$ for all $x \geqslant 0$.*

Proposition 5. *Let $\pi \in$ TLPS. Then $\overline{U}_x^{\pi} \leqslant \overline{U}_x^{\mathrm{PS}}$ for all $x \geqslant 0$.*

Proof. Let $\pi' \in$ 2PS with the same threshold as π. Since

$$\overline{U}_x^{\pi} = \lim_{t\to\infty} \frac{1}{t} \int_0^t U_x^{\pi}(s)\,\mathrm{d}s,$$

Proposition 2implies that $\overline{U}_x^{\pi} \leqslant \overline{U}_x^{\pi'}$ for all $x \geqslant 0$. The claim follows now from Proposition 4. □

Theorem 1. *Let $\pi \in$ TLPS.*

(i) *If the hazard rate $h(x)$ is decreasing, then $\overline{T}^{\pi} \leqslant \overline{T}^{\mathrm{PS}}$.*
(ii) *If the hazard rate $h(x)$ is increasing and bounded, then $\overline{T}^{\pi} \geqslant \overline{T}^{\mathrm{PS}}$.*

Proof. These results follow immediately from Propositions 3 and 5. □

3. New results

In this section we present the new results concerning the comparison of MLPS disciplines. Similarly as in the previous section, the results are grouped into two subsections.

3.1. Sample path results

Consider a single server queueing system starting empty at time $t=0$ and obeying a scheduling discipline $\pi \in$ MLPS. Assume that the jobs arrive one at a time. The notation used is the same as in Section 2.1.

Proposition 6. *Let $N \geqslant 1$, $\pi \in (N+1)$PS with thresholds $\{a_1, \ldots, a_N\}$, and $\pi' \in N$PS with thresholds*

$\{a_1, \ldots, a_{N-1}\}$. *Then* $U_x^{\pi}(t) \leqslant U_x^{\pi'}(t)$ *for all* $x \leqslant a_N$ *and* $t \geqslant 0$.

Proof. First we note that, since the two disciplines follow the same rule as regards the jobs with attained service time less than a_{N-1}, we surely have, for all $t \geqslant 0$,

$$i \in \mathcal{N}^{\pi}_{a_{N-1}}(t) \Rightarrow X_i^{\pi}(t) = X_i^{\pi'}(t). \quad (2)$$

Then we claim that, for all $t \geqslant 0$,

$$i \in \mathcal{N}^{\pi}_{a_N}(t) \backslash \mathcal{N}^{\pi}_{a_{N-1}}(t) \Rightarrow X_i^{\pi}(t) \geqslant X_i^{\pi'}(t). \quad (3)$$

Eqs. (2) and (3) guarantee that $\mathcal{N}^{\pi}_x(t) \subset \mathcal{N}^{\pi'}_x(t)$ for all $x \leqslant a_N$ and $t \geqslant 0$. This, together with (2) and (3), implies that, for all $x \leqslant a_N$ and $t \geqslant 0$,

$$\begin{aligned} U_x^{\pi}(t) &= \sum_{i \in \mathcal{N}^{\pi}_x(t)} (\min\{S_i, x\} - X_i^{\pi}(t)) \\ &\leqslant \sum_{i \in \mathcal{N}^{\pi'}_x(t)} (\min\{S_i, x\} - X_i^{\pi'}(t)) = U_x^{\pi'}(t). \end{aligned}$$

Thus, it remains to prove that (3) is true for all t. The proof given below is an induction with respect to arrival epochs A_k.

1. During the interval $[0, A_1)$ both systems are empty. Thus (3) is trivially true for all $t < A_1$.

2. Let $k \in \{1, 2, \ldots\}$, and assume that (3) is true for all $t < A_k$. We will show that it is also true in the interval $[A_k, A_{k+1})$.

We divide the interval $[A_k, A_{k+1})$ into three consecutive periods I_1, I_2, and I_3, with the following starting (b) and ending (e) points:

$$\begin{aligned} &I_1^{\mathrm{b}} = A_k, \quad I_1^{\mathrm{e}} = \sup\{I_1^{\mathrm{b}} < t \leqslant A_{k+1} | N^{\pi}_{a_{N-1}}(t) > 0\}, \\ &I_2^{\mathrm{b}} = I_1^{\mathrm{e}}, \quad I_2^{\mathrm{e}} = \sup\{I_2^{\mathrm{b}} < t \leqslant A_{k+1} | N^{\pi}_{a_N}(t) > 0\}, \\ &I_3^{\mathrm{b}} = I_2^{\mathrm{e}}, \quad I_3^{\mathrm{e}} = A_{k+1}. \end{aligned}$$

We note that during interval I_1 both systems give service only to those jobs whose attained service is less than a_{N-1}. During interval I_2 there are no longer any such jobs in either system. The system with discipline π gives service to those jobs whose attained service is at least a_{N-1} but less than a_N, while in the other system all the remaining jobs are served at the same time. Finally, in the interval I_3, when there are no longer any jobs with attained service less than a_N in the system with discipline π, all the remaining jobs are served at the same time also in that system. We further note that I_1 is always of positive length, whereas I_2 and I_3 may vanish. The three intervals $I_1 - I_3$ are considered in 2.1–2.3, respectively.

2.1. Consider first the interval I_1. Due to the induction assumption and the fact that, during this interval, strict priority is given (in both systems) to those jobs with attained service time less than a_{N-1}, we have, for all $t \in I_1$ and $i \in \mathcal{N}^{\pi}_{a_N}(t) \backslash \mathcal{N}^{\pi}_{a_{N-1}}(t)$,

$$X_i^{\pi}(t) = X_i^{\pi}((A_k)^-) \geqslant X_i^{\pi'}((A_k)^-) = X_i^{\pi'}(t).$$

Thus, (3) is true for the whole interval I_1. This is enough if the interval I_1 ends at time A_{k+1} when a new job arrives. Otherwise we have to consider, at least, the interval I_2, too.

2.2. Consider then the interval I_2. During this interval, strict priority is given to those jobs with attained service time at least a_{N-1} but less than a_N in the system with discipline π. From (2) and 2.1, we deduce that $X_i^{\pi}(I_2^{\mathrm{b}}) \geqslant X_i^{\pi'}(I_2^{\mathrm{b}})$ for all $i \in \mathcal{N}^{\pi}_{a_N}(I_2^{\mathrm{b}})$ implying that

$$\mathcal{N}^{\pi}_{a_N}(I_2^{\mathrm{b}}) \subset \mathcal{N}^{\pi'}_{a_N}(I_2^{\mathrm{b}}).$$

From time I_2^{b} on, the set $\mathcal{N}^{\pi}_{a_N}(t)$ remains the same until a new job arrives or one of the jobs in this set reaches level a_N or leaves the system. In this subinterval, we have

$$\begin{aligned} \mathcal{N}^{\pi}_{a_N}(t) &= \mathcal{N}^{\pi}_{a_N}(I_2^{\mathrm{b}}) \subset \mathcal{N}^{\pi'}_{a_N}(I_2^{\mathrm{b}}) \\ &= \mathcal{N}^{\pi'}_{a_N}(t) \subset \mathcal{N}^{\pi'}(t), \end{aligned}$$

and the jobs $i \in \mathcal{N}^{\pi}_{a_N}(t)$ in the system with discipline π are served with rate

$$(X_i^{\pi})'(t) = 1/N^{\pi}_{a_N}(t),$$

while in the system with discipline π' they get service with rate

$$(X_i^{\pi'})'(t) = 1/N^{\pi'}(t) \leqslant 1/N^{\pi}_{a_N}(t).$$

Thus, we have $X_i^{\pi}(t) \geqslant X_i^{\pi'}(t)$ for all $i \in \mathcal{N}^{\pi}_{a_N}(t)$ and t in this subinterval. Continuing similarly, it is easy to see that (3) is true for all I_2. This is enough if the interval I_2 ends at time A_{k+1} when a new job arrives. Otherwise we have to consider the final interval I_3, too.

2.3. Consider finally the interval I_3. By definition, $\mathcal{N}^{\pi}_{a_N}(t) = \emptyset$ for all $t \in I_3$. Thus, (3) is trivially true for all $t \in I_3$. $\square$

3.2. Mean value results

Consider a stable M/G/1 queue obeying a scheduling discipline $\pi \in$ MLPS. The notation used is the same as in Section 2.2.

It is well known that the mean delay of a job with service time $x > 0$ in a PS system reads as

$$\overline{T}^{\mathrm{PS}}(x) = \frac{x}{1-\rho}.$$

According to [9, Eqs. (4.27) and (4.39)], the corresponding mean delay in a system with scheduling discipline $\pi \in (N+1)$PS with thresholds $\{a_1, \ldots, a_N\}$ satisfies, for all $x > a_N$,

$$\overline{T}^{\pi}(x) = \overline{T}^{\mathrm{FB}}(a_N) + \frac{\alpha_N(x - a_N)}{1 - \rho_{a_N}}.$$

Here $\rho_{a_N} = \lambda E[\min\{S, a_N\}]$ refers to the "truncated load", and $\alpha_N(x)$ is such that $\alpha'_N(x) = \frac{d}{dx}\alpha_N(x)$ satisfies the following integral equation:

$$\alpha'_N(x) = \frac{\lambda}{1-\rho_{a_N}} \int_0^x \alpha'_N(y)\overline{F}(a_N + x - y)\,\mathrm{d}y + \frac{\lambda}{1-\rho_{a_N}} \int_0^\infty \alpha'_N(y)\overline{F}(a_N + x + y)\,\mathrm{d}y + c_N(x) + 1 \tag{4}$$

with $c_N(x) \geqslant 0$. In addition, $\alpha_N(x)$ is increasing implying that $\alpha'_N(x) \geqslant 0$ for all $x > 0$. Note further that $\overline{T}^{\pi}(x)$ is differentiable, at least, for all $x > a_N$.

Proposition 7. *Let* $\pi \in (N+1)$PS *with thresholds* $\{a_1, \ldots, a_N\}$. *Then* $(\overline{T}^{\pi})'(x) \geqslant (\overline{T}^{\mathrm{PS}})'(x)$ *for all* $x > a_N$.

Proof. This is proved similarly as the corresponding result for the two-level case, see the derivation of the latter part of Eq. (13) in [1]. $\square$

Proposition 8. *Let* $\pi \in (N+1)$PS. *Then* $\overline{U}^{\pi}_x \leqslant \overline{U}^{\mathrm{PS}}_x$ *for all* $x \geqslant 0$.

Proof. The claim is proved by induction.

1. For $N = 1$, the claim is the same as in Proposition 4.

2. Let then $N > 1$ and assume that the claim is true for any $\pi' \in N$PS.

Let $\pi \in (N+1)$PS with thresholds $\{a_1, \ldots, a_N\}$ and $\pi' \in N$PS with thresholds $\{a_1, \ldots, a_{N-1}\}$. By Proposition 6 and the induction assumption above, we have, for all $x \leqslant a_N$,

$$\overline{U}^{\pi}_x \leqslant \overline{U}^{\pi'}_x \leqslant \overline{U}^{\mathrm{PS}}_x.$$

Define then

$$x^* = \inf\{x \geqslant a_N \,|\, \overline{T}^{\pi}(x) \geqslant \overline{T}^{\mathrm{PS}}(x)\}.$$

By (1) we have, for all $x \geqslant a_N$,

$$\overline{U}^{\pi}_x = \overline{U}^{\pi}_{a_N} + \lambda \int_{a_N}^x \overline{F}(t)\overline{T}^{\pi}(t)\,\mathrm{d}t.$$

Thus, $\overline{U}^{\pi}_x \leqslant \overline{U}^{\mathrm{PS}}_x$ for all $x \leqslant x^*$. In particular, we have

$$\overline{U}^{\pi}_{x^*} \leqslant \overline{U}^{\mathrm{PS}}_{x^*}.$$

On the other hand, by definition, $\overline{T}^{\pi}((x^*)^+) \geqslant \overline{T}^{\mathrm{PS}}(x^*)$. Together with Proposition 7 this implies that, for all $x > x^*$,

$$(\overline{U}^{\pi}_x)' = \lambda\overline{F}(x)\overline{T}^{\pi}(x) \geqslant \lambda\overline{F}(x)\overline{T}^{\mathrm{PS}}(x) = (\overline{U}^{\mathrm{PS}}_x)'.$$

Finally, since both π and PS are work conserving disciplines, for which the mean unfinished work is equal, we have

$$\overline{U}^{\pi}_\infty = \overline{U}^{\mathrm{PS}}_\infty.$$

These last three formulas together with the fact that $\overline{U}^{\pi}_x$ is a continuous function of x guarantee that $\overline{U}^{\pi}_x \leqslant \overline{U}^{\mathrm{PS}}_x$ for all $x > x^*$, which completes the proof. $\square$

Proposition 9. *Let* $\pi \in$ MLPS. *Then* $\overline{U}^{\pi}_x \leqslant \overline{U}^{\mathrm{PS}}_x$ *for all* $x \geqslant 0$.

Proof. Let $\pi' \in (N+1)$PS with the same thresholds as π. Since

$$\overline{U}^{\pi}_x = \lim_{t\to\infty} \frac{1}{t} \int_0^t U^{\pi}_x(s)\,\mathrm{d}s.$$

Proposition 2 implies that $\overline{U}^{\pi}_x \leqslant \overline{U}^{\pi'}_x$ for all $x \geqslant 0$. The claim follows now from Proposition 8. $\square$

Theorem 2. *Let* $\pi \in$ MLPS.

(i) *If the hazard rate* $h(x)$ *is decreasing, then* $\overline{T}^{\pi} \leqslant \overline{T}^{\mathrm{PS}}$.
(ii) *If the hazard rate* $h(x)$ *is increasing and bounded, then* $\overline{T}^{\pi} \geqslant \overline{T}^{\mathrm{PS}}$.

Proof. This follows immediately from Propositions 3 and 9. □

4. Conclusions

What still remains to be proved is the plausible claim that, roughly said, an MLPS discipline with the internal disciplines belonging to the set {FB, PS} is the better, the more levels there are. The key question here is the following "level splitting problem". Let $N \geqslant 1$ and $\pi \in (N+1)$PS with thresholds $\{a_1, \ldots, a_N\}$. Furthermore, let $n \in \{1, \ldots, N\}$ and $\pi' \in N$PS with thresholds $\{a_1, \ldots, a_{n-1}, a_{n+1}, \ldots, a_N\}$. Then prove that $\overline{U}_x^{\pi} \leqslant \overline{U}_x^{\pi'}$ for all $x \geqslant 0$.

References

[1] S. Aalto, U. Ayesta, E. Nyberg-Oksanen, Two-level processor-sharing scheduling disciplines: mean delay analysis, in: Proceedings of ACM SIGMETRICS/PERFORMANCE, New York, 2004, pp. 97–105.
[2] K. Avrachenkov, U. Ayesta, P. Brown, E. Nyberg, Differentiation between short and long TCP flows: predictability of the response time, in: Proceedings of IEEE Infocom, Hong Kong, 2004.
[3] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, J. Roberts, Statistical bandwidth sharing: a study of congestion at flow level, in: Proceedings of ACM SIGCOMM, San Diego, CA, 2001, pp. 111–122.
[4] M.E. Crovella, A. Bestavros, Self-similarity in World Wide Web traffic: evidence and possible causes, in: Proceedings of ACM SIGMETRICS, Philadelphia, PA, 1996, pp. 160–169.
[5] A. Feldmann, W. Whitt, Fitting mixtures of exponentials to long-tail distributions to analyze network performance models, in: Proceedings of IEEE Infocom, Kobe, Japan, 1997, pp. 1096–1104.
[6] H. Feng, V. Misra, Mixed scheduling disciplines for network flows, ACM SIGMETRICS Performance Evaluation Review, vol. 31, 2003, pp. 36–39.
[7] L. Guo, I. Matta, Differentiated control of web traffic: a numerical analysis, in: Proceedings of SPIE ITCOM'2002: Scalability and Traffic Control in IP Networks, Boston, MA, 2002.
[8] D.P. Heyman, T.V. Lakshman, A.L. Neidhardt, A new method for analysing feedback-based protocols with applications to engineering Web traffic over the Internet, in: Proceedings of ACM SIGMETRICS, Seattle, WA, 1997, pp. 24–38.
[9] Queueing Systems, L. Kleinrock, Volume II: Computer Applications, Wiley, New York, 1976.
[10] M. Nuijens, The foreground–background queue, Ph.D. Thesis, University of Amsterdam, 2004.
[11] I.A. Rai, G. Urvoy-Keller, M.K. Vernon, E.W. Biersack, Performance analysis of LAS-based scheduling disciplines in a packet switched network, in: Proceedings of ACM SIGMETRICS/PERFORMANCE, New York, 2004, pp. 106–117.
[12] R. Righter, J.G. Shanthikumar, Scheduling multiclass single server queueing systems to stochastically maximize the number of successful departures, Probab. Eng. Inform. Sci. 3 (1989) 323–333.
[13] R. Righter, J.G. Shanthikumar, G. Yamazaki, On extremal service disciplines in single-stage queueing systems, J. Appl. Probab. 27 (1990) 409–416.
[14] A. Wierman, N. Bansal, M. Harchol-Balter, A note on comparing response times in the M/GI/1/FB and M/GI/1/PS queues, Oper. Res. Lett. 32 (2004) 73–76.
[15] S.F. Yashkov, Processor-sharing queues: some progress in analysis, Queueing Systems 2 (1987) 1–17.