

FB_∞ : An Efficient Scheduling Policy for Edge Routers to Speedup the Internet Access

Idris A. Rai, Guillaume Urvoy-Keller, Ernst W. Biersack

Technical Report TR-02.04.1

Institut Eurecom

2229, route des Crêtes

06904 Sophia-Antipolis, France

{rai,urvoy,erbi}@eurecom.fr

Abstract

Recent Internet traffic measurements reveal that Internet traffic exhibits high coefficient of variability (CoV). That is, the Internet traffic consists of many small jobs and very few large jobs, and less than 1% of the largest jobs constitute more than half of the load. Consequently, we propose to use policies that take advantage of this attribute to favor small jobs over large jobs. If such a policy is implemented in an edge router, this would mean that short jobs such as HTTP sessions will see their latency reduced.

The shortest remaining processing time first (SRPT) scheduling policy has been known to be an optimal policy in minimizing the mean response time. Recent work [2] has shown that for job size distributions with high coefficient of variability (CoV), SRPT favors small jobs without unfairly penalizing large jobs. An implementation of SRPT requires that the sizes of all jobs be known, which can not be assumed in most networking environments. In this paper, we analyze the *Foreground-Background-Infinity* (FB_∞) scheduling policy, which is a priority policy that is known to favor small jobs the most among the scheduling policies that do not require the knowledge of job sizes. However, when evaluated under the $M/M/1$ queueing model, FB_∞ has been shown to highly penalize many large jobs in favor of small jobs. In this paper, we analyze the $M/G/1/FB_\infty$ queue the objective being to investigate the fairness of FB_∞ by comparing its slowdown to the slowdown offered by PS, to quantify the response time improvement that FB_∞ offers when used instead of FIFO, and to compare FB_∞ to an optimal policy SRPT, for service distributions G with varying CoVs. Finally, we consider FB_∞ under overload where we analyze its stability and derive its expression for the

conditional mean response time.

1 Introduction

Recently, the Internet traffic measurements reveal that the Internet traffic exhibits high variability at different levels [5, 6, 16]. Consequently, there is a need to revisit scheduling policies taking into account the observed high variability attribute of the Internet traffic. In this paper, we analyze the $M/G/1/FB_\infty$ queueing policy and compare the results for job size distributions with high variance to job size distributions with low variance.

We consider the *response time* and the *slowdown* as performance metrics. Response time is the overall time a job spends in the system. Slowdown is the normalized response time defined as the ratio of the response time of a job to the size of the job. In particular, we use the **conditional mean response time** and the **conditional mean slowdown**, which (for a job of size x) are defined as $E[T(x)] \triangleq E[(T|X = x)]$ and $E[S(x)] \triangleq E[(S|X = x)]$ respectively. We also use the mean response time and the mean slowdown defined as $E[T] \triangleq \int_0^\infty E[T(x)]f(x)dx$ and $E[S] \triangleq \int_0^\infty E[S(x)]f(x)dx$ respectively. Slowdown is a useful metric to analyze fairness [2]. For instance, the slowdowns of all jobs under a well known fair policy PS are equal.

Scheduling policies that favor small jobs are known to offer lower mean response times than other policies. Therefore, they are more appropriate to schedule highly varying job sizes with most of the jobs small. Shortest job first (SJF) is a nonpreemptive policy that is proven to achieve the minimum mean response time [12, 23]. The preemptive version of the SJF policy is called the *shortest remaining processing time* (SRPT). In [9, 21], SRPT is proven to be the optimal policy in minimizing the mean response time. Moreover, it is demonstrated that at some load values, there exist job size distributions with high variances under which SRPT does not penalize long jobs at all as compared to PS [2], i.e. all jobs have a lower mean slowdown under SRPT than under PS. Implementations of these policies however, are limited to environments where the job sizes are known. Hence, SRPT is not suitable for network routers.

The FB_∞ scheduling policy is a multi-level scheduling policy that also favors small jobs. The subscript ∞ represents an infinite number of queues in the policy. The FB_∞ policy is also called as the shortest elapsed time policy (SET) [3] because it is a preemptive policy that gives service to the job in the system who has received the least service. In the event of ties, the FB_∞ policy requires that the set of jobs

having received the least service time share the processor in a processor-sharing mode [12]. Also, a newly arriving job always preempts the job currently in service and retains the processor until it departs, or until the next arrival appears, or until it has obtained an amount of service equal to that received by the job preempted on arrival, whichever occurs first. As a result, FB_∞ favors small jobs without requiring to know the sizes. An implementation of FB_∞ requires the elapsed service time of each job, which is available from the processor.

The FB_∞ scheduling algorithm has been known to be the most discriminatory policy in favor of small jobs among the policies that do not require the knowledge of job sizes ([12], pp. 172). Despite, FB_∞ has not been used in practice, the main reason being the belief that FB_∞ favors small jobs at the expense of an increased response time for large jobs ([12], pp. 197-198). This objection was observed when FB_∞ is analyzed using M/M/1 model. Also, it is not clear how significant are the improvements of FB_∞ over traditional FIFO or PS policies in terms of reducing the mean response time. Similar objections were raised for SRPT, however it is demonstrated [2] that the performance of SRPT highly depends on the job size distribution. This paper investigate these objections for the FB_∞ policy: FB_∞ is compared to PS in order to evaluate its fairness and its comparison with FIFO shows how much the response time of jobs under FB_∞ is reduced if FB_∞ were deployed in the routers. Additionally, the paper presents an analytical comparison between FB_∞ and the optimal policy SRPT and analyzes the stability of FB_∞ under overload. It turns out that, similar to SRPT, the variance of a job size distribution is also a basic factor on which the performance of FB_∞ in terms of response time and slowdown depends.

We see an immediate application of FB_∞ in the edge routers where scalability due to number of flows is not an issue. We also observe that the reduction of the response time attained by FB_∞ at the edge routers directly results to the reduction in end-to-end delay if the core resources are overprovisioned or guaranteed. Hence, FB_∞ is a suitable policy for edge routers of a VPN with MPLS based core [15, 25].

The outline of the paper is as follows: we present a review of related previous work in the next section and discuss the mathematical background necessary for the analysis carried out in the rest of the paper in Section 3. In Section 4, we analytically compare the FB_∞ policy to the PS, SRPT, and nonpreemptive policies (NPP). In Section 5, we investigate the performance of FB_∞ under overload, and we conclude the paper in Section 6.

2 Previous Work

The high variability of the Internet traffic is observed at different traffic levels, i.e., connection durations, file sizes, FTP transfers, session durations, and session sizes [5, 16]. This attribute is explained by the observation that the Internet traffic consists of many small jobs and more than half of the traffic load is constituted by less than 1% of the largest jobs. This has been referred to as the heavy-tailed property [11, 2, 6], and Pareto distributions have been commonly used to model this attribute. Downey shows in [7] that the Internet dataset complies better to a lognormal distribution than a heavy-tailed distribution. However, recent work in [10] reveals that it is not easy to decide whether the Internet traffic fits heavy-tailed distributions or not. In particular, the results in [10] show that the Internet traffic can fit many distributions with high variance. In fact, in addition to Pareto and lognormal distributions, hyper-exponential, Weibull [5], inverse Gaussian, and hybrid of lognormal and Pareto distributions have been shown to model different traffic traces when their *coefficient of variability* (CoV) is greater than 1 [1, 5, 8]. The CoV, defined as the ratio of the standard deviation to the mean of a distribution, is a common metric to measure the variability of a distribution. The analysis that we conduct in this paper considers job size distributions with varying CoV values while emphasizing the results for distributions with high CoVs.

The expressions for the conditional mean response time $E[T(x)]$ of the FB_∞ are derived in [24, 20, 3, 13]. The response time ($E[T(x)]$) formulas for $M/G/1/FB_\infty$ however, are complex and difficult to evaluate numerically. As a result, very little has been done to analytically evaluate the FB_∞ policy while considering the variance of the job size distribution. In ([3], pp. 188-187), an intuitive result that compares the mean waiting times (the expected delay in the queues) of FB_∞ to PS is given. The result depends on the coefficient of variability and states that the mean waiting time of FB_∞ is lower than that of PS for distributions with coefficient of variability greater than 1, and it is higher for distributions with coefficient of variability less than or equal to 1. In ([12], pp. 196-197), Kleinrock compares the mean waiting times of FB_∞ with that of FIFO for an $M/M/1$ queue model. The results show that FB_∞ offers lower waiting times for short jobs while large jobs suffer a significant penalty. The analysis of $M/G/1/FB_\infty$ queue at steady state is available in [18, 19]. These works derive the steady state distribution of the attained services in the queue. The work in [18, 19] does not consider the variance of the service time distribution. In a recent work [2], a qualitative comparison between FB_∞ and SRPT is presented that states that for a system at underload ($\rho < 1$), every job has a higher expected response time (hence slowdown) under FB_∞ than under SRPT.

3 Mathematical Background

We use the bounded Pareto distribution $BP(k, p, \alpha)$ (where k and p are the minimum and maximum job sizes and α is the exponent of the power law) as a typical example of high variance empirical job sizes and the exponential distribution to represent low variance empirical job sizes. The density functions of the bounded Pareto $f(x)_{BP}$ and the exponential distributions $f(x)_{Exp}$ are given as:

$$f(x)_{BP} = \frac{\alpha k^\alpha}{1 - (k/p)^\alpha} x^{-\alpha-1}, \quad k \leq x \leq p, 0 \leq \alpha \leq 2 \quad (1)$$

$$f(x)_{Exp} = \mu e^{-\mu x}, \quad x \geq 0, \mu \geq 0 \quad (2)$$

The BP distribution can have a very high CoV, whereas the CoV of the exponential distribution is always 1.

Let the average job arrival rate be λ . Assume a job size distribution X with a probability mass function $f(x)$. The abbreviation c.f.m.f.v is used to denote continuous, finite mean, and finite variance. Given the cumulative distribution function as $F(x) \triangleq \int_0^x f(t)dt$, we denote the survivor function of X as $F^c(x) \triangleq 1 - F(x)$. We define $m_n(x) \forall n$ as $m_n(x) \triangleq \int_0^x t^n f(t)dt$. Thus $m_1 \triangleq m_1(\infty)$ is the mean and $m_2 \triangleq m_2(\infty)$ is the second moment of the job size distribution. The load of jobs with sizes less than or equal to x is given as $\rho(x) \triangleq \lambda \int_0^x t f(t)dt$, and $\rho \triangleq \rho(\infty)$ is the total load in the system.

Throughout this paper we assume an M/G/1 queue, where G is a c.f.m.f.v distribution. We make the Poisson arrival hypothesis at the flow level and not at the packet level for which it is known to be unrealistic [17]. The expression of the $E[T(x)]$ for M/G/1/SRPT [22] can be decomposed into the conditional mean waiting time $E[W(x)]$ (the time when a job arrives at the system until it receives the service for the first time) and the conditional mean residence time $E[R(x)]$ (the time a job takes to complete service once the server starts serving it). Hence,

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

$$E[W(x)]_{SRPT} = \frac{\lambda(m_2(x) + x^2 F^c(x))}{2(1 - \rho(x))^2} \quad (3)$$

$$E[R(x)]_{SRPT} = \int_0^x \frac{1}{1 - \rho(t)} d(t) \quad (4)$$

Similarly, we decompose the expression of the $E[T(x)]$ for M/G/1/ FB_∞ into two terms. We denote the first term as $E[\tilde{W}(x)]_{FB_\infty}$ and the second term as $E[\tilde{R}(x)]_{FB_\infty}$. Note that these terms do not denote the conditional mean waiting and residence times for FB_∞ , rather they are introduced to ease the analysis in

Section 4.2 where we compare FB_∞ to SRPT. The formulas for $E[T(x)]_{FB_\infty}$ are given in [20],

$$\begin{aligned} E[T(x)]_{FB_\infty} &= E[\tilde{W}(x)]_{FB_\infty} + E[\tilde{R}(x)]_{FB_\infty} \\ E[\tilde{W}(x)]_{FB_\infty} &= \frac{\lambda(m_2(x) + x^2 F^c(x))}{2(1 - \rho(x) - \lambda x F^c(x))^2} \end{aligned} \quad (5)$$

$$E[\tilde{R}(x)]_{FB_\infty} = \frac{x}{1 - \rho(x) - \lambda x F^c(x)} \quad (6)$$

Finally, the formulas of $E[T(x)]$ for the M/G/1/PS and M/G/1/NPP [4] (where NPP stands for any non-preemptive policy) are:

$$E[T(x)]_{PS} = \frac{x}{1 - \rho} \quad (7)$$

$$E[T(x)]_{NPP} = \frac{\lambda m_2}{2(1 - \rho)} + x \quad (8)$$

The definition of $E[S(x)]$ reveals that for two scheduling policies A and B, the ratio $\frac{E[T(x)]_A}{E[T(x)]_B} = \frac{E[T(x)]_A/x}{E[T(x)]_B/x} = \frac{E[S(x)]_A}{E[S(x)]_B}$.

4 Analytical Results

No analysis on M/G/1/ FB_∞ , for G with varying variances, has been carried out previously. The analysis for M/G/1/ FB_∞ is tedious as it involves nested integrals of complicated functions. Here, we provide elements to compare M/G/1/ FB_∞ to M/G/1/PS, M/G/1/NPP, and M/G/1/SRPT systems.

4.1 Comparison between FB_∞ and PS

Processor sharing is a well known fair policy since at any load it gives the same conditional mean slowdown to all jobs, i.e., $E[S(x)]_{PS} = \frac{1}{(1-\rho)} \forall x$. In this section, we investigate the fairness of the FB_∞ policy by comparing its slowdown to the slowdown offered by PS. We begin the comparison for a general c.f.m.f.v job size distribution and then for job size distributions with high and low CoVs.

4.1.1 Comparison for a general distribution

We define the function ϕ as $\phi(x) \triangleq \lambda \int_0^x t f(t) dt + \lambda x F^c(x)$. If we integrate $\int_0^x t f(t) dt$ by parts in the definition of ϕ we obtain $\phi(x) = \lambda \int_0^x F^c(t) dt$. The following result holds for the conditional mean response time of a job of size x :

Theorem 1 For all c.f.m.f.v job size distributions and load $\rho < 1$,

$$E[T(x)]_{FB_\infty} \leq (1 - \rho) \frac{2 - \phi(x)}{2(1 - \phi(x))^2} E[T(x)]_{PS} \quad (9)$$

$$E[S(x)]_{FB_\infty} \leq (1 - \rho) \frac{2 - \phi(x)}{2(1 - \phi(x))^2} E[S(x)]_{PS} \quad (10)$$

Proof.

$$\begin{aligned} E[T(x)]_{FB_\infty} &= \frac{\lambda \int_0^x t^2 f(t) dt + \lambda x^2 F^c(x)}{2(1 - \phi(x))^2} + \frac{x}{1 - \phi(x)} \\ &\leq \frac{\lambda x \int_0^x t f(t) dt + \lambda x^2 F^c(x)}{2(1 - \phi(x))^2} + \frac{x}{1 - \phi(x)} \quad \left[\int_0^x t^2 f(t) dt \leq x \int_0^x t f(t) dt \right] \\ &= E[T(x)]_{PS} \frac{1 - \rho}{1 - \phi(x)} \left(\frac{\lambda \int_0^x t f(t) dt + \lambda x F^c(x)}{2(1 - \phi(x))} + 1 \right) \quad \left[E[T(x)]_{PS} = \frac{x}{1 - \rho} \right] \\ &= E[T(x)]_{PS} \frac{1 - \rho}{1 - \phi(x)} \left(\frac{\phi(x)}{2(1 - \phi(x))} + 1 \right) \quad [\text{By definition of } \phi(x)] \\ &= (1 - \rho) \frac{2 - \phi(x)}{2(1 - \phi(x))^2} E[T(x)]_{PS} \end{aligned} \quad (11)$$

Equation (10) follows directly by dividing both sides of Equation (11) by x . \square

Next, we use Theorem 1 to elaborate on the comparison between the mean slowdown of FB_∞ and PS.

We first introduce a lemma that we need in the proof of Theorem 2.

Lemma 1 For any load $\rho < 1$,

$$\frac{2 - \phi(x)}{2(1 - \phi(x))^2} \leq \frac{2 - \rho}{2(1 - \rho)^2} \quad (12)$$

Proof.

Function $\phi(x)$ is an increasing function for $x \in [0, \infty]$ since $\frac{d\phi(x)}{dx} = \lambda F^c(x) \geq 0$. Also, function $\frac{2-u}{2(1-u)^2}$ is an increasing function for $u \in [0, 1]$. Hence, $\frac{2-\phi(x)}{2(1-\phi(x))^2}$ is an increasing function of x and upper bounded by $\frac{2-\rho}{2(1-\rho)^2}$ since $\phi(x) \in [0, \rho] \subset [0, 1]$. \square

Theorem 2 For all c.f.m.f.v job size distributions and load $\rho < 1$,

$$E[S]_{FB_\infty} \leq \frac{2 - \rho}{2(1 - \rho)} E[S]_{PS} \quad (13)$$

Proof.

$$\begin{aligned}
E[S]_{FB_\infty} &= \int_0^{+\infty} \frac{E[T(x)]_{FB_\infty}}{x} f(x) dx \\
&\leq \int_0^{+\infty} \frac{2 - \phi(x)}{2(1 - \phi(x))^2} f(x) dx \quad [\text{Theorem 1}] \tag{14}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{2 - \rho}{2(1 - \rho)^2} \int_0^{+\infty} f(x) dx \quad [\text{Lemma 1}] \\
&= \frac{2 - \rho}{2(1 - \rho)^2} \\
&= \frac{2 - \rho}{2(1 - \rho)} E[S]_{PS} \quad [\text{Since } E[S]_{PS} = \frac{1}{1 - \rho}] \tag{15}
\end{aligned}$$

□

Corollary 1 For all c.f.m.f.v job size distributions and load $\rho < 1$,

$$E[T]_{FB_\infty} \leq \frac{2 - \rho}{2(1 - \rho)} E[T]_{PS} \tag{16}$$

Proof. The proof is similar to the proof of Theorem 2. □

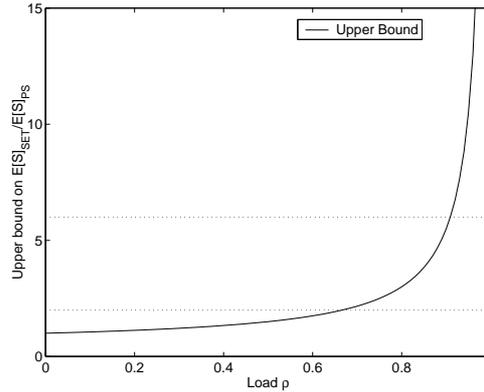


Figure 1: Upper bound on the mean slowdown ratio, $\frac{E[S]_{FB_\infty}}{E[S]_{PS}}$

We illustrate the result of Theorem 2 in Figure 1. We can see that the ratio between the average slowdown of FB_∞ and PS is bounded by a value that is less than 2 for $\rho \leq 0.66$ and less than 6 for $\rho \leq 0.9$. This result clearly supports the fact that using FB_∞ instead of PS results in a better slowdown for small jobs without affecting too much larger jobs since the average slowdown over small and large jobs under PS and FB_∞ remains fairly close for most system loads. Note that the real ratio between the average slowdowns of FB_∞ and PS is below the value of $\frac{2 - \rho}{2(1 - \rho)}$ used in Figure 1 due to the rather crude majorization of

$\phi(x)$ applied in Lemma 1. We will see in the next sections that the fairness of FB_∞ highly depends on the CoV of the job size distribution. In particular, for the case of distributions with high CoVs, a large percentile of jobs have lower slowdown under FB_∞ than under PS.

4.1.2 Distribution dependent comparison

We now analyze the fairness of FB_∞ for two specific job size distributions. Throughout the paper, we use the exponential distribution with a mean of $2.3 * 10^3$ and $CoV = 1$ and the bounded Pareto distribution $BP(332, 10^{10}, 1.1)$ with mean of 3000 and $CoV = 283.9$ as typical examples of job size distributions with a low and a high CoV respectively.

Using Equation (9) (resp. Equation (10)), we identify a lower bound on the percentile¹ of jobs that have a smaller conditional mean response time (resp. slowdown) under FB_∞ than under PS. Since the functions $\phi(x)$ and $\beta(u) \triangleq (1 - \rho) \frac{2-u}{2(1-u)^2}$ are increasing functions of x and u respectively, $\beta(\phi)$ is also increasing of x . Moreover, $\beta(0) = 1 - \rho < 1$ (we study the system in underload). Thus: $\exists x_o = \max\{x' \mid \forall x \in [0, x'], \beta(\phi(x)) \leq 1\}$. The latter means that, according to Equation (9) (resp. Equation (10)), jobs of size $x \leq x_o$ will have a smaller conditional mean response time (resp. slowdown) under FB_∞ than under PS.

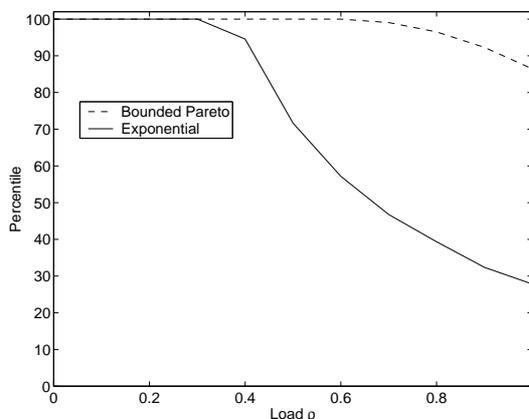


Figure 2: Percentiles of jobs for which $E[S(x)]_{FB_\infty} \leq E[S(x)]_{PS}$ for $BP(332, 10^{10}, 1.1)$ and $Exp(2.3 * 10^3)$ distributions

In Figure 2, we indicate the lower bound of percentile corresponding to x_o for both distributions. The

¹ x is called the 100β percentile for a given random variable X with cumulative distribution function $F(x)$ if the function $Q(\beta)$ is such that $Q(\beta) = \inf_{x \in X} \{F(x) \geq \beta\}$.

Figure clearly illustrates the interest of FB_∞ for the case of a job size distribution with a high CoV since most of the jobs perform better under FB_∞ than under PS. For the BP distribution, independent of the load ρ , at least the jobs in the 85th percentile will have a smaller slowdown under FB_∞ than under PS. For the exponential distribution, the percentile of jobs where FB_∞ outperforms PS decreases rapidly with increasing load for values of load $\rho \geq 0.3$.

However, for the case of the exponential distribution the average slowdown of FB_∞ is always better than the average slowdown of PS as we show in the next theorem.

Theorem 3 *For an exponential job size distribution and at any load $\rho \leq 1$*

$$E[S]_{FB_\infty} \leq E[S]_{PS} \quad (17)$$

Proof. Following the same reasoning as in Theorem 2, one obtains:

$$\begin{aligned} E[S]_{FB_\infty} &= \int_0^{+\infty} \frac{E[T(x)]_{FB_\infty}}{x} f(x) dx \\ &\leq \int_0^{+\infty} \frac{x(2 - \phi(x))}{2(1 - \phi(x))^2} \frac{f(x)}{x} dx \quad [\text{Equation (14)}] \end{aligned} \quad (18)$$

For exponential job sizes, we have $\phi(x) = \int_0^x \lambda F^c(t) dt = \rho(1 - e^{-\mu x}) = \rho F(x)$ and $f(x) = \mu e^{-\mu x}$.

Using these facts and the definition of a function $h(\phi(x)) \triangleq \frac{2 - \phi(x)}{2(1 - \phi(x))^2}$ in Equation 18 we obtain:

$$E[S]_{FB_\infty} \leq \frac{1}{\rho} \int_0^{+\infty} h(\rho F(t)) \rho f(t) dt$$

Replacing $\rho F(t)$ by u in the above equation we get:

$$\begin{aligned} E[S]_{FB_\infty} &\leq \frac{1}{\rho} \int_0^{+\rho} h(u) du \\ &= \frac{1}{2\rho} \left(-\ln(1 - \rho) + \frac{\rho}{1 - \rho} \right) \\ &= \frac{1}{2\rho} \left(-(1 - \rho) \ln(1 - \rho) + \rho \right) E[S]_{PS} \quad [\text{Since } E[S]_{PS} = \frac{1}{1 - \rho}] \end{aligned} \quad (19)$$

A straightforward study of the derivative of ψ defined as $\psi(\rho) \triangleq \frac{(\rho - (-1 - \rho) \ln(1 - \rho))}{2\rho}$ indicates that ψ is a decreasing function and $\psi(\rho) \sim_{\rho \rightarrow 0} \frac{2\rho - \rho^2}{2\rho} \rightarrow 1$. Thus, $\forall \rho \leq 1, \psi(\rho) \leq 1$. \square

Similar result for the BP distribution is not easy to compute. However, it is interesting to note that for $BP(332, 10^{10}, 1.1)$ at load $\rho = 0.9$, more than 99% of the jobs have a lower mean slowdown under FB_∞ than under PS as illustrated in Figure 3. Moreover, the mean slowdowns of less than 1% of the

largest jobs under FB_∞ is only about 10% larger than their mean slowdowns under PS. Hence, for job size distributions with a high CoV, FB_∞ results in a moderate penalty for the very few largest jobs while significantly favoring small jobs.

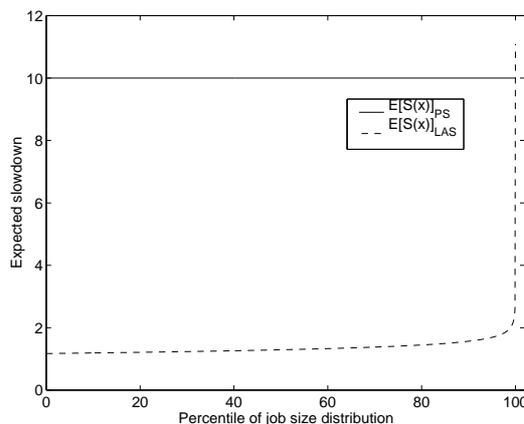


Figure 3: Conditional mean slowdown as a function of percentile for $BP(332, 10^{10}, 1.1)$ job size distribution, at load $\rho = 0.9$

4.2 Quantitative comparison between FB_∞ and SRPT

It is stated in [2] that for any job size distribution and at any load $\rho < 1$ every job x has higher mean response time ($E[T(x)]$) under FB_∞ than under SRPT. However, it was not elaborated how much higher the mean slowdown of a job under FB_∞ can be as compared to under SRPT. In this section, we compare FB_∞ and SRPT quantitatively. We show that at load $\rho < 1$, the conditional mean response time of a job under FB_∞ is quite close to the mean response time under SRPT. In particular, a job under FB_∞ experiences a higher conditional mean response time than under SRPT for job size distributions with low CoVs than for job size distributions with high CoVs.

Theorem 4 Let $\phi(x) \triangleq \rho(x) + x\lambda F^c(x)$, then for all c.f.m.f.v job size distributions and at load $\rho < 1$,

$$E[T(x)]_{SRPT} \leq E[T(x)]_{FB_\infty} \leq \left(\frac{1 - \rho(x)}{1 - \phi(x)} \right)^2 E[T(x)]_{SRPT} \quad (20)$$

Proof.

Since $\rho(x) \leq \phi(x) \leq \rho$, we obtain directly from Equations (3) and (4) and Equations (5) and (6): $E[\tilde{R}(x)]_{FB_\infty} \geq E[R(x)]_{SRPT}$ and $E[\tilde{W}(x)]_{FB_\infty} \geq E[W(x)]_{SRPT}$. Hence the left-hand side inequality

ity. We begin the proof of the right-hand side inequality by studying $E[\tilde{R}(x)]_{FB_\infty} - E[R(x)]_{SRPT}$:

$$\begin{aligned}
E[\tilde{R}(x)]_{FB_\infty} - E[R(x)]_{SRPT} &= \frac{x}{1 - \phi(x)} - \int_0^x \frac{dt}{1 - \rho(t)} \\
&= \int_0^x \frac{dt}{1 - \phi(x)} - \int_0^x \frac{dt}{1 - \rho(t)} \\
&= \int_0^x \frac{\phi(x) - \rho(t)}{1 - \phi(x)} \frac{1}{1 - \rho(t)} dt \\
&\leq \frac{1}{x} \int_0^x \frac{\phi(x) - \rho(t)}{1 - \phi(x)} dt \underbrace{\int_0^x \frac{dt}{1 - \rho(t)}}_{E[R(x)]_{SRPT}} \tag{21}
\end{aligned}$$

The last inequality (Equation (21)) results from the Chebyshev integral inequality [14] applied to $f(t) = \frac{\phi(x) - \rho(t)}{1 - \phi(x)}$, which is a decreasing function of t and $g(t) = \frac{1}{1 - \rho(t)}$, which is an increasing function of t .

Besides:

$$\begin{aligned}
\frac{1}{x} \int_0^x \frac{\phi(x) - \rho(t)}{1 - \phi(x)} dt &= \left([t(\phi(x) - \rho(t))]_0^x + \int_0^x \lambda t^2 f(t) dt \right) \\
&= \frac{1}{x(1 - \phi(x))} (x(\lambda x F^c(x)) + \lambda m_2(x)) \\
&= \frac{\lambda}{x(1 - \phi(x))} (x^2 F^c(x) + m_2(x)) \\
&\leq \frac{\lambda}{x(1 - \phi(x))} (x^2 F^c(x) + x \int_0^x t f(t) dt) \quad \left[\int_0^x t^2 f(t) dt \leq x \int_0^x t f(t) dt \right] \\
&= \frac{x\phi(x)}{x(1 - \phi(x))} \\
&= \frac{\phi(x)}{1 - \phi(x)} \tag{22}
\end{aligned}$$

Using Equation (22) in Equation (21), one obtains:

$$\begin{aligned}
E[\tilde{R}(x)]_{FB_\infty} &\leq \left(1 + \frac{\phi(x)}{1 - \phi(x)} \right) E[R(x)]_{SRPT} \\
&= \frac{1}{1 - \phi(x)} E[R(x)]_{SRPT} \tag{23}
\end{aligned}$$

Consider now $E[\tilde{W}(x)]_{FB_\infty}$ and $E[W(x)]_{SRPT}$, we have:

$$E[\tilde{W}(x)]_{FB_\infty} = \frac{(1 - \rho(x))^2}{(1 - \phi(x))^2} E[W(x)]_{SRPT} \tag{24}$$

Adding Equations (23) and (24), we obtain:

$$\begin{aligned}
E[T(x)]_{FB_\infty} &\leq \max \left(\frac{(1 - \rho(x))^2}{(1 - \phi(x))^2}, \frac{1}{1 - \phi(x)} \right) E[T(x)]_{SRPT} \\
&= \frac{1}{1 - \phi(x)} \max \left(\frac{(1 - \rho(x))^2}{1 - \phi(x)}, 1 \right) E[T(x)]_{SRPT} \tag{25}
\end{aligned}$$

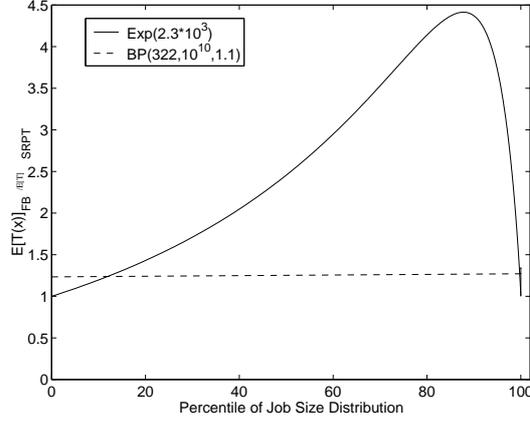


Figure 4: $\frac{E[T(x)]_{FB_\infty}}{E[T(x)]_{SRPT}}$ as a function of percentile of job size distributions, at load $\rho = 0.9$

Since $(1 - \rho(x))^2 = 1 - 2\rho(x) + \rho(x)^2 = 1 - \rho(x) + \underbrace{\rho(x)^2 - \rho(x)}_{=\rho(x)(1-\rho(x)) \geq 0}$ and $1 - \phi(x) = 1 - \rho(x) - \underbrace{x F^c(x)}_{\geq 0}$, then $\frac{(1-\rho(x))^2}{1-\phi(x)} \geq 1$ and the right-hand side of the Theorem follows directly from Equation (25). \square

Figure 4 elaborates the result of Theorem 4 for the $BP(332, 10^{10}, 1.1)$ distribution and the exponential distribution with a mean of $2.3 * 10^3$ as a function of percentiles of job sizes. We observe that the ratio between $E[T(x)]_{FB_\infty}$ and $E[T(x)]_{SRPT}$ highly depends on the CoV of a job size distribution. We see that for the bounded Pareto distribution, the ratio $\frac{E[T(x)]_{FB_\infty}}{E[T(x)]_{SRPT}}$ is about 1.25 for all jobs, which shows that the mean response times $E[T(x)]$ of jobs offered by FB_∞ are quite close to the mean response times $E[T(x)]$ of the jobs under SRPT. On the other hand, most of the exponentially distributed jobs experience significantly higher conditional mean response times $E[T(x)]_{FB_\infty}$ under FB_∞ than under SRPT. Observe that the ratio $\frac{E[T(x)]_{FB_\infty}}{E[T(x)]_{SRPT}}$ approaches 4.5 for some large jobs. Note also that as $x \rightarrow 100th$ percentile job size, the ratio $\frac{E[T(x)]_{FB_\infty}}{E[T(x)]_{SRPT}} \rightarrow 1$ because $\lim_{x \rightarrow \infty} \phi(x) = \rho(x)$.

4.3 Nonpreemptive polices (NPP)

To see the benefits that FB_∞ can offer when used in routers, we ought to compare it to the policies that are currently implemented. It is well known that the mean response time $E[T]$ is the same for all nonpreemptive policies [4]. Examples of nonpreemptive policies include FIFO, LIFO, and RANDOM. Hence, the comparison that we conduct here applies to the FIFO policy, which is a nonpreemptive policy that is currently widely implemented in the routers. From Equation (8), the mean response time for

nonpreemptive policies is given as:

$$E[T]_{NPP} = \frac{\lambda m_2}{2(1 - \rho)} + E(X)$$

In ([3], pp. 188), the following relation between the mean waiting times of NPP and PS is given:

$$E[W]_{PS} = E[W]_{NPP} - \frac{\rho[C^2(X) - 1]}{2(1 - \rho)}E(X) \quad (26)$$

where $C(X)$ is the coefficient of variability (CoV) and $E(X)$ is the mean job size. The mean response time, $E[T]$, is given as $E[T] = E[W] + E(X)$. Using this relation in Equation (26) we obtain the relation between the mean response times of the PS and NPP policies as:

$$E[T]_{PS} = E[T]_{NPP} - \frac{\rho[C^2(X) - 1]}{2(1 - \rho)}E(X). \quad (27)$$

Replacing $E[T]_{PS}$ in Corollary 1 by the right hand side of Equation (27) allows us to bound the mean response time of FB_∞ as follows:

$$E[T]_{FB_\infty} \leq \frac{(2 - \rho)}{2(1 - \rho)}E[T]_{NPP} - \frac{\rho(2 - \rho)[C^2(X) - 1]}{4(1 - \rho)^2}E(X) \quad (28)$$

We note that the bound of the mean response time of FB_∞ is a function of the mean response time of a nonpreemptive policy, the load ρ , and the CoV. This bound is interesting since it enables us to compare the performance of FB_∞ relative to that of any nonpreemptive policy for a large range of distributions.

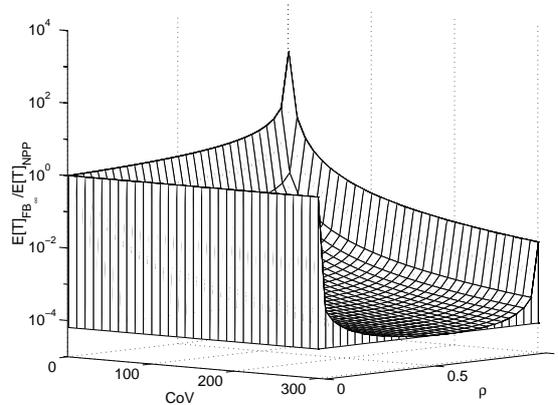


Figure 5: Upper bound on the mean response time ratio, $\frac{E[T]_{FB_\infty}}{E[T]_{NPP}}$, as a function of load ρ and coefficient of variability

Figure 5 shows the upper bound on the ratio of the mean response time of FB_∞ to the mean response time of a nonpreemptive policy as a function of $CoV \geq 1$ and load $\rho < 1$. Observe that FB_∞ has a higher response time than nonpreemptive policies only for distributions with CoVs close to 1. In this case,

we also observe that the mean response time (resp. slowdown) of FB_∞ increases to large values with increasing load. On the other hand, the mean response time of FB_∞ is lower than that of nonpreemptive policies for high variance distributions at all load values. For a given load ρ , the ratio $\frac{E[T]_{FB_\infty}}{E[T]_{NPP}}$ decreases with increasing CoV.

In summary, we see that FB_∞ is a more appropriate policy than any NPP policy, in particular the FIFO policy, when the job size distribution has a high variance. This is to be expected because in contrast to FB_∞ , the service of a job under FIFO is not interrupted until the job leaves the system and so large jobs are favored over small jobs.

5 FB_∞ under Overload

In real systems, it may happen that jobs arrive to the system at a higher rate than the rate at which they are serviced. This is termed as overload condition under which load $\rho > 1$. To the best of our knowledge, no analysis of FB_∞ under overload has been conducted. Here, we show that FB_∞ is stable for some job sizes at $\rho > 1$ and we derive the formulas of the conditional mean response time for FB_∞ under overload. SRPT is also proven to be stable under overload [2]. In this section, we also compare FB_∞ to SRPT under overload.

Theorem 5 *Let $\theta(x) \triangleq m_1(x) + xF^c(x)$, λ be the job mean arrival rate, and $f(t)$ be a service time distribution with a mean of m_1 . Then, for any load $\rho = \lambda m_1 > 1$, every job size $x \leq x_{FB_\infty}(\lambda)$ with $x_{FB_\infty}(\lambda) \triangleq \max\{x \mid \theta(x) \leq \frac{1}{\lambda}\}$ has a finite conditional mean response time under FB_∞ , whereas every job of size $x > x_{FB_\infty}(\lambda)$ experiences an infinite response time.*

Proof. The load offered to the server when FB_∞ system is under overload is $\rho = \lambda m_1 > 1$. However, the effective load $\rho_{\text{effective}}$ that corresponds to the work serviced by the server is equal to 1. Let $\theta_{\text{effective}} \triangleq \frac{\rho_{\text{effective}}}{\lambda} = \frac{1}{\lambda}$. Then, $\theta_{\text{effective}}$ is the expected service offered to the set of jobs that access the server under overload. This set depends on the policy. With FB_∞ , every newly arriving job in the system gets an immediate access to the server. On the average a job receives a service of $\theta_{\text{effective}}$, but since some jobs require less than $\theta_{\text{effective}}$, other jobs may receive more service than $\theta_{\text{effective}}$, namely up to $x_{FB_\infty}(\lambda)$ defined as:

$$x_{FB_\infty}(\lambda) = \max\{x \mid \theta(x) \leq \frac{1}{\lambda}\} \quad (29)$$

Hence, for FB_∞ under overload, one obtains $\rho_{\text{effective}} = \lambda\theta(x_{FB_\infty}) = 1$. \square

Corollary 2 *For SRPT under overload, every job of size $x \leq x_{\text{SRPT}}(\lambda)$ with $x_{\text{SRPT}}(\lambda) \triangleq \max\{x \mid m_1(x) \leq \frac{1}{\lambda}\}$ has a finite conditional mean response time, whereas every job of size $x > x_{\text{SRPT}}(\lambda)$ experiences an infinite response time*

Proof. The reasoning for SRPT is different from the one for FB_∞ . Note that the service of a job under SRPT is not affected by the arrival of jobs with larger sizes. Hence, a set of jobs with size less than or equal to x contributes a system load of $\rho(x) = \lambda \int_0^x t f(t) dt$. Thus under overload, only jobs with sizes less than or equal to $x_{\text{SRPT}}(\lambda)$ receive service, where $x_{\text{SRPT}}(\lambda) = \max\{x \mid \rho(x) \leq \rho_{\text{effective}} = 1\}$, which is equivalent to $x_{\text{SRPT}}(\lambda) = \max\{x \mid m_1(x) \leq \frac{1}{\lambda}\}$. The jobs with size greater than $x_{\text{SRPT}}(\lambda)$ can not access the server since the jobs with sizes less than $x_{\text{SRPT}}(\lambda)$ always preempt them to maintain $\rho_{\text{effective}} = 1$. \square

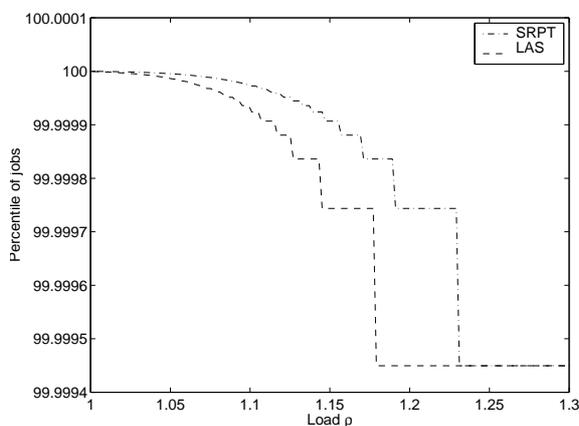


Figure 6: *Percentile of the largest job sizes that can be serviced under overload for $BP(332, 10^{10}, 1.1)$ job size distribution*

Note that $x_{\text{SRPT}}(\lambda) \geq x_{FB_\infty}(\lambda)$ since with SRPT under overload, a job accesses the server only if its service can be entirely completed, which is not the case with FB_∞ . We observe in Figure 6 that for $BP(332, 10^{10}, 1.1)$, the maximum job size that SRPT can service under overload is larger than the job size that FB_∞ can service at a given load up to $\rho \leq 1.23$. For a load above 1.23, we note that for the job size distribution considered, the maximum job that both scheduling policies can service is the same. It is worth mentioning that the percentile of each job that is serviced by both policies under overload is above the 99 percentile. This result is to be expected since for the job size distribution considered, more than

99% of the jobs are small jobs and the fraction of load that these small jobs contribute to the system load is less than 50% of the total load.

Now we compute the formulas for the conditional mean response time of the jobs under overload. For the case of the SRPT policy, the formulas in underload and overload are different [2]. This may be explained by the two observations that all the jobs of size $x \leq x_{\text{SRPT}}(\lambda)$ with $\rho(x_{\text{SRPT}}) \triangleq \int_0^{x_{\text{SRPT}}(\lambda)} t f(t) dt = 1$ have a finite response time, and all the jobs of size $x > x_{\text{SRPT}}(\lambda)$ receive no service at all. Therefore, in overload, the SRPT system works as if there are no jobs of size $x > x_{\text{SRPT}}(\lambda)$. Hence, the moment² $m_2(x) + x^2 F^c(x)$ that appears in the formulas of $E[T(x)]_{\text{SRPT}}$ is computed only for the jobs of size $x \leq x_{\text{SRPT}}(\lambda)$, whereas in underload, $E[T(x)]_{\text{SRPT}}$ is computed for all jobs.

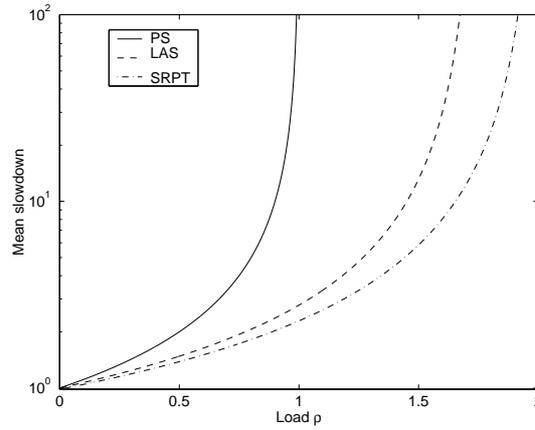


Figure 7: Mean slowdown for the 99th percentile job for $BP(322, 10^{10}, 1.1)$ as a function of load

For FB_∞ under overload, all jobs can receive up to $x_{FB_\infty}(\lambda)$ of service (if their initial requirement is larger than $x_{FB_\infty}(\lambda)$, they receive only $x_{FB_\infty}(\lambda)$). Therefore, the moments² $m_2(x) + x^2 F^c(x)$ and $m_1(x) + x F^c(x)$ that appear in the formulas for $E[T(x)]_{FB_\infty}$ must be computed over all job sizes. As a consequence, the formulas for the conditional mean response time in overload and underload are identical:

Theorem 6 *The conditional mean response time of a job size x under FB_∞ is:*

$$E[T(x)]_{FB_\infty} = \begin{cases} \frac{\lambda(m_2(x) + x^2(1-F(x)))}{2(1-\rho(x) - \lambda x(1-F(x)))^2} + \frac{x}{1-\rho(x) - \lambda x(1-F(x))} & \text{if } x \leq x_{FB_\infty}(\lambda) \\ +\infty & \text{if } x > x_{FB_\infty}(\lambda) \end{cases}$$

² $m_2(x) + x^2 F^c(x)$ and $m_1(x) + x F^c(x)$ are the moments of a truncated distribution $f_x(y)$ (with $f_x(y) = f(y)$ if $y < x$, $f_x(y) = F^c(x)$ if $y = x$, and $f_x(y) = 0$ if $y > x$) that account for the contribution of all jobs of size y to the response time of the job of size x (see ([12], pp. 173).

Figure 7 shows the mean slowdown of the 99th percentile job of the $BP(332, 10^{10}, 1.1)$ job size distribution under PS, FB_∞ , and SRPT under overload. At load $\rho = 1.4$, the mean slowdown of the job under PS is infinity, whereas it is 10 and 4 under FB_∞ and under SRPT respectively. It is obvious that FB_∞ becomes unstable at lower load values than does SRPT. However, we see that FB_∞ is quite close to the optimal policy SRPT in terms of the mean slowdown of a job even under overload.

6 Conclusion

The goal of this paper is to analyze the $M/G/1/FB_\infty$ queuing model in order to evaluate the performance of the FB_∞ for job size distributions with different coefficient of variability (CoV). We show through analysis and numerical evaluation that the CoV of a job size distribution is important in determining the performance of FB_∞ in terms of response time and slowdown. In particular, we demonstrated that the percentage of jobs that have higher slowdowns under FB_∞ than under PS is lower for job sizes with a high CoV than job sizes with a low CoV. For the case of the exponential job size distribution, we proved that the mean slowdown for FB_∞ is always less than or equal to the mean slowdown for PS. Even for a general job size distribution, we showed that for moderate load values, the mean slowdown $E[S]$ of FB_∞ remains fairly close to the mean slowdown of PS.

FIFO is a nonpreemptive policy that is currently implemented in most routers, hence its comparison with FB_∞ is important to analyze the benefits in case FB_∞ were deployed in the routers. We obtained an upper bound that compares the mean response time of FB_∞ to the mean response time of nonpreemptive policies. We showed that the FB_∞ policy offers a lower mean response time than nonpreemptive policies for job size distributions with high CoVs. We also provide comparison of FB_∞ to the optimal policy SRPT. We demonstrated that $E[T(x)]_{FB_\infty}$ is closer to the $E[T(x)]_{SRPT}$ for job size distributions with high CoVs than job sizes with low CoVs.

We also proved that FB_∞ is stable under overload and obtained the expression for the conditional mean response time $E[T(x)]$ for FB_∞ under overload.

While the immediate application of the FB_∞ policy is seen in the edge routers, its ability to service some jobs under overload is a potential advantage over FIFO or PS when FB_∞ is deployed in the core routers. Hence, in the future work, we will investigate the impact of the FB_∞ policy in the core routers of the network in terms of reducing end to end latency.

References

- [1] V. Almeida et al., “Characterizing Reference Locality in the WWW”, In *Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems (PDIS'96)*, pp. 92–103, 1996.
- [2] N. Bansal and M. Harchol-Balter, “Analysis of SRPT Scheduling: Investigating Unfairness”, In *Sigmetrics 2001 / Performance 2001*, pp. 279–290, June 2001.
- [3] E. G. Coffman and P. J. Denning, *Operating Systems Theory*, Prentice-Hall Inc., 1973.
- [4] R. W. Conway, W. L. Maxwell, and W. Louis, *Theory of Scheduling*, Addison-Wesley Publishing Company, 1967.
- [5] M. Crovella and A. Bestavros, “Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes”, *IEEE/ACM Transactions on Networking*, pp. 835–846, December 1997.
- [6] M. E. Crovella, “Performance Evaluation with Heavy Tailed Distributions”, In *Job Scheduling Strategies for Parallel Processing 2001 (JSSPP)*, pp. 1–10, 2001.
- [7] A. B. Downey, “The structural cause of file size distributions”, In *Proc. MASCOTS 2000I*, August 2001.
- [8] M. J. Fischer et al., “Analyzing the Waiting Time Process in Internet Queueing Systems With the Transform Approximation Method”, *The Telecommunications Review*, pp. 21–32, 2001.
- [9] J. E. Gehrke et al., “Scheduling to Minimize Average Stretch Online”, In *40th Annual Symposium on Foundation of Computer Science*, pp. 422–433, 1999.
- [10] W. Gong et al., “On the Tails of Web File Size Distributions”, In *Proc. of 39-th Allerton Conference on Communication, Control, and Computing*, October 2001.
- [11] M. Harchol-Balter, “The Effect of Heavy-Tailed Job Size. Distributions on Computer System Design”, In *Proc. of ASA-IMS Conf. on Applications of Heavy Tailed Distributions in Economics*, June 1999.
- [12] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications*, Wiley, New York, 1976.
- [13] H. Krayl, E. J. Neuhold, and C. Unger, *Grundlagen der Betriebssysteme*, Walter de Gruyter, Berlin, New York, 1975.

- [14] D. S. Mitrinovic, *Analytic Inequalities*, Springer-Verlag, 1970.
- [15] K. Muthukrishnan et al., “A Core MPLS IP VPN Architecture”, Request for Comments RFC 2917, Internet Engineering Task Force, September 2000.
- [16] V. Paxson and S. Floyd, “Wide-Area Traffic: The failure of Poisson Modelling”, *IEEE/ACM Transactions on Networking*, 3:226–244, June 1995.
- [17] J. Roberts, “Traffic Theory and the Internet”, In *IEEE Comm. Magazine*, pp. 94–99, 2001.
- [18] R. Schassberger, “A Detailed Steady State Analysis of the M/G/1 Queue under various Time Sharing Disciplines”, In P. J. C. In O.J. Boxma G. Iazeolla, editor, *Computer Performance and Reliability*, pp. 431–442, North Holland, 1987.
- [19] R. Schassberger, “The Steady State Distribution of Spent Service Times Present in the M/G/1 Foreground-Background Processor-sharing Queue”, *Journal of Applied Probability*, 25(7):194–203, 1988.
- [20] L. E. Schrage, “The queue M/G/1 with feedback to lower priority queues”, *Management Science*, 13(7):466–474, 1967.
- [21] L. E. Schrage, “A Proof of the Optimality of the Shortest Remaining Service Time Discipline”, *Operations Research*, 16:670–690, 1968.
- [22] L. E. Schrage and L. W. Miller, “The queue M/G/1 with the shortest processing remaining time discipline”, *Operations Research*, 14:670–684, 1966.
- [23] A. Silberschatz et al., *Applied Operating System Concepts*, John Wiley and Sons, first edition, 2000.
- [24] R. W. Wolff, “Time Sharing with Priorities”, *SIAM Journal of Applied Mathematics*, 19(3):566–574, 1970.
- [25] P. Zhang and R. Kantola, “Building MPLS VPNs with QoS Routing Capability”, In *Fifth International Symposium on Interworking*, pp. 292–301, October 2000.