Batch arrival M/G/1 Processor Sharing with application to Multilevel Processor Sharing scheduling

Konstantin Avrachenkov, Urtzi Ayesta, Patrick Brown

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Batch arrival $M/G/1$ Processor Sharing with application to Multilevel Processor Sharing scheduling

Konstantin Avrachenkov — Urtzi Ayesta — Patrick Brown

N° 5043

Decembre 2003

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5043--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Batch arrival $M/G/1$ Processor Sharing with application to Multilevel Processor Sharing scheduling

Konstantin Avrachenkov*, Urtzi Ayesta† , Patrick Brown‡

Thème 1 — Réseaux et systèmes
Projet Maestro

Rapport de recherche n° 5043 — Decembre 2003 — 21 pages

Abstract: We analyze an $M/G/1$ Processor-Sharing queue with Batch arrivals. Our analysis is based on the integral equation derived by Kleinrock, Muntz and Rodemich. Using the contraction mapping principle, we demonstrate the existence and uniqueness of a solution to the integral equation. Then we provide asymptotical analysis as well as tight bounds for the expected response time conditioned on the job size. In particular, the asymptotics for large size jobs depends only on the first moment of the job size distribution and on the first two moments of the batch size distribution. That is, similarly to the Processor Sharing with single arrivals, in the $M/G/1-PS$ with batch arrivals the expected conditional response time is finite even when the job size distribution has infinite second moment. Finally, we show how the present results can be applied to the Multilevel Processor Sharing scheduling.

Key-words: Processor Sharing, Batch arrivals, Work conservation, Multilevel Processor Sharing

* INRIA Sophia Antipolis, k.avrachenkov@sophia.inria.fr
† INRIA Sophia Antipolis and France Telecom R&D, Urtzi.Ayesta@{sophia.inria.fr,francetelecom.com}
‡ France Telecom R&D, Patrick.Brown@francetelecom.com

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

$M/G/1$ "Processor Sharing" avec des arrivées en rafales avec application a la théorie d'ordonnancement "Multilevel Processor Sharing"

Résumé : Nous analysons une file d'attente $M/G/1$ Processor-Sharing avec des arrivées en rafales. Notre analyse se base sur l'équation intégrale de Kleinrock, Muntz et Rodemich. En utilisant le principe de contraction, nous démontrons l'existence et l'unicité de sa solution. Ensuite, nous fournissons une analyse asymptotique ainsi que des bornes serrées pour le temps de réponse moyen conditionné sur la quantité de travail requise. En particulier, l'asymptote pour des tâches très longues dépend seulement du premier moment de la distribution du temps de traitement et des deux premiers moments de la distribution de la taille des rafales. Pareillement à la file Processor Sharing avec des arrivées individuelles, dans la file $M/G/1$ Processor Sharing avec des arrivées groupées, le temps de réponse moyen demeure fini même lorsque le deuxième moment de la distribution du temps de traitement est infini. Finalement, nous montrons comment ces résultats peuvent être appliqués à l'analyse de politiques d'ordonnancement basées sur le service écoulé.

Mots-clés : Processor Sharing, arrivées en rafales, Loi de conservation, Multilevel Processor Sharing

1 Introduction and Motivation

The $M/G/1-PS$ queue with batch arrivals ($M/G/1-BPS$) has not been fully characterized yet. Kleinrock et al. [1, 2] showed that the derivative of the expected response time conditioned on the job size satisfies an integral equation. Furthermore, they obtained an analytic solution for the job size distributions of the type $\overline{F}(x) = q(\tau)e^{-\mu x}$ where $q(\tau)$ is polynomial. Bansal [3], using the Kleinrock's integral equation, obtained the Laplace transform of the expected conditional response time for hyperexponential distributions and more generally for distributions with rational Laplace Transforms. More recently, Feng and Misra [4] provided bounds for the expected conditional response time. Their bounds depend on the second moment of the job size distribution. Rege and Sengupta [5] found the distribution of the expected conditional response time for a tagged customer, given the service times of all customer in the system.

One of the main motivations to study the $M/G/1-BPS$ queue is its application to attained service based scheduling. Attained service based scheduling has recently received a fairly big attention in connection with the differentiation of Short and Long flows in the Internet [6, 7, 8, 4, 9]. Kleinrock et al. [1, 10, 2] introduced a quite general set of size-based scheduling termed as Multilevel Processor Sharing ($MLPS$). In $MLPS$, jobs are served with a discipline that will depend on their attained amount of service. That is, based on their attained service, jobs are classified into different classes. Jobs within the same class are served either with $FIFO$ or PS or Least Attained Service (LAS) policy. The classes themselves are served according to the LAS policy, that is, the class that contains jobs with smallest attained service is served first. It turns out, that when PS is used to serve jobs in any of the classes, the expected conditional response time in this class can be expressed as a function of the expected conditional response in an $M/G/1-BPS$.

The organization of the paper is as follows: First we prove the existence and uniqueness of a solution to the Kleinrock's integral equation. Second we show that under natural conditions, the expected conditional response time has an asymptote and we give an analytical expression for the slope and the bias of the asymptote. In particular, this asymptote provides a tight upper bound for the expected conditional response time for large jobs. Yet another upper bound is obtained for small jobs. Combining these two bounds we obtain a very good characterization of the expected conditional response time for all job sizes. In particular these bounds are insensitive to the job size distribution and depend only on the distribution through the first moment. Finally, as an example of the application of these results, we show that in the case of $MLPS$ schedulers, the expected conditional response time has an asymptote.

2 Analysis of the Batch $M/G/1-PS$ queue

2.1 Model and Notation

Let us denote $T_{BPS}(x)$ the conditional response time for a job of size x in an $M/G/1-PS$ system with batch arrivals. Let $T'_{BPS}(x)$ be its derivative. Kleinrock et al. [1, 2] has shown that $T'_{BPS}(x)$ is a solution of the following integral equation

$$\begin{aligned} T'_{BPS}(x) &= \lambda E[N]\int_0^\infty T'_{BPS}(y)\overline{F}(x+y)dy \\ &+\lambda E[N]\int_0^x T'_{BPS}(y)\overline{F}(x-y)dy \\ &+b\overline{F}(x)+1, \end{aligned} \tag{1}$$

where λ is the batch arrival rate, $E[N]$ is the average batch size, b is the average number of jobs that arrive in addition to the tagged job $b+1=E[N^2]/E[N]$ and $\overline{F}(x)=1-F(x)$ is the complementary distribution function. The load in the BPS system is given by $\rho = \lambda E[N]E[X]$.

2.2 Fixed Point Approach to the Kleinrock's integral equation

Theorem 1 shows that there exists a unique solution to the integral equation (1).

Theorem 1 *Let the service time distribution have finite mean, the batch size distribution have a finite second moment and $\rho < 1$. Then there exists a unique solution of the integral equation (1).*

Proof 1 *We consider the fixed point iterations*

$$\begin{aligned} T'_{k+1}(x) &= \lambda E[N]\int_0^\infty T'_k(y)\overline{F}(x+y)dy \\ &+\lambda E[N]\int_0^x T'_k(y)\overline{F}(x-y)dy \\ &+b\overline{F}(x)+1, \quad k=0,1,... \end{aligned} \tag{2}$$

on the complete functional space of continuous bounded non-negative functions $\mathcal{C}[0,\infty)$ with the supremum metric. Let $\|T'\| = \sup_x\{T'(x)\} < \infty$. Define the linear integral operator $\mathcal{A}[\beta(x)]$ as follows:

$$\begin{aligned} \mathcal{A}[\beta(x)] &= \lambda E[N]\int_0^\infty \beta(y)\overline{F}(x+y)dy \\ &+\lambda E[N]\int_0^x \beta(y)\overline{F}(x-y)dy + b\overline{F}(x)+1 \end{aligned}$$

Clearly the operator $\mathcal{A}[\beta(x)]$ maps the space $\mathcal{C}[0,\infty)$ into itself.

If we show that the linear integral operator $\mathcal{A}[\beta(x)]$ is a contraction, then the integral equation (3) has a unique solution in $\mathcal{C}[0,\infty)$. Let us denote as d the distance in the metric space $\mathcal{C}[0,\infty)$, that is, $d(\beta_1,\beta_2)=\sup_x|\beta_1(x)-\beta_2(x)|$. We show now that the linear operator $\mathcal{A}[\beta(x)]$ is indeed a contraction mapping on $\mathcal{C}[0,\infty)$.

$$\begin{aligned} d(\mathcal{A}[\beta_1],\mathcal{A}[\beta_2]) &= \sup_x\{|\mathcal{A}[\beta_1]-\mathcal{A}[\beta_2]|\} \\ &\leq \lambda E[N]\sup_x\{|\mathcal{A}[\beta_1]-\mathcal{A}[\beta_2]|\} \\ &\quad \sup_x\left(\int_0^\infty \overline{F}(x+y)dy + \int_0^x \overline{F}(x-y)dy\right) \\ &= \lambda E[N] d(\beta_1,\beta_2)E[X] \\ &= \rho d(\beta_1,\beta_2). \end{aligned}$$

Thus, the mapping is a contraction if $\rho<1$.

Theorem 1 implies that we can apply the Fixed Point Iterations (2) for the solution of the integral equation (1). A numerical example will be provided in Section 3.

For the ensuing analysis, it will be convenient to remove the constant component of the solution of equation (1), hence we note that the solution of the integral equation (1) is equivalent to the solution of the following integral equation

$$\begin{aligned} \delta T'(x) &= \lambda E[N]\int_0^\infty \delta T'(y)\overline{F}(x+y)dy \\ &\quad +\lambda E[N]\int_0^x \delta T'(y)\overline{F}(x-y)dy \\ &\quad +b\overline{F}(x), \end{aligned} \tag{3}$$

where

$$\delta T'(x) := T'_{BPS}(x) - \frac{1}{1-\rho}. \tag{4}$$

2.3 Asymptotic Analysis

It is known that in a queue, under any work conserving discipline, the total unfinished work in the system does not depend on the particular scheduling policy being used. This fact has been widely exploited since it poses a constraint on the average conditional response time $T(x)$ of the system among the scheduling disciplines that are service time independent, i.e., disciplines that do not take advantage of the length of the jobs when choosing which job to serve first.

Lemma 1 *[11, 12] In an ergodic queue, under any work conserving and service-time independent scheduling discipline, the expected conditional response time satisfies*

$$\lambda \int_0^\infty T(x)\overline{F}(x)dx = \overline{W} \tag{5}$$

where λ *is the job arrival rate and* $\overline{W}$ *is the time-average unfinished work in the system.*

The interest of Lemma 1 lies in the fact that if the expected conditional response time is known for a particular scheduling discipline, then one can compute the average unfinished work in the system $\overline{W}$. Since this quantity is independent of the scheduling discipline, the expected conditional time for any other scheduling discipline have to satisfy equation (5).

Let $\overline{W}^B$ be the expected unfinished work in the case of Poisson batch arrival queue. In order to apply Lemma 1 to the the Poisson batch arrival system, we note that the job arrival rate is $\lambda E[N]$, thus

$$\lambda E[N] \int_0^\infty T(x)\overline{F}(x)dx = \overline{W}^B. \tag{6}$$

The expected unfinished work $\overline{W}^B$ in a Poisson batch arrival queue can be easily computed [13, 14].The basic step is to consider a $FIFO$ discipline and to define the random variable $Y = \sum_i^N X_i$, where N is the size of the batch and X_i is the size of the i-th job. Then the expected unfinished work can be computed directly by the Pollaczek-Khinchin formula. The expressions given in [13, 14] become more transparent if they are expressed as a function of b, namely,

$$\overline{W}^B = \overline{W}^{Batch-FIFO} = \frac{\lambda E[Y^2]}{2(1-\lambda E[Y])} = \frac{\lambda E[N]E[X^2]}{2(1-\rho)} + \frac{bE[X]\rho}{2(1-\rho)}. \tag{7}$$

We illustrate Lemma 1 with two particular examples.

Example 1 *We consider an* $FIFO$ *queue with Poisson batch arrivals. The expected response time of a job is the sum of its own service requirements, the service time of jobs of the same batch that are ahead of him in the batch and the amount of unfinished work he finds in the system upon arrival. A more detailed analysis can be found in [13, 15]. Expressing their result as a function of* b *we obtain* $T_{Batch-FIFO}(x) = x + \frac{b}{2}E[X] + \overline{W}^B$.

$$\begin{aligned}
\lambda E[N] \int_0^\infty T_{Batch-FIFO}(x)\overline{F}(x)dx &= \lambda E[N] \int_0^\infty \left(x + \frac{bE[X]}{2} + \overline{W}^B\right)\overline{F}(x)dx \\
&= \lambda E[N]\frac{E[X^2]}{2} + \frac{bE[X]}{2}\rho + \overline{W}^B\rho \\
&= (1-\rho)\overline{W}^B + \rho\overline{W}^B = \overline{W}^B
\end{aligned}$$

Example 2 *Let us consider now a BPS queue with exponentially distributed file sizes. This is the only distribution for which there exists an analytical expression for the expected conditional response time $T_{BPS_{exp}}$. Then it is known that [1, 2, 5]*

$$T_{BPS_{exp}} = \frac{x}{1-\rho} + \frac{b(2-\rho)E[X]}{2(1-\rho)^2}(1 - e^{\frac{-(1-\rho)}{E[X]}x}). \tag{8}$$

Then, proceeding similarly as in the previous example we obtain

$$\begin{aligned}
\lambda E[N]\int_0^\infty T_{BPS_{exp}}(x)\overline{F}(x)dx &= \frac{\lambda E[N]E[X^2]}{2(1-\rho)} + \lambda E[N]\frac{b(2-\rho)E[X]}{2(1-\rho)^2}E[X] \\
&\quad -\lambda E[N]\frac{b(2-\rho)E[X]}{2(1-\rho)^2}\int_0^\infty e^{\frac{-(2-\rho)}{E[X]}x}dx \\
&= \frac{\lambda E[N]E[X^2]}{2(1-\rho)} + \frac{b(2-\rho)\rho E[X]}{2(1-\rho)^2}\rho - \frac{b(2-\rho)\rho}{2(1-\rho)^2}\frac{E[X]}{(2-\rho)} \\
&= \frac{\lambda E[N]E[X^2]}{2(1-\rho)} + \frac{b\rho E[X](2-\rho-1)}{2(1-\rho)^2} = \overline{W}^B
\end{aligned}$$

In the following Lemma, we take advantage of equation (6) to obtain a result that is crucial for the ensuing analysis.

Lemma 2 *Let $\delta T(x) = T_{BPS}(x) - \frac{x}{1-\rho}$, and let the service time distribution have a finite mean, the batch size distribution have a finite second moment and $\rho < 1$. Then it holds that*

$$\lambda E[N]\int_0^\infty \delta T(x)\overline{F}(x)dx = \frac{bE[X]\rho}{2(1-\rho)}. \tag{9}$$

Proof 2 *Let X be a random variable with complementary distribution function $\overline{F}(x)$ and density function $f(x)$. The second moment of X is allowed to be infinite. We consider the truncated random variable X_t at t, that is $X_t = \min\{X, t\}$. The complementary distribution of the truncated random variable is*

$$\overline{F}_t(x) = \begin{cases} \overline{F}(x), & x \le t, \\ 0, & otherwise. \end{cases}$$

The mean of the truncated random variable is given by

$$E[X_t] = \int_0^\infty \overline{F}_t(y)dy = \int_0^t \overline{F}(y)dy$$

We note that the second moment of the truncated random variable is always finite.

$$E[X_t^2] = \int_0^t 2y\overline{F}(y)dy < \infty$$

Let $T^t_{BPS}(x)$ be the expected conditional response time in a BPS queue in the case when jobs are distributed according to the random variable X_t and $\rho_t = \lambda E[N]E[X_t]$. Then from equation 7 and Lemma 1 we have

$$\begin{aligned}
\frac{\lambda E[N]E[X_t^2]}{2(1-\rho_t)} + \frac{\rho E[X_t]b}{2(1-\rho_t)} &= \lambda E[N]\int_0^\infty T^t_{BPS}(x)\overline{F_t}(x)dx \\
&= \lambda E[N]\int_0^\infty \left(\frac{x}{1-\rho_t} + \delta T_t(x)\right)\overline{F_t}(x)dx \\
&= \frac{\lambda E[N]E[X_t^2]}{2(1-\rho_t)} + \lambda E[N]\int_0^\infty \delta T_t(x)\overline{F_t}(x)dx
\end{aligned}$$

Consequently, we have that $\lambda E[N]\int_0^\infty \delta T_t(x)\overline{F_t}(x)dx = \frac{\rho E[X_t]b}{2(1-\rho_t)}$. Taking the limit when $t \to \infty$ we obtain

$$\lambda E[N]\int_0^\infty \delta T(x)\overline{F}(x)dx = \frac{bE[X]\rho}{2(1-\rho)}$$

Let us prove now another Lemma.

Lemma 3 *Let the service time distribution have a finite mean, the batch size distribution have a finite second moment and $\rho < 1$. Then, $\delta T(x) = T_{BPS}(x) - \frac{x}{1-\rho}$ is increasing with respect to x.*

Proof 3 *Let us show that $\delta T'(x) = T'_{BPS}(x) - \frac{1}{1-\rho} \geq 0$.*

$$\begin{aligned}
\inf_{x\geq 0}\{\delta T'(x)\} &= \inf_{x\geq 0}\{\lambda E[N]\int_0^\infty \delta T'(y)\overline{F}(x+y)dy + \lambda E[N]\int_0^x \delta T'(y)\overline{F}(x-y)dy + b\overline{F}(x)\} \\
&\geq \lambda E[N]\inf_{y\geq 0}\{\delta T'(y)\}\left(\int_0^\infty \overline{F}(x+y)dy + \int_0^x \overline{F}(x-y)dy\right) \\
&= \lambda E[N]\inf_{y\geq 0}\{\delta T'(y)\}E[X] \\
&= \rho\inf\{\delta T'(x)\}
\end{aligned}$$

Hence $\delta T'(x) \geq 0$ and in particular $T'_{BPS}(x) - \frac{1}{1-\rho} \geq 0$.

A direct consequence of Lemma 3 is that $\delta T(x) \geq 0\ \forall x \geq 0$. Next we obtain an upper bound for the expected conditional response time.

Lemma 4 *Let the service time distribution have a finite mean, the batch size distribution have a finite second moment and $\rho < 1$. The upper bound for the expected conditional response time is given by*

$$T_{BPS}(z) \leq \frac{z}{1-\rho} + \frac{b\,(\rho E[X] + 2E[X_z](1-\rho))}{2(1-\rho)(1-\rho_z)}$$

Proof 4 *Let us consider the first term on the right hand-side of equation (3). Integrating by parts we have*

$$\begin{aligned}\lambda E[N]\int_0^\infty \delta T'(y)\overline{F}(x+y)dy &= \lambda E[N]\left(\delta T(y)\overline{F}(x+y)\right)|_{y=0}^{y=\infty} \\ &+\lambda E[N]\int_0^\infty \delta T(y)f(x+y)dy\end{aligned}$$

Noting that $\delta T(0)=0$ *in a Processor Sharing system and that as a consequence of Lemma 2* $\lim_{y\to\infty}\delta T(y)\overline{F}(x+y)=0$ *we have*

$$\lambda E[N]\int_0^\infty \delta T'(y)\overline{F}(x+y)dy = \lambda E[N]\int_0^\infty \delta T(y)f(x+y)dy$$

Using the integral equation (3) and the fact that $\delta T(z)=\int_0^z \delta T'(x)dx$*, we can write*

$$\begin{aligned}\int_0^z \delta T'(x)dx &= \lambda E[N]\int_0^z\int_0^\infty \delta T'(y)\overline{F}(x+y)dydx \\ &+\lambda E[N]\int_0^z\int_0^x \delta T'(y)\overline{F}(x-y)dydx + b\int_0^z \overline{F}(x)dx \\ &= \lambda E[N]\int_0^z\int_0^\infty \delta T(y)f(x+y)dydx \\ &+\lambda E[N]\int_0^z \delta T'(y)\int_y^z \overline{F}(x-y)dxdy + bE[X_z] \\ &= \lambda E[N]\int_0^\infty \delta T(y)\int_0^z f(x+y)dxdy \\ &+\lambda E[N]\int_0^z \delta T'(y)\int_0^{z-y}\overline{F}(h)dhdy + bE[X_z] \\ &= \lambda E[N]\int_0^\infty \delta T(y)\left(\overline{F}(y)-\overline{F}(y+z)\right)dx \\ &+\lambda E[N]\int_0^z \delta T'(y)\int_0^{z-y}\overline{F}(h)dhdy + bE[X_z] \\ &\quad \textit{Next by Lemma 3 it follows that} \\ &\le \lambda E[N]\int_0^\infty \delta T(y)\overline{F}(y)dy \\ &+\lambda E[N]E[X_z]\int_0^z \delta T'(y)dy + bE[X_z] \\ (1-\lambda E[N]E[X_z])\delta T(z) &\le \lambda E[N]\int_0^\infty \delta T(y)\overline{F}(y)dy + bE[X_z].\end{aligned}$$

Substituting the result obtained in Lemma 2 and taking into account that $\rho_z < 1$, we get

$$\delta T(z) = \int_0^z \delta T'(x)dx \le \frac{b\left(\rho E[X] + 2E[X_z](1-\rho)\right)}{2(1-\rho)(1-\rho_z)}.$$

Consequently, we obtain the upper bound for $T_{BPS}(z)$

$$\begin{aligned} T_{BPS}(z) &= \frac{z}{1-\rho} + \delta T(z) \\ &\le \frac{z}{1-\rho} + \frac{b\left(\rho E[X] + 2E[X_z](1-\rho)\right)}{2(1-\rho)(1-\rho_z)}. \end{aligned}$$

In the next Theorem we state the main result of this paper. Namely we show that $T_{BPS}(x)$ has an asymptote. This result will be useful afterwards to provide tight upper bounds on the expected conditional and unconditional response times.

Theorem 2 *Let the service time distribution have a finite mean, the batch size distribution have a finite second moment and $\rho < 1$. The conditional response time for the BPS queue has an asymptote with slope $1/(1-\rho)$ and bias*

$$\lim_{x\to\infty}\left(T_{BPS}(x) - \frac{x}{1-\rho}\right) = \frac{bE[X](2-\rho)}{2(1-\rho)^2}.$$

Proof 5 *Let us show that there exists an asymptote. From Lemma 4 we know that $T_{BPS}(x) - \frac{x}{1-\rho}$ is upper bounded and from Lemma 3 that $T_{BPS}(x) - \frac{x}{1-\rho}$ is increasing with respect to x. Consequently $\lim_{x\to\infty} T_{BPS}(x) - \frac{x}{1-\rho}$ exists. This justifies the following calculation of the asymptote bias. Proceeding in a similar way as in the proof of Lemma 4, we can write*

$$\begin{aligned} \lim_{x\to\infty}\left(T_{BPS}(x) - \frac{x}{1-\rho}\right) &= \int_0^\infty \delta T'(x)dx = \\ &= \lambda E[N]\int_0^\infty\int_0^\infty \delta T'(y)\overline{F}(x+y)dydx \\ &\quad +\lambda E[N]\int_0^\infty\int_0^x \delta T'(y)\overline{F}(x-y)dydx + b\int_0^\infty \overline{F}(x)dx \\ &= \lambda E[N]\int_0^\infty\int_0^\infty \delta T(y)f(x+y)dydx \\ &\quad +\lambda E[N]\int_0^\infty \delta T'(y)\int_y^z \overline{F}(x-y)dxdy + bE[X] \\ &= \lambda E[N]\int_0^\infty \delta T(y)\int_0^\infty f(x+y)dxdy \\ &\quad +\lambda E[N]\int_0^\infty \delta T'(y)\int_0^\infty \overline{F}(h)dhdy + bE[X] \end{aligned}$$

$$
\begin{aligned}
&= \lambda E[N]\int_0^\infty \delta T(y)\overline{F}(y)dy \\
&\quad +\lambda E[N]E[X]\int_0^\infty \delta T'(y)dy + bE[X] \\
&= \frac{bE[X]\rho}{2(1-\rho)} + \lambda E[N]E[X]\int_0^\infty \delta T'(y)dy + bE[X]
\end{aligned}
$$

Solving the equation for $\int_0^\infty \delta T'(y)dy$, *we obtain*

$$\lim_{x\to\infty}\left(T_{BPS}(x) - \frac{x}{1-\rho}\right) = \frac{bE[X](2-\rho)}{2(1-\rho)^2}$$

Interestingly, we observe that the value of the bias is insensitive with respect to the job size distribution, that is, it depends on the distribution only through the first moment.

Corollary 1 *Let the service time distribution have a finite mean, the batch size distribution have a finite second moment and* $\rho < 1$, *then the slowdown of* $M/G/1-BPS$ *is given by*

$$\lim_{x\to\infty}\frac{T_{BPS}(x)}{x} = \frac{1}{1-\rho}$$

Proof 6 *The result is a direct consequence of Theorem 2.*

Corollary 1 shows that in the BPS queue, very large jobs obtain service at the same rate they would in the equivalent PS queue, that is $\lim_{x\to\infty}\frac{T_{BPS}(x)}{x} = \frac{T_{PS}(x)}{x}$.

2.4 Bounds

In this section, we use the results obtained in the preceding section to obtain tight upper bounds for the expected conditional response time as well as for the average unconditional response time. We start by providing upper and lower bounds for the expected conditional response time.

Theorem 3 *The lower and upper bounds for the expected conditional response time in the* BPS *queue are given by:*

$$\frac{x}{1-\rho} \le T_{BPS}(x) \le \min\{\frac{b+1}{1-\rho}x, \frac{x}{1-\rho} + \frac{bE[X](2-\rho)}{2(1-\rho)^2}\}.$$

The bounds on the right hand part of the inequality intersect at the point $x^* = \frac{E[X](2-\rho)}{2(1-\rho)}$.

Proof 7 *Since* $T'_{BPS}(x) - \frac{1}{1-\rho} \ge 0$, $T_{BPS}(x)$ *approaches the asymptotic from below. Hence for large job sizes we obtain a bound that is asymptotically tight. Thus, from Lemma 2 we have:*

$$T_{BPS}(z) \le \frac{z}{1-\rho} + \frac{bE[X](2-\rho)}{2(1-\rho)^2} \tag{10}$$

Clearly, the upper bound (10) is not appropriate for small job sizes, since we know that $T_{BPS}(0) = 0$.

Thus, for small values of x, *the function* $T_{BPS}(x)$ *can be approximated by calculating an upper bound of its derivative. Let us estimate* $\sup_{x\geq 0}\{\delta T'_{BPS}(x)\}$

$$\begin{aligned}
\sup_{x\geq 0}\{\delta T'_{BPS}(x)\} &= \sup_{x\geq 0}\{\lambda E[N]\int_0^\infty \delta T'_{BPS}(y)\overline{F}(x+y)dy \\
&\quad +\lambda E[N]\int_0^x \delta T'_{BPS}(y)\overline{F}(x-y)dy + b\overline{F}(x) + 1\} \\
&\leq \lambda E[N]\sup_{x\geq 0}\{\delta T'_{BPS}(x)\}\left(\int_0^\infty \overline{F}(x+y)dy + \int_0^x \overline{F}(x-y)dy\right) + b + 1 \\
&= \lambda E[N]\sup_{x\geq 0}\{\delta T'_{BPS}(x)\}\int_0^\infty \overline{F}(z)dz + b + 1 \\
&= \lambda E[N]\sup_{x\geq 0}\{\delta T'_{BPS}(x)\}E[X] + b + 1 = \rho\sup_{x\geq 0}\{\delta T'_{BPS}(x)\} + b + 1
\end{aligned}$$

and hence $\sup\{\delta T'_{BPS}(x)\} \leq \frac{b+1}{1-\rho}$. *Noting that* $T_{BPS}(0) = 0$ *and integrating between* 0 *and* x *we obtain another upper bound, that is,*

$$T_{BPS}(x) \leq \frac{1+b}{1-\rho}x \tag{11}$$

Equating the bounds (10) and (11), we find the intersection point

$$\frac{x}{1-\rho} + \frac{bE[X](2-\rho)}{2(1-\rho)^2} = \frac{1+b}{1-\rho}x \Longrightarrow x^* = \frac{E[X](2-\rho)}{2(1-\rho)}. \tag{12}$$

Thus, we have

$$T_{BPS}(x) \leq \min\{\frac{b+1}{1-\rho}x, \frac{x}{1-\rho} + \frac{bE[X](2-\rho)}{2(1-\rho)^2}\}.$$

The lower bound is a direct consequence of the inequality $T'_{BPS}(x) - \frac{1}{1-\rho} \geq 0$.

In Section 3 we show that the upper bound characterizes quite closely the expected conditional response time for large jobs in the BPS queue. In the next theorem, we use the upper bound of the expected conditional response time to provide an upper bound on the unconditional response time.

Theorem 4 *The lower and upper bounds for the expected unconditional response time in a* BPS *queue are given by:*

$$\frac{E[X]}{1-\rho} \leq E[T_{BPS}] \leq \frac{E[X]}{1-\rho} + \frac{b}{1-\rho}E[X_{x^*}].$$

where x^* *is the same as in Theorems 3.*

Proof 8 *The lower bound is straightforward from the lower bound in Theorem 3. Now we calculate the upper bound.*

$$
\begin{aligned}
E[T_{BPS}] &= \int_0^\infty T_{BPS}(x)f(x)dx \\
&\le \int_0^\infty \frac{x}{1-\rho}f(x)dx + \int_0^{x^*} \frac{bx}{1-\rho}f(x)dx \\
&\quad + \int_{x^*}^\infty \frac{bE[X](2-\rho)}{2(1-\rho)^2}f(x)dx \\
&= \frac{E[X]}{1-\rho} + \frac{b}{1-\rho}E[X_{x^*}] - \frac{b}{1-\rho}\frac{E[X](2-\rho)}{2(1-\rho)}\overline{F}(x^*) \\
&\quad + \frac{bE[X](2-\rho)}{2(1-\rho)^2}\overline{F}(x^*) \\
&= \frac{E[X]}{1-\rho} + \frac{b}{1-\rho}E[X_{x^*}]
\end{aligned}
$$

2.5 Reducing the upper bound iteratively

We observe that combining the results of Theorem 3 and 4 with the fact that

$$
\begin{aligned}
T'_{BPS}(0) &= \lambda E[N]\int_0^\infty T'_{BPS}(y)\overline{F}(y)dy + b + 1 \\
&= \lambda E[N]\int_0^\infty T_{BPS}(y)f(y)dy + b + 1 = \lambda E[N]E[T_{BPS}] + b + 1, \qquad (13)
\end{aligned}
$$

it is possible to perform successive iterations to lower the upper bound on $E[T_{BPS}]$. Note that $T'_{BPS}(0) \ge T'_{BPS}(x)\ \forall x$. If we plug the upper bound obtained in Theorem 4 in (13) , we get $T'_{BPS}(0) \le \frac{1+b}{1-\rho} - \frac{b(\rho-\rho_{x^*})}{1-\rho}$. Hence, we can lower the upper bound, that is,

$$T_{BPS}(x) \le \left(\frac{1+b}{1-\rho} - \frac{b(\rho-\rho_{x^*})}{1-\rho}\right)x,$$

which is clearly more accurate than the previous one.

Then, in the spirit of Theorems 3 and 4, we can write

$$T_{BPS}(x) \le \min\{\left(\frac{1+b}{1-\rho} - \frac{b(\rho-\rho_{x^*})}{1-\rho}\right)x, \frac{x}{1-\rho} + \frac{bE[X](2-\rho)}{2(1-\rho)^2}\}$$

with intersection at $x_2^* = \frac{E[X](2-\rho)}{2(1-\rho)(1-\rho+\rho_{x^*})}$, and also

Similarly,

$$E[T_{BPS}] \le \frac{E[X]}{1-\rho} + \frac{bE[X_{x_2^*}]}{1-\rho} - \frac{bE[X](2-\rho)(\rho-\rho_{x^*})\overline{F}(x_2^*)}{2(1-\rho)^2(1+\rho_{x^*}-\rho)}.$$

3 Numerical examples

In this section we provide some numerical examples of the results of the preceding sections. In order to compute numerically $T_{BPS}(x)$ (or its derivative) for a general distribution we perform the Fixed Point Iterations (2). Indeed, we have shown in Theorem 1 that the Fixed Point Iterations (2) will converge to the solution of equation (1).

First of all, we consider the case of exponentially distributed file sizes and we demonstrate a high speed of convergence of the Fixed Point Iterations. Taking the derivative of the expected conditional response time for the exponential distribution case (see equation (8)), we obtain

$$\frac{dT_{BPS_{exp}}(x)}{dx} = \frac{1}{1-\rho} + \frac{b(2-\rho)}{2(1-\rho)} e^{\frac{-(1-\rho)}{E[X]}x}. \tag{14}$$

In Figure 1 we depict equation (14) and the Fixed Point Iterations (1st, 6th and 11th) of equation (2). We take $E[N] = 2$, $b = 5$, $E[X] = 20$ and $\rho = 0.7$. We note that the Fixed Point Iterations converge very rapidly to the analytic solution.

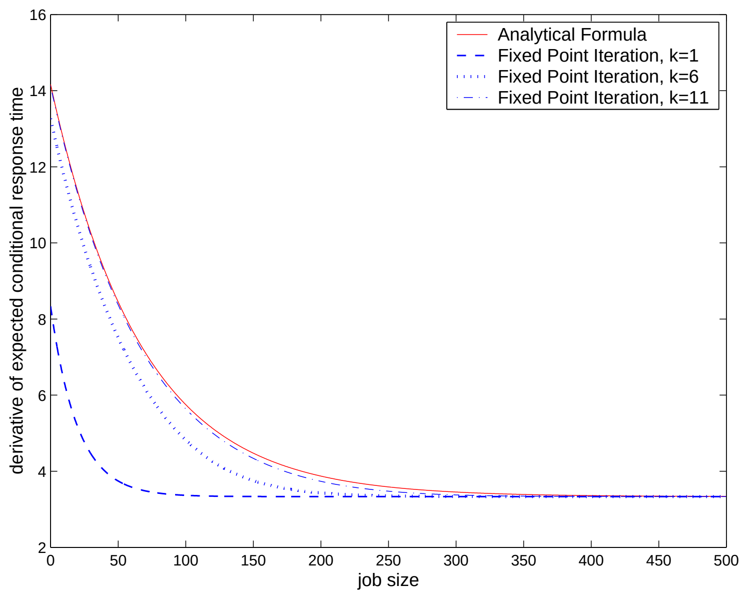

Figure 1: Converge of $T'_{BPS}(x)$ for exponential distribution: Analytical formula (14) and Fixed Point Iterations of equation (2)

In Figure 2 we plot the value of $T_{BPS}(x)$ obtained by Fixed Point Iterations for the case of Pareto distribution with infinite variance $(1 < \alpha < 2)$. We also plot the upper bound for the conditional response time of Theorem 3. The Pareto distribution is $F(x) = 1 - \frac{k^\alpha}{x^\alpha}$ and the parameters are $k = 10$, $\alpha = 1.5$, $E[N] = 3$, $b = 10$ and $\rho = \{0.3, 0.7\}$.

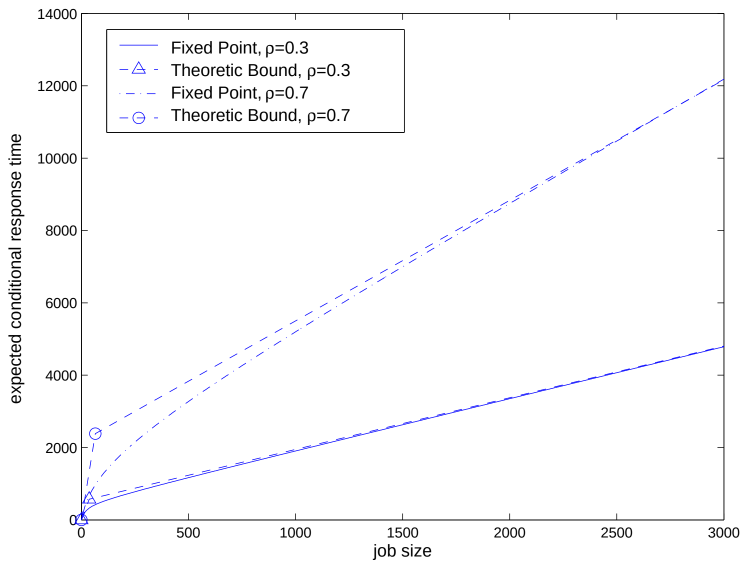

Figure 2: T_{BPS} for Pareto distribution: Fixed Point iterations and upper bound of Theorem 3

We consider now the upper bound for the unconditional expected response time obtained in Theorem 4. In the case of a general distribution, we can compare the upper bound of Theorem 4 with the numerical value of the expected unconditional response time obtained by the Fixed Point Iterations. In Figure 3, we consider a Pareto distribution, and we plot the upper bound provided in Theorem 4 and the numerical calculation of the expected unconditional response time for different loads. As in the previous numerical example with Pareto distribution, we take $k = 10$, $\alpha = 1.5$, $E[N] = 3$ and $b = 10$.

The tightness of the upper bound provided in Theorem 4 depends on the characteristics of the job size distribution. After performing extensive numerical analysis, we conclude that Theorem 4 provides a quite tight upper bound approximately up to load 0.6.

We note that it can be power consuming to calculate accurately the value $T_{BPS}(x)$ and $E[T_{BPS}]$ for distributions with infinite second moment, for example, Pareto if $1 < \alpha < 2$. In this case, we emphasize that the bounds provided in Theorems 3 and 4 are useful to characterize in a simple way and with good accuracy the performance of the BPS queue when the job size has a heavy-tail distribution.

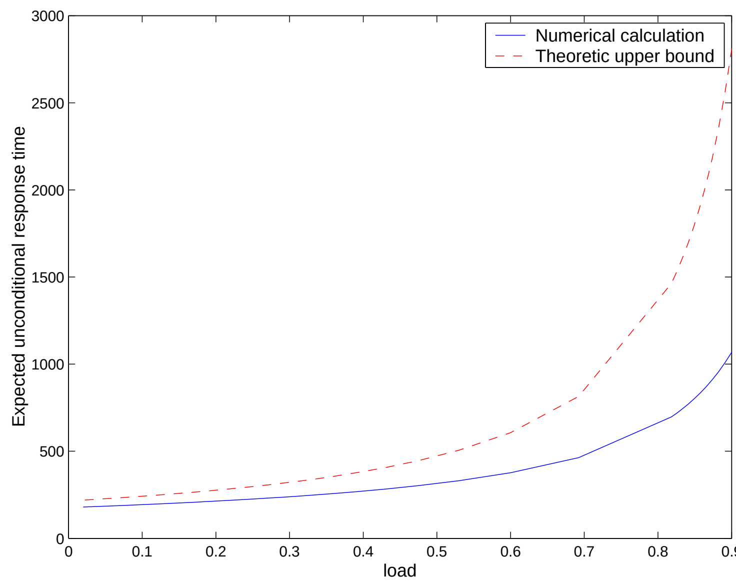

Figure 3: Expected unconditional response time for Pareto file size distribution: Fixed Point iterations and Theorem 4

4 Application to Multilevel Processor Sharing

One of the classical results of queuing theory says that when information on the size of the jobs is available for the server, the *Shortest Remaining Processing Time SRPT* scheduling discipline is optimal with respect to the expected unconditional response time of the system [16]. In some scenarios, this information is not available to the server, for instance, in computer networks the file size is not known in advance. Hence, scheduling disciplines that only take advantage of the attained service of jobs have drawn significant attention recently. The performance of the attained service based scheduling depends on the characteristics of the distribution function. For instance, it is known that when the distribution of the job size has a decreasing hazard rate $\mu(x) = f(x)/\overline{F}(x)$, the *Least Attained Service LAS* scheduling minimizes the expected unconditional response time among all disciplines that do not know the job's total size [17, 18] . In few words, we can describe *LAS* as the scheduler who gives full service to the job who has obtained the least amount of service.

It is clear that choosing an appropriate scheduling policy may significantly improve the performance of the system. In the current TCP/IP architecture of the Internet the length of a flow is not known in advance. This, coupled with the fact that no job obtains preferential treatment, have led researchers to propose *Processor Sharing PS* as a good mathematical abstraction for the bandwidth allocation that the network provides [19].

It has been widely reported that whereas most of the connections are made up of few packets, most of the data is carried by some few large connections [20]. This type of distributions (pareto, hyperexponential) have decreasing hazard rate. Therefore, from the theoretic point of view, it seems that giving priority to short flows might improve the overall performance of the system.

Even though the apparent desirable properties of LAS, its deployment does not seem to be a simple task. Hence, researchers have recently analyzed and proposed different size-based scheduling disciplines that aim to improve the performance that the current network architecture provides [6, 7, 4, 8, 9]. In the next section, we describe these schedulers as particular examples of the set of *Multilevel Processor-Sharing Scheduling* $MLPS$ disciplines introduced by Kleinrock [1] and we show how the results presented here for the BPS queue can be applied.

4.1 Multilevel Processor Sharing Scheduling $MLPS$

The framework of $MLPS$ allows us to define a very large class of scheduling disciplines. Let a_i be a set of numbers such that

$$0 = a_0 < a_1 < \ldots < a_N < a_{N+1} = \infty$$

We consider $N+1$ scheduling disciplines, where D_i is the discipline which is used to serve jobs whose attained service τ belongs to level i,that is $a_{i-1} < \tau \leq a_i$. We permit D_i to be either LAS or PS. Intervals are served according to a LAS discipline with respect to each other, that is, at any instant of time, the processor will give full service to the jobs belonging to the lowest nonempty level. For instance, let us consider a two level $MLPS$ scheduler with threshold at a, jobs with attained service smaller than a are served with PS and jobs that have attained more service than a with LAS. If there are in the system jobs who have attained less service than a, those jobs receive full service and they will be served according to a PS discipline. When there are no such jobs, the $MLPS$ scheduler will give full service to those jobs who have attained more service than a, in this case following a LAS scheduler. As soon as there is a new arrival, the server will interrupt serving jobs with attained service greater than a and start serving a new arrival.

Let x be some value belonging to the i-th interval, i.e. $a_{i-1} < x \leq a_i$ and $T^{MLPS}(x)$ the expected conditional response time. An important characteristic of $MLPS$ disciplines is that $T^{MLPS}(x)$ is independent of the scheduling discipline utilized in intervals $j \neq i$ [1].

Kleinrock showed that when PS is used in the i-th interval, this interval behaves as an $M/G/1-PS$ queue with batch arrivals (BPS) and interruptions due to arrivals of higher priority. When LAS is used, the i-th interval is equivalent to a LAS scheduler.

4.2 Truncated and Residual random variables

Let us introduce notations that will be used in the next section. Given a random variable X that takes values in $[0, \infty)$, we consider the truncated random variables $X_{0,a} = \min\{a, X\}$

with density function

$$f_{0,a}(x) = f(x)1\{x < a\} + \overline{F}(a)\delta_a(x)$$

We consider as well the residual random variables $X_{a,\infty} = \{X - a | X \geq a\}$ with density distribution is given by $f_{a,\infty}(x) = \frac{f(x+a)}{\overline{F}(a)}$ $\forall x \geq 0$. The mean is given by

$$E[X_{a,\infty}] = \int_0^\infty \frac{\overline{F}(x+a)}{\overline{F}(x)} dx = \frac{E[X] - E[X_{0,a}]}{\overline{F}(a)}$$

We use the notation $E[T^S_{0,a}]$ and $E[T^S_{a,\infty}]$ to denote the average waiting time in a queue $S \in \{PS, LAS, BPS\}$ with truncated and residual random variable respectively.

4.3 Asymptotic Analysis

We study in this section the effect of giving priority to short jobs on very large jobs. Following the arguments in Section 4.1, the response time for very large jobs in an $MLPS$ system will depend only on the scheduling discipline deployed in the last interval. We denote as $MLPS_S$ an $MLPS$ discipline that utilizes $\{LAS, PS\}$ scheduling in the last interval $a < x < \infty$.

An undesirable property of LAS relies on the fact that in the case of distributions with infinite second moment, there is no asymptote and the expected conditional response time for large jobs deviates from PS [9]. As a consequence, the same result will hold for $MLPS_{LAS}$ disciplines. For example, in the case of Pareto distribution with infinite second moment, the asymptotics of an $MLPS_{LAS}$ has the following form

$$\overline{T}^{LAS}(x) = \frac{1}{1-\rho} x + \frac{\lambda k^\alpha}{(1-\rho)^2(2-\alpha)} x^{2-\alpha} + \mathrm{o}(x^{2-\alpha}).$$

There is no asymptote in this case, even though the limit $\lim_{x\to\infty} \frac{T^{MLPS_{LAS}(x)}}{x}$ exists. This implies that the performance of LAS deviates increasingly from PS performance with the increase of the file size.

We consider now an $MLPS_{PS}$ discipline. From [1] we observe that $T^{MLPS_{PS}}(x)$ if $x \leq a$ will depend on the scheduling disciplines utilized in the corresponding lower priority interval and that if $x > a$

$$T_{MLPS_{PS}}(x) = \frac{a + \overline{W}_{0,a}}{1 - \rho_{0,a}} + \frac{T^{BPS}(x-a)}{1-\rho_{0,a}}, \tag{15}$$

where $\overline{W}_{0,a} = \frac{\lambda E[X^2_{0,a}]}{2(1-\rho_{0,a})}$. We note that $\overline{W}_{0,a}$ is always finite.

In the next theorem, we show that $MLPS_{PS}$ has an asymptote with slope $1/(1-\rho)$ even with service time distributions that have an infinite second moment.

Proposition 1 *Let the service time distribution have finite mean and $\rho < 1$, then the response time for the $MLPS_{PS}$ queue has an asymptote with slope $1/(1-\rho)$ and bias*

$$\lim_{x\to\infty}\left(T_{MLPS_{PS}}(x)-\frac{x}{1-\rho}\right) = \frac{\overline{W}_{0,a}}{1-\rho_{0,a}}+\frac{a(\rho_{0,a}-\rho)}{(1-\rho)(1-\rho_{0,a})} +\frac{bE[X_{a,\infty}](2-(\rho+\rho_{0,a}))}{2(1-\rho)^2},$$

where $\rho_{0,a}=\lambda E[X_{0,a}]$.

Proof 9 *From Theorem 2 and equation (15) it follows that*

$$T_{MLPS_{PS}}(x)=\frac{a+\overline{W}_{0,a}}{1-\rho_{0,a}}+\frac{1}{1-\rho_{0,a}}\frac{x-a}{1-\rho_{a,\infty}}+\frac{1}{1-\rho_{0,a}}\frac{bE[X_{a,\infty}](2-\rho_{a,\infty})}{2(1-\rho_{a,\infty})^2}+o(1)$$

where $\rho_{a,\infty}$ *is the load in the last interval, that is,* $\rho_{a,\infty}=\lambda E[N]E[X_{a,\infty}]$. $E[N]$ *is the mean fraction of flows that reach the low priority queue after a busy period of the high priority queues* $E[N]=\overline{F}(a)/(1-\rho_{0,a})$. *Thus we have*

$$\frac{1}{1-\rho_{a,\infty}}=\frac{1}{1-\lambda\frac{\overline{F}(a)}{1-\rho_{0,a}}\frac{E[X]-E[X_{0,a}]}{\overline{F}(a)}}=\frac{1-\rho_{0,a}}{1-\rho}$$

and similarly $2-\rho_{a,\infty}=\frac{2-(\rho+\rho_{0,a})}{1-\rho_{0,a}}$.

Then we obtain

$$\begin{aligned} T_{MLPS_{PS}}(x) &= \frac{a+\overline{W}_{0,a}}{1-\rho_{0,a}}+\frac{x-a}{1-\rho}+\frac{bE[X_{a,\infty}](2-\rho_{a,\infty})}{2(1-\rho_{a,\infty})(1-\rho)}+o(1)\\ &= \frac{x}{1-\rho}+\frac{\overline{W}_{0,a}}{1-\rho_{0,a}}+\frac{a(\rho_{0,a}-\rho)}{(1-\rho)(1-\rho_{0,a})}+\frac{bE[X_{a,\infty}](2-(\rho+\rho_{0,a}))}{2(1-\rho)^2}+o(1)\end{aligned}$$

This result shows the robustness of $MLPS_{PS}$ disciplines against distributions that have an infinite second moment.

References

[1] Leonard Kleinrock, *Queueing Systems, vol. 2*, John Wiley and Sons, 1976.

[2] L. Kleinrock, R.R. Muntz, and E. Rodemich, "The processor sharing queueing model for time-shared systems with bulk arrivals,," *Networks Journal*, , no. 1, pp. 1–13, 1971.

[3] Nikhil Bansal, "Analysis of the M/G/1 processor-sharing queue with bulk arrivals," *Operations Research Letters*, vol. 31(5), September 2003.

[4] Hanhua Feng and Vishal Misra, "An analysis of scheduling disciplines for network flows," Tech. Rep. Unpublished document, Columbia University, August 2003.

[5] Kiran M. Rege and Bhaskar Sengupta, "The M/G/1 processor-sharing queue with bulk arrivals," in *Proceedings of Modelling and Evaluation of ATM Networks 1993*, 1993, pp. 417–432.

[6] Liang Guo and Ibrahim 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.

[7] Liang Guo and Ibrahim Matta, "Scheduling flows with unknown sizes: Approximate analysis," Tech. Rep. BU-CS-2002-009, Boston University, Mar. 2002.

[8] 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, International Conference on Measurement and Modeling of Computer Systems*, 2003.

[9] K.E. Avrachenkov, U.Ayesta, P.Brown, and E.Nyberg, "Differentiation between short and long TCP flows: Predictability of the response time," in *Proceedings of INFOCOM*, Hong Kong, 2004.

[10] L. Kleinrock and R.R. Muntz, "Multilevel processor-sharing queueing models for time-shared models,," in *Sixth International Teletraffic Congress*, August, 1970, pp. 341/1–341/8.

[11] T.M. O'Donovan, "Distribution of attained service and residual service in general queueing systems," *Operations Research*, , no. 22, pp. 570–575, 1974.

[12] F. Baccelli and P. Bremaud, *Elements of queuing theory: Palm martingale calculus and stochastic recurrences*, Springer, 2003.

[13] Robert Cooper, *Introduction to Queueing Theory*, North-Holland, 1981.

[14] Chatschik Bisdikian, "A note on the conservation law for queues with batch arrivals," *IEEE Transactions on Communications*, vol. 41, no. 6, pp. 832–835, 1993.

[15] S. K. Bose, *Introduction to Queuing Systems*, Kluwer/Plenum, 2001.

[16] L.E. Schrage, "The queue M/G/1 with feedback to lower priority queues," *Management Science*, , no. 13, pp. 466–471, 1967.

[17] S.F. Yashkov, "Processor-sharing queues: Some progress in analysis," *Queueing Systems*, vol. 2, pp. 1–17, 1987.

[18] R. Righter and G. Shanthikumar, "Scheduling multiclass single server queueing systems to stochastically maximize the number of successful departures," *Prob. Eng. Inf Sciences*, vol. 3, pp. 323–334, 1989.

[19] L. Massoulié and J. Roberts, "Bandwidth sharing and admission control for elastic traffic," *Telecommunication Systems*, , no. 15, pp. 185–201, 2000.

[20] Lian Guo and Ibrahim Matta, "The war between mice and elephants," in *Proceedings of the 9th IEEE International Conference on Network Protocols ICNP'01*, 2001.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399