A Square Root Formula for the Rate of Non-Persistent TCP Flows

François Baccelli, David R. Mcdonald

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Square Root Formula for the Rate of Non-Persistent TCP Flows

François Baccelli — David R. McDonald

N° 5301

Août 2004

THÈME 1

ISSN 0249-6399 ISRN INRIA/RR--5301--FR+ENG

A Square Root Formula for the Rate of Non-Persistent TCP Flows

François Baccelli* , David R. McDonald †

Thème 1 — Réseaux et systèmes
Projet TREC

Rapport de recherche n° 5301 — Août 2004 — 27 pages

Abstract: In this paper, we derive a closed form formula for the average rate attained by a non persistent TCP source which alternates between idle periods and download periods subject to a fixed packet loss probability. We also derive closed form expressions for the mean time to transfer a file and for the distribution of the transmission rate. Several distributions for the file sizes and idle times are considered including heavy tailed distributions. The formula for the mean transmission rate is shown to boil down to the classical square root mean value formula for persistent flows when the average file size tends to infinity. Using fixed point methods, these formulae can be applied to predict bandwidth sharing among competing HTTP flows subject to Active Queue Management.

Key-words: TCP, congestion control, flow control, additive increase–multiplicative decrease algorithm, IP traffic, on-off flow, HTTP, throughput.

* INRIA-ENS, 45 rue d'Ulm 75005, Paris, France, francois.baccelli@ens.fr
† University of Ottawa, 585 King Edward Av. Ottawa, Ontario, Canada K1N 6N5 and INRIA-ENS dmdsg@mathstat.uottawa.ca

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Une Formule en Racine Carrée pour le Débit de Flots TCP Non-Persistants

Résumé : Cet article donne une formule explicite pour le débit moyen atteint par un flot TCP non persistant qui alterne entre des périodes de téléchargement et de lecture et dont les paquets subissent des pertes à taux constant. Il contient aussi des formules explicites pour la durée moyenne de téléchargement d'un fichier et pour la distribution du débit de transfert des données. Plusieurs classes de distributions de tailles de fichiers et de durées de lecture sont considérées, et notamment des distributions à queues lourdes. La formule du débit moyen se réduit à celle bien connue pour les flux persistants lorsque la taille des fichiers tend vers l'infini. Combinées à des méthodes de point fixe, ces formules peuvent être utilisées pour prédire le partage de la bande passante réalisé entre plusieurs flots HTTP par un mécanisme de contrôle actif de file d'attente de type AQM.

Mots-clés : TCP, contrôle de congestion, contrôle de flux algorithme de croissance additive et décroissance multiplicative, trafic IP, trafic IP, flot on-off, HTTP, débit.

1 Introduction

The most basic formula for predicting the performance of long lived TCP flows is the square root formula, see [13]. This formula shows that the mean window size is inversely proportional to the square root of the probability a packet is dropped. Since the transmission rate of a source is the window size divided by the round trip time this formula determines the mean bandwidth allocated to a TCP flow. This formula assumes a regime with a constant drop probability. Such a regime might arise with certain Active Queue Management (AQM) schemes which stabilize the flows through losses at a congested buffer.

The aim of this paper is to give a generalization of this square root formula for the mean transmission rate of non persistent flows subject to such a constant drop probability. By non-persistent flows, we mean flows that alternate between ON periods when files of random sizes are downloaded and OFF periods which consist of think times of random durations. This ON-OFF structure is the simplest possible model for flows generated when a user consults a web site. Clicking on a link starts a download of a page via HTTP. The user then peruses the page in the following think period before clicking on a new link. Within this context, it is well known that it is appropriate to assume that the distribution of file sizes and OFF periods have heavy tails (e.g. Pareto file sizes and Weibull or lognormal OFF periods, as for example in [7]).

Earlier papers analyzing non-persistent TCP flows include [10], [14], [11], [9], [7] and [6].

[10] proposes a version of the Engset model which is shown to be insensitive w.r.t the file size distribution. The effect of TCP is modeled as a constant transfer rate calculated from the study of TCP's bandwidth sharing for a fixed number of persistent flows that are exactly in phase.

In the models considered in [11] and [6], losses take place at certain congestion epochs and the inter-congestion periods are dynamically changing with traffic. In [11], the flows contributing to the traffic are all in phase (they all react together at the same time and in the same way, a case which is referred to as full synchronization). Analytical results are derived in the low load, large file case where TCP's bandwidth sharing can be approximated by a completely fair allocation. [6] addresses the case where only a proportion of the flows lose packets at congestion epochs (case with partial synchronization) but does not lead to closed form expressions for the stationary rate.

[7] extends the utility function approach initially developed for the representation of the bandwidth sharing of persistent TCP flows to the non persistent-case.

[14] and [9] use the processor sharing heuristic to model the case of a Poisson point process of sessions, each associated with a file download having a general distribution.

To the best of our knowledge, none of these earlier papers provides a closed form formula for the mean rate of a non-persistent flow that takes into account all the following features:

- key features of TCP like slow start and the actual AIMD dynamics;
- networking parameters such as the RTT and the network packet loss;
- application parameters like the distributions of the file sizes and of the think times.

Such a formula is provided in §2 for the case of exponential file sizes and exponential think times and some heavy tailed cases. This mean rate formula is complemented by expressions of interest to users, like the mean time to transfer a file or the mean time to transfer a file of a given size.

Section 3 goes beyond mean values and provides a formula for the stationary distribution of the rate, which allows one to give a formula for the proportion of the time where the user gets a rate larger than some predefined value. Section 4 gathers a few applications of this formula to the prediction of bandwidth sharing for both homogeneous and heterogeneous cases. Finally, Section 5 proposes a way of estimating some parameters of the slow start model. All mathematical proofs, which are primarily based on Laplace and Mellin transform techniques, are deferred to the appendix.

The general model considered throughout the paper is as follows. We assume a non-persistent flow is silent for a random time with distribution F and mean $1/\beta$. After the silence period the flow transmits a file. The distribution of file sizes is G with a mean $1/\mu$. The flow is subject to a constant packet loss probability p. It is assumed that each flow implements the congestion avoidance phase of TCP Reno: the transmission rate increases at rate $1/R^2$ during the transmission of a file, where R is the round trip time (RTT) of packets and is cut in two when a packet is lost. When the file has been transmitted the transmission rate is reset to zero.

Within this framework we propose to represent slow start by an instantaneous jump of the rate of some random size at the epochs when a flow switches from OFF to ON, as was already done in [6]. The rationale for this simplified model is that the associated exponential growth phase is quite quick compared to the congestion avoidance phase so that it can to a first approximation, be seen as a jump from 0 to some random value with distribution H and density h that can be estimated as will be seen in §5.2.

Throughout the paper, we will assume as in the classical square root formula [13] that

- there is a constant packet loss probability p;
- the RTT is constant.

These constants can nevertheless be solutions of certain fixed point equations as customary within the framework of TCP analysis.

The trajectory of a typical flow is depicted in Figure 1.

2 Mean Values

2.1 Notation

We will assume that file sizes have a general distribution G with mean μ^{-1} and the think times have a general distribution F with mean β^{-1}. The size of the jump representing slow start is a random variable with law H. We will denote by $s(z)$ the stationary density of the stationary rate of a flow and by ν the stationary probability that this flow is inactive.

The first subsections of the present section focus on the case where G is exponential with parameter μ.

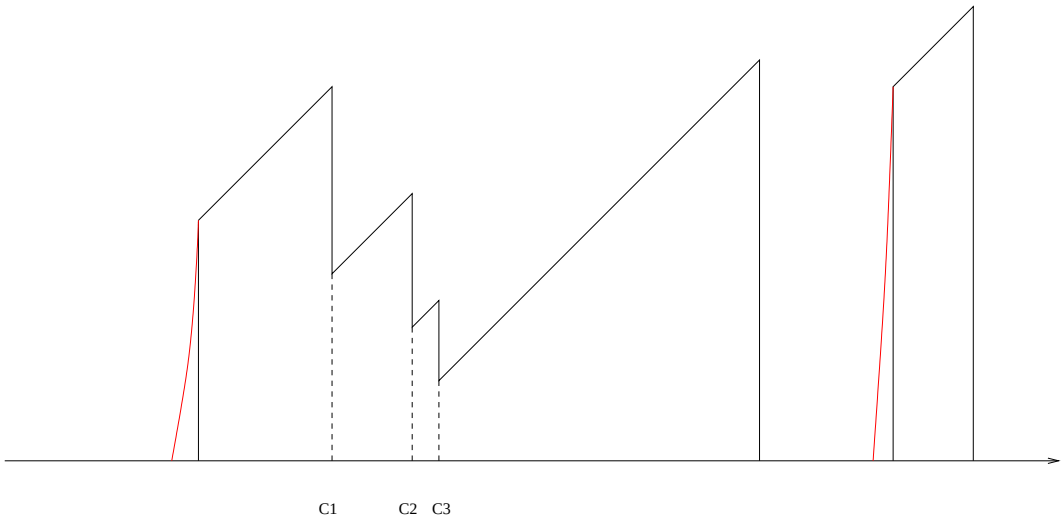

Figure 1: Evolution of the rate of a non persistent flow with time. The exponential slow start phase is in red. The jump approximation in black.

2.2 Throughput

Any distribution function Φ on the positive real line has an associated Mellin transform (see e.g. [12], [3]):

$$\widehat{\phi}(u) = \int_0^\infty z^{u-1}\Phi(dz) \text{ where } u \geq 1. \tag{1}$$

If Φ has a density ϕ then $\widehat{\phi}(u) = \int_0^\infty \phi(z)z^{u-1}dz$. Below we denote by $\widehat{h}(u)$ the Mellin transform of H.

Let also

$$\Pi_k(u) = \prod_{l=0}^{k-1}\left(1 - \frac{p}{p+\mu}2^{-u-2l}\right). \tag{2}$$

Theorem 1 *When G is exponential, the mean time T (or mean latency) to transfer a file is:*

$$\begin{aligned} T &= \frac{\sqrt{\pi}}{\mu}\frac{\Pi_\infty(1)}{\Pi_\infty(2)}\sqrt{\frac{(p+\mu)R^2}{2}} \\ &+ \frac{\sqrt{\pi}R^2}{2}\sum_{k=0}^{\infty}\left(\Pi_k(2)\frac{\Pi_\infty(1)}{\Pi_\infty(2)}\widehat{h}(2k+3)\frac{\left(\frac{(p+\mu)R^2}{2}\right)^{k+\frac{1}{2}}}{(k+1)!} - \Pi_k(1)\widehat{h}(2k+2)\frac{\left(\frac{(p+\mu)R^2}{2}\right)^{k}}{\Gamma(k+\frac{3}{2})}\right). \end{aligned} \tag{3}$$

The mean rate of a stationary flow is

$$M = \frac{\frac{1}{\mu}}{\frac{1}{\beta} + T} \tag{4}$$

and the probability that a flow is ON is

$$\nu = 1 - M\frac{\mu}{\beta}. \tag{5}$$

Figure 2 shows the how mean throughput expression in Formula (4) varies with p and with H.

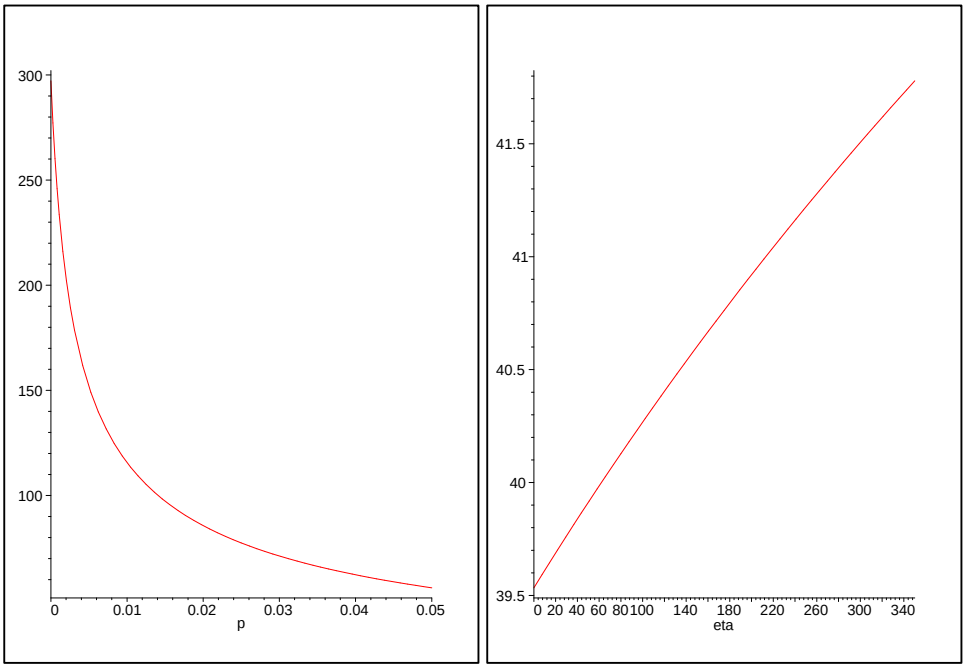

Figure 2: Dependence of Formula (4) on p (left) and H (right). Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 2000$ pkts. On the left, p varies and $H = \delta(100)$ (the delta function at 100 pkts) and on the right $p = 10\%$ while $H = \delta(\eta)$ where η varies.

Figure 3 plots the mean time to transfer a file for two values η of $H = \delta(\eta)$.

Figure 4 plots the mean time to transfer a file divided by the mean file size in function of the mean file size. The shape of the curves of Figure 4 can be explained as follows:

- For $H = \delta(0)$ and the mean file size m small, $T \sim R\sqrt{2m}$ so that $T/m \sim R\sqrt{2/m}$ gets large.

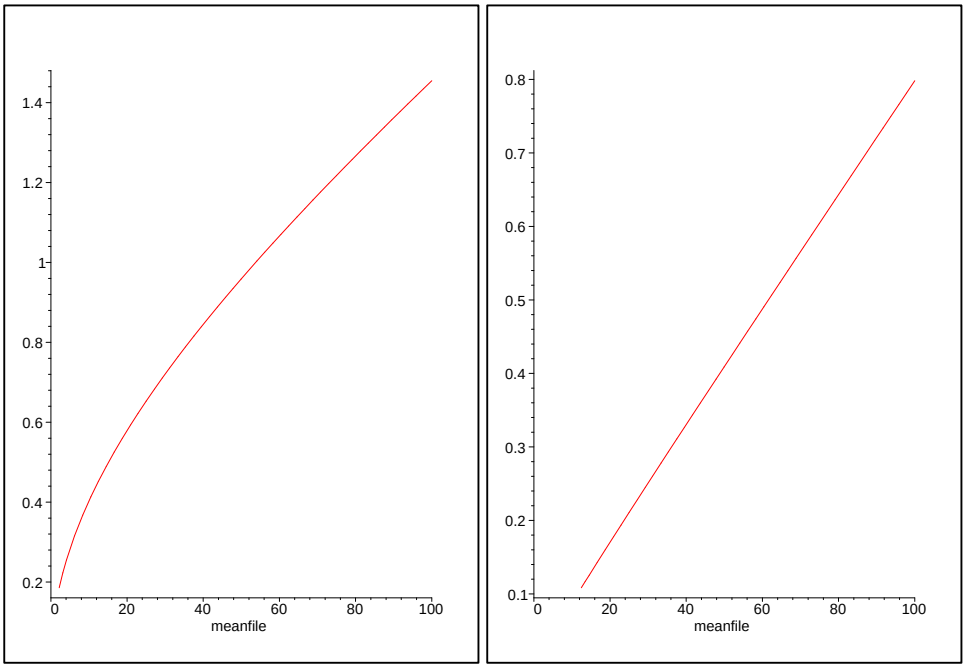

Figure 3: Mean time to transfer a file for $H = \delta(0)$ (left) and $H = \delta(100)$ (right). Here $R = 0.1$ s., $1/\beta = 2$ s., $p = 1\%$ and $1/\mu$ varies.

- For $H = \delta(\eta)$ with $\eta > 0$ and the mean file size m small,

$$T \sim \sqrt{\eta^2 R^4 + 2R^2 m} - \eta R^2 \sim \frac{m}{\eta}$$

  so that T/m tends to $1/\eta$.

2.3 Latency

A question of some practical importance is the mean value $\lambda(t)$ of the delay to transfer a file of size t within this setting. By a direct conditioning argument, we get that

$$\int_{x=0}^{\infty} \mu e^{-\mu x} \lambda(x) dx = T(\mu), \tag{6}$$

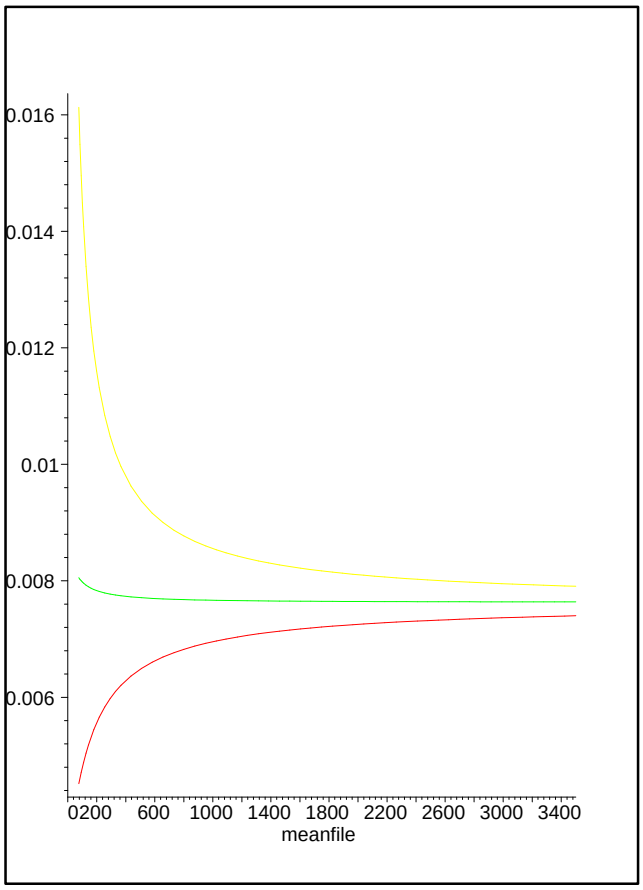

Figure 4: Mean time T to transfer a file as given by (3) divided by mean file size μ^{-1} for various values of η (deterministic case). Here $R = 0.1$ s., $1/\beta = 2$, $p = 1\%$, $\eta = 200$ (red), 100 (green) and 0 (yellow) and $1/\mu$ varies.

with $T(\mu)$ the function of μ given in closed form by (3). So $\lambda(x)$ is in fact obtained from the inverse Laplace transform of the function $T(\mu)$ at point x by the formula:

$$\lambda(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{1}{\mu} e^{x\mu} T(\mu) d\mu, \tag{7}$$

where the last expression is valid for all $c > 0$. Several other numerical methods for Laplace inversion could possible be used for computing $\lambda(x)$. Direct inversion will be investigated in §2.4.2 below.

2.4 Special cases

2.4.1 Large File Sizes

When μ tends to zero in (4), we immediately get that the mean value given in Theorem 1 tends to the classical persistent flow square roor formula.

Corollary 1 *For G exponential, when the mean file size tends to infinity, M tends to*

$$\frac{\alpha}{R\sqrt{p}} \text{ where } \alpha = \sqrt{\frac{2}{\pi}} \frac{\prod_{l=1}^{\infty}\left(1-4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-24^{-l}\right)} \sim 1.309 \tag{8}$$

which corresponds to the results of [5] (see Formula (4.7), page 90) and ν tends to 0.

2.4.2 No Slow Start

The case with no slow start is obtained when taking $\widehat{h}(u) = 0$.

Corollary 2 *For G exponential, in the special case when there is no slow start,*

$$M = \frac{1}{\frac{\mu}{\beta} + \sqrt{\frac{\pi}{2}} R \frac{\prod_{l=1}^{\infty}\left(1-\frac{2p}{p+\mu}4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-\frac{p}{p+\mu}4^{-l}\right)}\sqrt{p+\mu}}. \tag{9}$$

The last expression is the simplest exact formula extending the square root formula [13] to the non persistent flow case. Figure 5 illustrates the joint dependency of this formula in p and μ.

Figure 6 shows the discrepancy between the square root formula (8) and Formula (9) can be arbitrarily large for small values of p.

One can use the fact that the values of the function

$$x \to \frac{\prod_{l=1}^{\infty}\left(1-2x4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-x4^{-l}\right)}$$

on the interval $[0,1]$ are within 2% of the values of the linear function having the same values at the end points to approximate Formula (9) by

$$M = \frac{1}{\frac{\mu}{\beta} + R\sqrt{p+\mu}\left(\sqrt{\frac{\pi}{2}} - \frac{p}{p+\mu}(\sqrt{\frac{\pi}{2}} - \frac{1}{\alpha})\right)}, \tag{10}$$

where α is the constant defined at (8) approximately equal to 1.309.

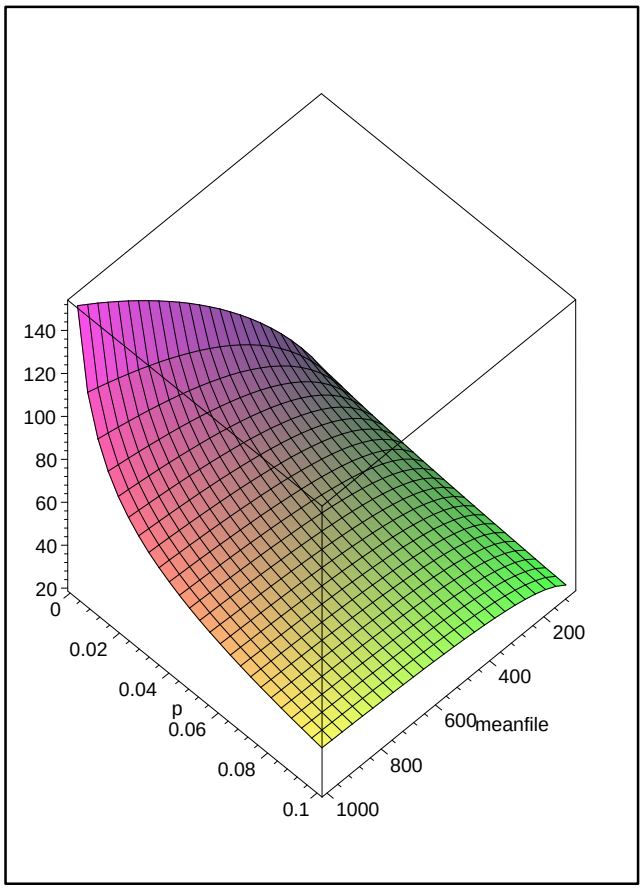

Figure 5: Dependence of Formula (9) in p and $m = 1/\mu$. Here $R = 0.1$ s. and $1/\beta = 2$ s.

Latency This special case allows for a more explicit inversion of the Laplace transform giving the latency $\lambda(x)$ to download a file of size x. Here

$$\frac{T(\mu)}{\mu} = \frac{1}{\mu^2}\sqrt{\frac{\pi}{2}}R\frac{\prod_{l=1}^{\infty}\left(1-\frac{2p}{p+\mu}4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-\frac{p}{p+\mu}4^{-l}\right)}\sqrt{p+\mu} = h(\mu)\left(\prod_{l=1}^{\infty}g_l(\mu)\right)f(\mu),$$

with

$$f(\mu) = \sqrt{\frac{\pi}{2}}R\frac{\sqrt{p+\mu}}{(p+\mu)^2}, \quad g_l(\mu) = \frac{p+\mu-2p4^{-l}}{p+\mu-p4^{-l}}, \quad h(\mu) = 1+\frac{2p}{\mu}+\frac{p^2}{\mu^2}.$$

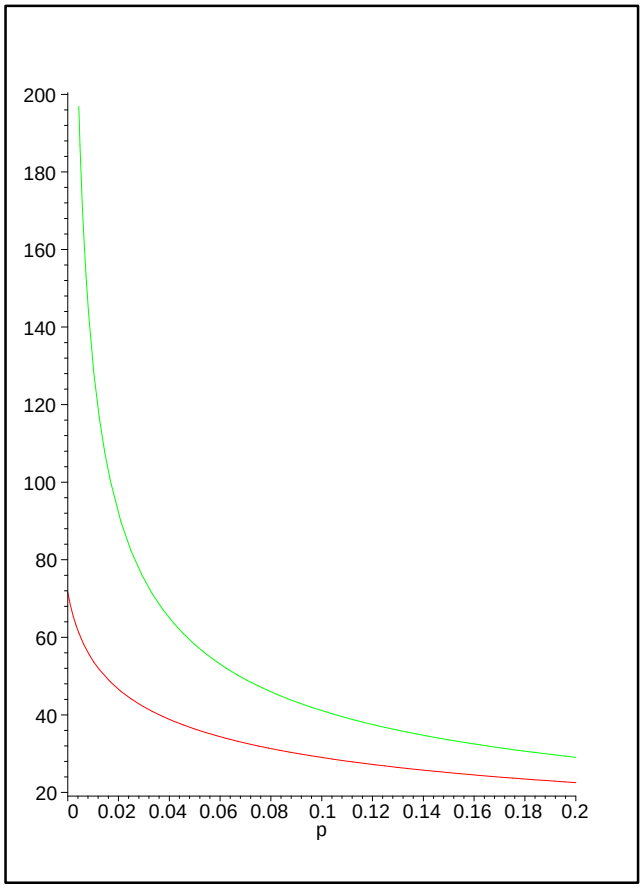

Figure 6: Comparison of Formula (9) in red and the square root formula (i.e. Formula (8)) in green. Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 1000$ pkts and p varies.

The inverse of $g_l(\mu)$ is

$$\gamma_l(x) = \delta_0(x) - p4^{-l}e^{-p(1-4^{-l})x} \equiv \delta_0(x) - p4^{-l}e_l(x),$$

with $\delta_0(x)$ the Dirac distribution at the origin, whereas that of $f(\mu)$ is

$$\phi(x) = R\sqrt{2x}e^{-px}.$$

The product $f(\mu)\prod_{l=1}^{\infty} g_l(\mu)$ in the Laplace plane, when inversed, leads to some inverse $\xi(x)$ which is a convolution of the inverses of the terms:

$$\begin{aligned}\xi(x) &= \phi(x) \\ &\quad -p\sum_{l=1}^{\infty} 4^{-l}\phi * e_l(x) \\ &\quad +p^2\sum_{l,m=1}^{\infty} 4^{-l-m}\phi * e_l * e_m(x) \\ &\quad -p^3\sum_{l,m,n=1}^{\infty} 4^{-l-m-n}\phi * e_l * e_m * e_n(x) \\ &\quad +p^4\ldots\end{aligned} \tag{11}$$

where * denotes convolution and where the multi-index sums bear on integers that are different (e.g. $l \neq m$, l, m, n all different etc.). It is easily checked that the last series is converging for all values of x and p. In what follows, we will call expansion of order k the sum of the k first terms of the last series.

Now, using the expression for $h(\mu)$, we get the following formula:

$$\lambda(x) = \xi(x) + 2p\int_0^x \xi(u)du + p^2\int_{u=0}^{x}\int_{v=0}^{u}\xi(v)dvdu. \tag{12}$$

We found no simple physical explanation for this formula. Approximations based on expansions of finite order are easily obtained via Maple. Figure 7 (obtained thanks to Maple) plots an approximation of order 3 for the $\lambda(x)$ function in the case with $R = .1$ s. and $p = 0.05$.

2.4.3 No Losses

Corollary 3 *When p is 0 and G is exponential, we still have (4) but with*

$$T = R\sqrt{\frac{\pi}{2\mu}} + \frac{\sqrt{\pi}R^2}{2}\sum_{k=0}^{\infty}\left(\widehat{h}(2k+3)\frac{\left(\frac{\mu R^2}{2}\right)^{k+\frac{1}{2}}}{(k+1)!} - \widehat{h}(2k+2)\frac{\left(\frac{\mu R^2}{2}\right)^{k}}{\Gamma(k+\frac{3}{2})}\right), \tag{13}$$

One can check that this is equivalent to

$$T = \int_{z=0}^{\infty}\int_{u=0}^{\infty} h(z)\mu e^{-\mu u}\left(\sqrt{z^2R^4 + 2uR^2} - zR^2\right)dudz,$$

which corresponds to the results of [6].

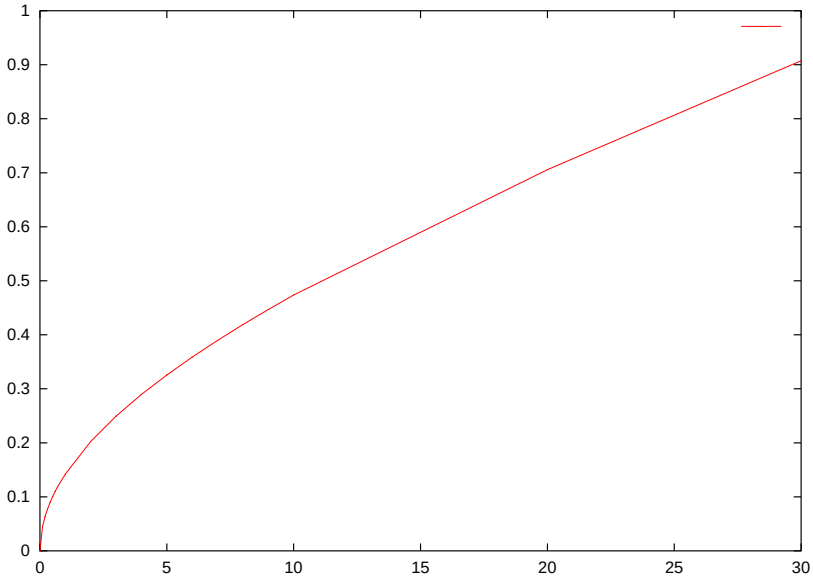

Figure 7: Expansion of order 3 for the latency of a file download in function of its size.

2.5 Heavy Tailed Case

Let us now take the distribution of the file sizes to be of the form

$$G(x) = \sum_{i=1}^{\infty} q_i (1 - e^{-\mu_i x}), \tag{14}$$

where $\{q_i\}$ is a probability and $\{\mu_i\}$ a sequence of positive real numbers such that $\sum_i q_i \mu_i^{-1} < \infty$, which guarantees that G has a finite first moment μ^{-1}.

An interesting instance is that with $q_i = Ai^{-\alpha}$ and $\mu_i = \mu/i$ with $\alpha > 2$. The associated mixture of exponentials is heavy tailed as shown by the inequality:

$$1 - G(k) \geq Ak^{-\alpha} e^{-\mu}.$$

Let

$$\Pi_k(i,u) = \prod_{l=0}^{k-1} \left(1 - \frac{p}{p+\mu_i} 2^{-u-2l}\right). \tag{15}$$

By arguments similar to those presented above, we get

Theorem 2 *Under the above assumptions, the mean rate $\widehat{s}(2)$ is still given by (4) with*

$$T = \sum_{i\geq 1} q_i \left(\frac{\sqrt{\pi}}{\mu_i} \frac{\Pi_\infty(i,1)}{\Pi_\infty(i,2)} \sqrt{\frac{(p+\mu_i)R^2}{2}} + \right.$$
$$\frac{\sqrt{\pi}R^2}{2} \sum_{k=0}^{\infty} \left(\Pi_k(i,2) \frac{\Pi_\infty(i,1)}{\Pi_\infty(i,2)} \widehat{h}(2k+3) \frac{\left(\frac{(p+\mu_i)R^2}{2}\right)^{k+\frac{1}{2}}}{(k+1)!} \right.$$
$$\left. \left. -\Pi_k(i,1)\widehat{h}(2k+2) \frac{\left(\frac{(p+\mu_i)R^2}{2}\right)^{k}}{\Gamma(k+\frac{3}{2})} \right) \right). \tag{16}$$

In the case when slow start is absent, (9) can be generalized to

$$M = \frac{1}{\frac{\mu}{\beta} + \sqrt{\frac{\pi}{2}} R \sum_i q_i \left(\frac{\prod_{l=1}^{\infty}\left(1-\frac{2p}{p+\mu_i}4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-\frac{p}{p+\mu_i}4^{-l}\right)} \sqrt{p+\mu_i} \right)}. \tag{17}$$

It turns out that these mean value formulae are almost insensitive to the heavy tailedness as shown by Figure 8 which compares the mean rate given by (17) and that given by (9) when taking the same mean values for the file size. These curves seem to overlap but this is not an exact insensitivity. For example, for $R = 0.1$ s., $1/\beta = 2$ s., $p = 0.5\%$, $q_i = Ai^{-4}$ and mean file size 200, (17) gives a mean rate of 56,699 whereas (9) gives 56.424. The difference (17)-(9) is displayed in Figure 9 in the case with $p = .5\%$.

Using the same arguments as in (10), one gets the following approximation for the rate in the heavy tailed case:

$$M = \frac{1}{\frac{\mu}{\beta} + R \sum_i q_i \sqrt{p+\mu_i} \left(\sqrt{\frac{\pi}{2}} - \frac{p}{p+\mu_i} (\sqrt{\frac{\pi}{2}} - \frac{1}{\alpha}) \right)}. \tag{18}$$

3 Distribution of the Tranmission Rate

In this section, we return to the general case with losses and slow start and we assume both the file sizes and the off periods are exponentially distributed.

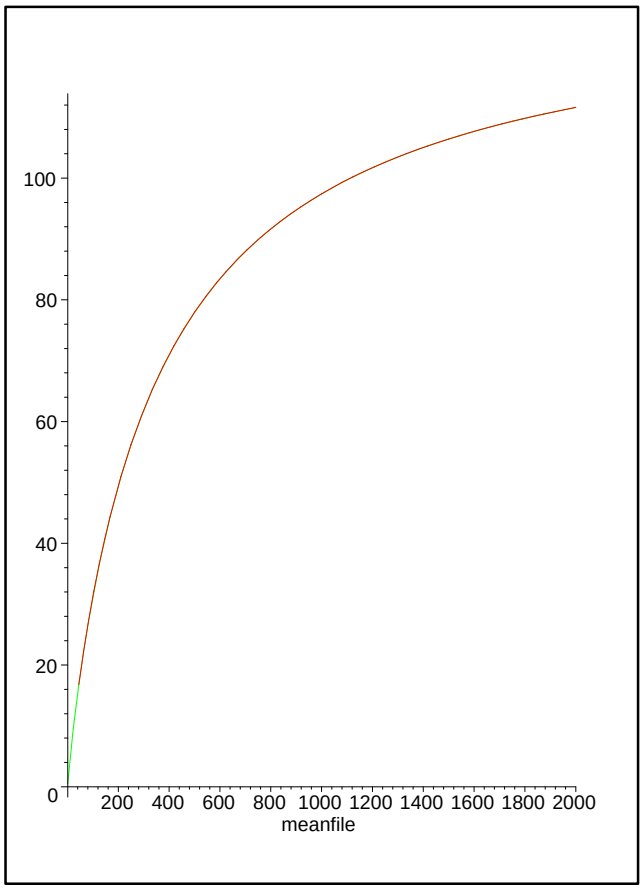

Figure 8: Comparison of (17) and (9). Here $R = 0.1$ s., $1/\beta = 2$, $p = 1\%$, $q_i = Ai^{-4}$.

3.1 Transforms and Distributions

Theorem 3 *Under the above exponential assumptions, the Mellin transform of the distribution of the transmission rates is given by the formula*

$$\begin{aligned}\widehat{s}(u) &= \widehat{s}(2)\Gamma(\frac{u}{2})\left(\frac{\Pi_\infty(u)}{\Pi_\infty(2)}\left(\frac{(p+\mu)R^2}{2}\right)^{1-\frac{u}{2}} + \right. \\ &\quad \frac{\mu R^2}{2}\sum_{k=0}^{\infty}\left(\Pi_k(2)\frac{\Pi_\infty(u)}{\Pi_\infty(2)}\frac{\widehat{h}(2k+3)}{(k+1)!}\left(\frac{(p+\mu)R^2}{2}\right)^{k+1-\frac{u}{2}}\right. \\ &\quad \left.\left. -\Pi_k(u)\frac{\widehat{h}(u+2k+1)}{\Gamma(\frac{u}{2}+k+1)}\left(\frac{(p+\mu)R^2}{2}\right)^{k}\right)\right). \qquad (19)\end{aligned}$$

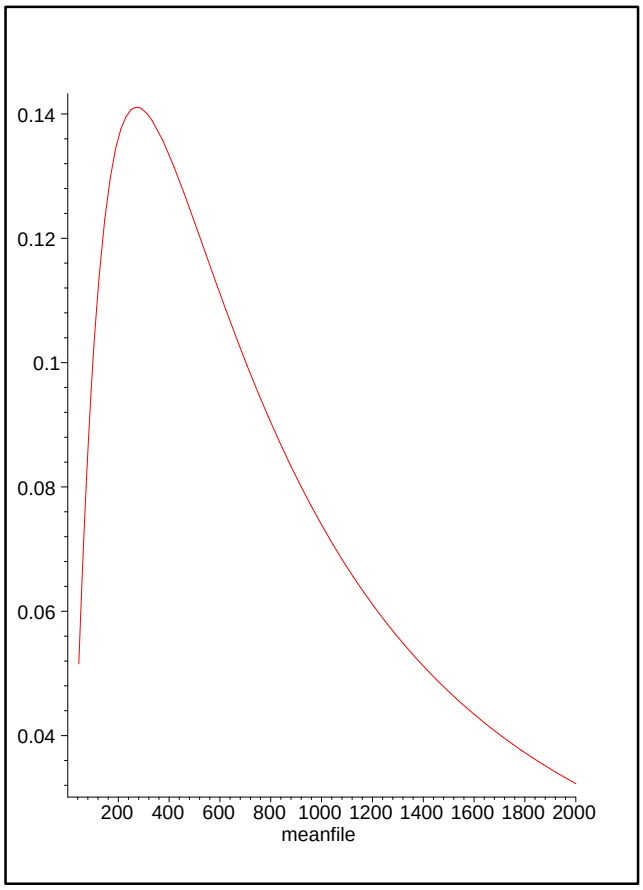

Figure 9: Difference of (17) and (9). Here $R = 0.1$ s., $1/\beta = 2$ s., $p = .5\%$, $q_i = Ai^{-4}$.

with $\widehat{s}(2) = M$ given by (4). In the case without slow start,

$$\widehat{s}(u) \quad = \quad \widehat{s}(2)\Gamma(\frac{u}{2})\frac{\Pi_\infty(u)}{\Pi_\infty(2)}\left(\frac{(p+\mu)R^2}{2}\right)^{1-\frac{u}{2}}, \tag{20}$$

with $\widehat{s}(2) = M$ given by (9). Let $\{a_n\}$ be the coefficients of the analytic expansion

$$\prod_{l=0}^{\infty}\left(1 - \frac{p}{p+\mu}2^{-2l}x\right) = \sum_n a_n x^n.$$

Then

$$s(z) = \frac{\widehat{s}(2)}{\Pi_\infty(2)}(p+\mu)R^2 \sum_{n\geq 0} a_n e^{-(\frac{(p+\mu)R^2}{2}4^n)z^2}. \tag{21}$$

The last expression is the substochastic density of the transmission rates of active flows. The relation $\nu = 1 - \widehat{s}(1)$ fully characterizes the distribution of the rates, which is composed of a mass at the origin and of an infinite (signed) mixture of truncated Gaussian laws.

Figures 10 and 11 plot the density $s(z)$ as given by Formula (21).

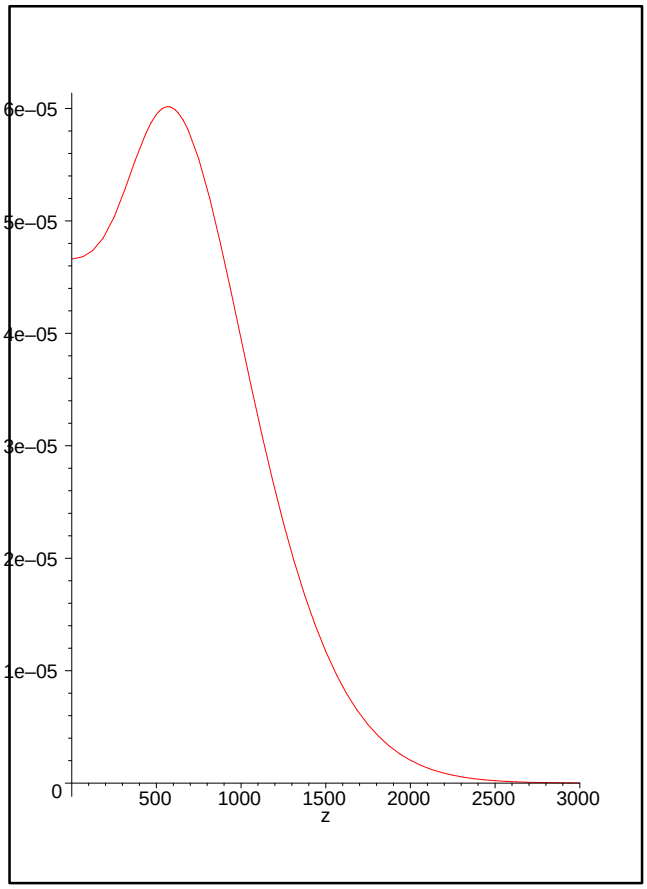

Figure 10: The density of the stationary rate (21). Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 100$ and $p = 1\%$.

From this expression, it is easy to derive the following result:

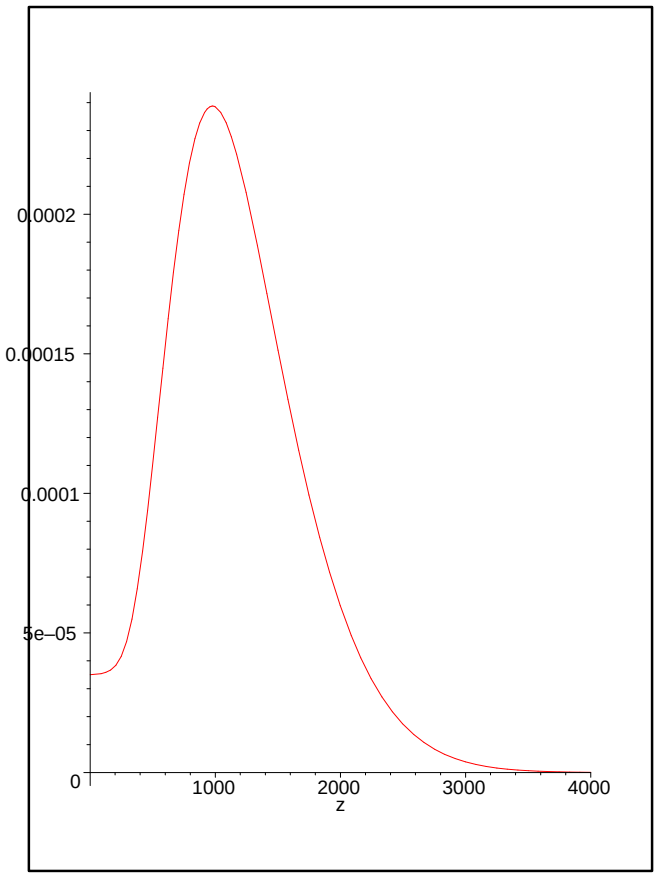

Figure 11: The density of the stationary rate (21). Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 1000$ and $p = 1\%$.

Corollary 4 *In the no slow start case, the probability that a flow gets a rate of at least X in steady state is*

$$\overline{S}(X) = \frac{\widehat{s}(2)}{\Pi_\infty(2)}(p+\mu)R^2 \sum_{n\geq 0} a_n \int_X^\infty e^{-(\frac{(p+\mu)R^2}{2}4^n)z^2} dz, \tag{22}$$

and the probability that a flow gets a rate of at least X given that it is active is $\overline{S}(X)/\widehat{s}(1)$.

The integral in the last formula can in turn easily be expressed in terms of the ERF function of Gaussian calculus.

3.2 The Mean Rate of a Flow at the End of a File Transfer

To illustrate the flexibility that the knowledge of distributions gives us, let us consider the estimation of the mean rate D of a flow just before the time when it switches from ON to OFF. The rate conservation principle, which equates the stationary rate of decrease and the stationary rate of increase of the flow (see [6]) gives

$$\frac{p\widehat{s}(3)}{2} + \frac{1}{T+1/\beta}D = \frac{\widehat{s}(1)}{R^2} + \frac{1}{T+1/\beta}\widehat{h}(2), \tag{23}$$

since the flow decreases by D every period of length $T + 1/\beta$ due to the completion of files and by

$$\int_0^\infty pzs(z)\frac{z}{2} = \frac{p\widehat{s}(3)}{2},$$

per unit of time due to multiplicative decrease, and it increases by $\widehat{h}(2)$ every period of length $T + 1/\beta$ due to slow start and by $(1-\nu)/R^2$ due to additive increase.

Equation (23) determines D since all other terms are known once distributions are known.

4 AQM Bandwidth Sharing

The setting of the present paper is adequate to represent certain AQM schemes that stabilize the flows through losses at a congested buffer. For instance RED, [8], is based on dropping packets with a probability proportional to the queue length. If RED stabilizes then there is essentially a constant drop probability. Another example of AQM scheme is DRED, [4], which is specifically designed to find a constant drop probability consistent with a target queue size. Other instances of TCP controlled networks with a constant drop probability are ADSL networks and certain wireless networks where there is an unavoidable and essentially random bit error rate at the physical layer that translates into random packet losses even when no congestion occurs.

The results of this application section are presented using the exponential model. There is no difficulty extending them to the heavy tailed setting of §2.5.

4.1 The two Regimes

Consider n statistically identical ON-OFF flows with parameters (μ, β, h, R) that share a common AQM link with capacity $C \cdot n$. Let

$$\rho \quad = \quad \frac{1/\mu}{\frac{1}{\beta}+T} \tag{24}$$

with

$$T = \int_{z=0}^{\infty}\int_{u=0}^{\infty} h(z)\mu e^{-\mu u}\left(\sqrt{z^2R^4+2uR^2} - zR^2\right)dudz$$

denote the mean value of the rate obtained by one flow in the absence of packet loss (see (13) and [6] for the derivation of this formula). In the case without slow start,

$$\rho = (\mu(\beta^{-1} + R\sqrt{\pi/(2\mu)}))^{-1}.$$

Assume n is large. Then using the mean-field approach of [5, 6], one can show that there are at least two possible regimes:

- The *stabilized congestion regime*, which is reached by the system whenever $\rho > C$ and where a positive drop probability is required to match the load brought by the flows and the capacity of the link. In this regime, the AQM scheme stabilizes to a constant buffer content b, to a constant packet loss probability p and to a mean rate per flow of $M[p]$, such that

$$M[p](1-p) = C \tag{25}$$

  with $M[p]$ the function of p defined in (4). Since the function $p \to M[p](1-p)$ is decreasing in p and tends to ρ when p tends to 0, the above equation defines a unique equilibrium point p^* whenever $\rho > C$.

- *The congestion-less regime*, which is reached when $\rho < C$, where the load brought by the flows is less than the link rate, and where each flow gets a mean rate of ρ.

Other and in particular oscillating regimes are also possible as suggested by the results of many other authors (see e.g. [5]).

4.2 Mice and Elephants

Assume that two classes of flows, share a common AQM link of capacity $C \cdot n$ and of drop probability function K.

There are $r_i.n$ flows of type i with RTT R_i, think time with mean β_i^{-1} and file size with mean μ_i^{-1}, $i = 1, 2$.

Again there are two regimes depending on whether the mean load without loss $r_1\rho_1 + r_2\rho_2$ with ρ_i defined as in (24) is larger or smaller than C.

If the mean load without loss is larger than C, the AQM algorithm stabilizes to a positive packet killing rate p (see [5, 6]), and the stationary rate of each class $\widehat{s}_i(2)$ is given by (4):

$$\widehat{s}_i(2) = \frac{1}{\frac{\mu_i}{\beta_i} + \sqrt{\frac{\pi}{2}} R_i \frac{\prod_{l=1}^{\infty}\left(1-\frac{2p}{p+\mu_i}4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-\frac{p}{p+\mu_i}4^{-l}\right)}\sqrt{p+\mu_i}}. \tag{26}$$

The rate of arrival of packets of class i to the link is

$$\int_0^\infty s_i(z) z (1-p) dz = (1-p)\widehat{s}_i(2)$$

so that one should have

$$(1-p)(r_1\widehat{s}_1(2)+r_2\widehat{s}_2(2))=C \tag{27}$$

at equilibrium. The LHS of the last equation is a decreasing function of p taking the value ∞ at $p=0$. Hence there is a unique p that solves it, say p^*, so that the AQM bandwidth sharing is that given by (26) with p replaced by p^*.

The stationary buffer content b of the RED queue is then that obtained from the RED function $K(b)$ by solving the equation $K(b)=p^*$.

Here is an example with two classes of flows: the elephant class with mean file size 1000 packets and the class with file m, where m is smaller than 1000, which can be called the mouse class when m is significantly smaller than 1000. In the model, 50 elephant flows and 50 mice flows compete for the bandwidth of a 2000 pkts/s link. The RTT is .1 s. and the think times have a 2 s. mean. There is no slow-start. Figure 12 shows the rate obtained by a typical elephant and a typical mouse when m varies, where the rate is here the emission rate (which counts the lost packets).

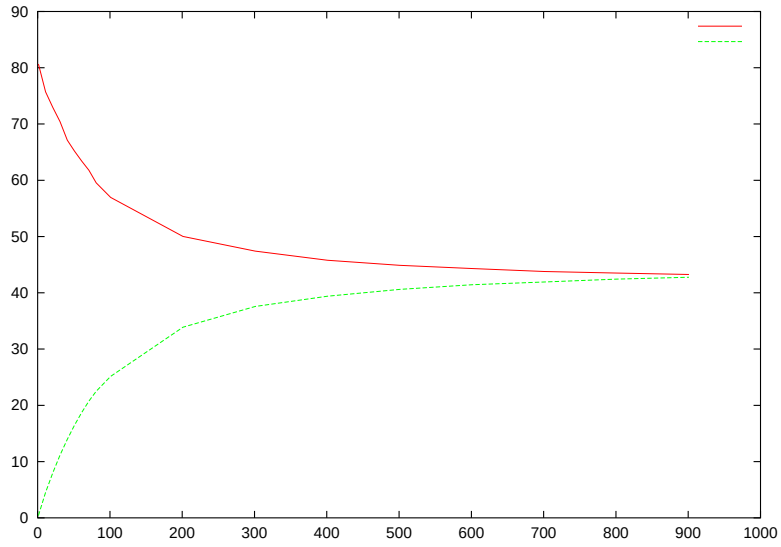

Figure 12: Mice vs. elephants: mean bandwidth sharing predicted by (27)

4.3 Long and Short RTTs

The setting is the same as in the last subsection, but now class 1 and 2 only differ in their RTT. Figure 13 illustrates the bandwidth sharing between 50 flows of class 1, with a RTT of 100 ms, and 50 flows of class 2, with a RTT that we let vary from 1 to 200 ms. The other parameters of the flows are the

same: mean think time of 2 s., mean file size of 1000 packets. Bandwidth sharing is predicted using Formula (27).

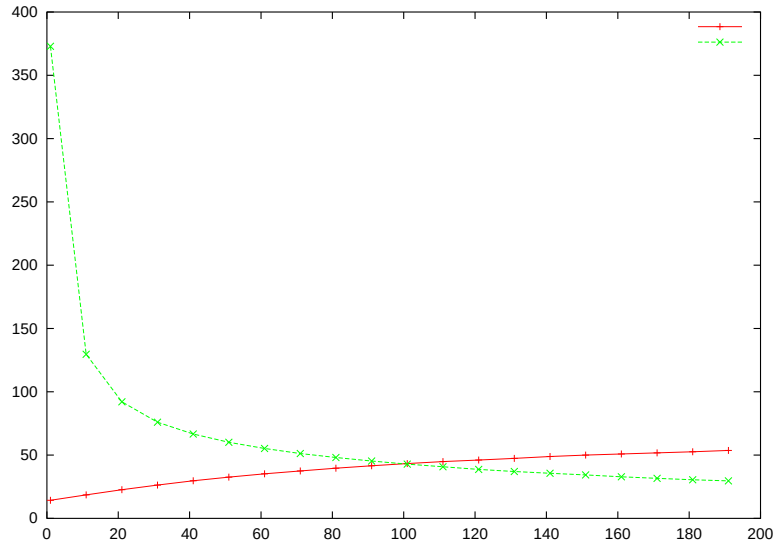

Figure 13: Short vs. long RTTs: mean bandwidth sharing predicted by (27)

More generally, consider two non-persistent flows with the same statistical properties and subject to the same packet loss probability p, and which only differ via their RTTs R_1 and R_2. It follows from (9) that the ratio of their mean rates m_1 and m_2 is of the form

$$\frac{m_1}{m_2} = \frac{a + bR_2}{a + bR_1} \tag{28}$$

with

$$a = \frac{\mu}{\beta}, \qquad b = \sqrt{\frac{\pi}{2}} \frac{\prod_{l=1}^{\infty} \left(1 - \frac{2p}{p+\mu} 4^{-l}\right)}{\prod_{l=1}^{\infty} \left(1 - \frac{p}{p+\mu} 4^{-l}\right)} \sqrt{p+\mu}.$$

5 Slow Start Model

5.1 Setting

We propose to focus on HTTP 1.1 where the files successively downloaded by a flow use the same TCP connection. This means that the successive downloads of this user are made from the same server and that the Keepalive Timer (usually 15 s.) does not expire (for the last point, see [1]).

IETF RFC 2581, see [2], stipulates the following rules for TCP:

- When the TCP connection is idle for more than one retransmission timeout (RTO, roughly a few RTTs), CWND is reduced to IW (initial window), which we will assume to correspond to decreasing the rate to 0.
- SSTHRESH is however kept to save information on the previous value of the congestion window. We propose here to take SSTHRESH equal to the mean value of the congestion window just after loss epochs.

This will allow us to estimate the law H of SSTHRESH as we shall see below.

5.2 Slow Start Jump Distribution

According to the principles stated above, we propose to take $H = \delta(\eta)$ with η equal to the expected value of the throughput just after a loss; i.e.

$$\eta = \frac{\int_0^\infty pzs(z)\frac{z}{2}dz}{\int_0^\infty pzs(z)dz} = \frac{\widehat{s}(3)}{2\widehat{s}(2)}. \tag{29}$$

Of course this is not constructive as η determines $\widehat{s}(u)$ (say through (19)) which in turn determines η via (29). To break this loop, we propose the following iterative scheme with $H = \delta_{\eta_1}$ with $\eta_1 = 0$:

1. $\widehat{s}_n(2)$ is obtained from (4) and $\widehat{s}_n(3)$ from (19), both with $H = \delta_{\eta_n}$;
2. η_{n+1} is defined by
$$\eta_{n+1} = \frac{\widehat{s}_n(3)}{2\widehat{s}_n(2)}.$$

The convergence of this scheme was tested using Maple. The scenario with exponential file sizes with mean value 300 pkts, exponential think times with mean 2 s. and with $R = .1$ s. and $p = .05$ leads to a mean rate $M = 40.3166$. The value given by the last procedure is $\eta = 32.59349528$ pkts/s. This fixed point is reached in less than 10 iterations.

6 Appendix

6.1 Proofs of the Results of §2

Consider first the case when the think times are exponential too. Let $\nu(t)$ be the probability that the HTTP flow is idle at time t. Let $s(z,t)dz$ be the probability that the transmission rate of the flow is in $[z, z+dz]$.

Using the same approach as in [5, 6] it is easy to see that $s(z,t)$ satisfies the partial differential equation

$$\begin{aligned}\frac{\partial s}{\partial t}(z,t) &= -\frac{1}{R^2}\frac{\partial s}{\partial z} - \mu z s(z,t) + \beta\nu(t)h(z) \\ &+ 4zp(t-R)s(2z,t) - zp(t-R)s(z,t),\end{aligned} \tag{30}$$

where h is the density of the slow start jump. In addition

$$\frac{d\nu}{dt}(t) = \int_0^\infty \mu z s(z,t) dz - \beta\nu(t). \tag{31}$$

and

$$\int_0^\infty s(z,t)dz = 1 - \nu(t). \tag{32}$$

We consider a steady state regime. The steady state density s satisfies the ODE

$$\begin{aligned} \frac{ds}{dz}(z) &= -\mu R^2 z s(z) + \beta\nu R^2 h(z) + 4zpR^2 s(2z) - zpR^2 s(z) \\ &= -\mu R^2 z s(z) + \mu R^2 \int_0^\infty \mu v s(v) dv h(z) + 4zpR^2 s(2z) - zpR^2 s(z) \end{aligned} \tag{33}$$

since

$$\int_0^\infty \mu z s(z) dz = \beta\nu. \tag{34}$$

Let $\widehat{s}(u)$ be the Mellin transform of $s(z)$. Multiplying both sides of (33) by z^u and integrating w.r.t. z, we get

$$u\widehat{s}(u) = pR^2\widehat{s}(u+2)\left(1 - 2^{-u}\right) + \mu R^2\widehat{s}(u+2) - \mu R^2\widehat{s}(2)\widehat{h}(u+1).$$

Let

$$\widehat{s}(u) = f(u)\Gamma(\frac{u}{2})\left(\frac{2}{(p+\mu)R^2}\right)^{\frac{u}{2}}.$$

Then

$$f(u) = f(u+2)(1 - \frac{p}{p+\mu}2^{-u}) - \frac{\mu R^2}{2}\widehat{s}(2)\widehat{h}(u+1)\frac{\left(\frac{(p+\mu)R^2}{2}\right)^{\frac{u}{2}}}{\Gamma(\frac{u}{2}+1)} \tag{35}$$

which implies that

$$f(u) = f(\infty)\Pi_\infty(u) - \frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(u)\widehat{h}(u+2k+1)\frac{\left((p+\mu)R^2/2\right)^{\frac{u}{2}+k}}{\Gamma(\frac{u}{2}+k+1)} \tag{36}$$

so that

$$\begin{aligned}\widehat{s}(u) &= f(\infty)\frac{\Gamma(\frac{u}{2})}{\left(\frac{(p+\mu)R^2}{2}\right)^{\frac{u}{2}}}\Pi_\infty(u) \\ &\quad -\frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(u)\widehat{h}(u+2k+1)\left(\frac{(p+\mu)R^2}{2}\right)^k\frac{\Gamma(\frac{u}{2})}{\Gamma(\frac{u}{2}+k+1)}. \qquad (37)\end{aligned}$$

Specializing (37) to $u=2$, we get

$$\begin{aligned}\widehat{s}(2) &= f(\infty)\left(\frac{2}{(p+\mu)R^2}\right)\Pi_\infty(2) \\ &\quad -\frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(2)\widehat{h}(2k+3)\left(\frac{(p+\mu)R^2}{2}\right)^k\frac{1}{(k+1)!}. \qquad (38)\end{aligned}$$

From (32), we also have

$$1=\widehat{s}(1)+\nu$$

which together with $\beta\nu=\mu\widehat{s}(2)$ implies

$$\widehat{s}(1)=1-\mu\widehat{s}(2)/\beta.$$

Hence, specializing (37) to $u=1$, we get

$$\begin{aligned}1-\mu\widehat{s}(2)/\beta &= f(\infty)\sqrt{\frac{2\pi}{(p+\mu)R^2}}\Pi_\infty(1) \\ &\quad -\frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(1)\widehat{h}(2k+2)\left(\frac{(p+\mu)R^2}{2}\right)^k\frac{\sqrt{\pi}}{\Gamma(k+\frac{3}{2})}. \qquad (39)\end{aligned}$$

When multiplying (38) by $\pi\Pi_\infty(1)$ and subtracting (39) multiplied by $\sqrt{\frac{2\pi}{(p+\mu)R^2}}\Pi_\infty(2)$, we get

$$\begin{aligned}&\widehat{s}(2)\pi\Pi_\infty(1)-\sqrt{\frac{2\pi}{(p+\mu)R^2}}\Pi_\infty(2)(1-\mu\widehat{s}(2)/\beta) \\ &\quad = -\pi\Pi_\infty(1)\frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(2)\widehat{h}(2k+3)\frac{\left(\frac{(p+\mu)R^2}{2}\right)^k}{(k+1)!} \\ &\quad\quad +\pi\Pi_\infty(2)\frac{\mu R^2\widehat{s}(2)}{2}\sum_{k=0}^{\infty}\Pi_k(1)\widehat{h}(2k+2)\frac{\left(\frac{(p+\mu)R^2}{2}\right)^{k-\frac{1}{2}}}{\Gamma(k+\frac{3}{2})}.\end{aligned}$$

Solve for $\widehat{s}(2)$ to get (4).

A regenerative theory argument allows one to identify the mean value T of a file transfer as given in (3) in the expression for the mean value of the stationary rate given above. Since the mean time to transfer is the same whatever the think time, the formulas that we have established in this particular exponential case are all insensitive to the distribution of the think times and are all valid as long as the mean think time is equal to β^{-1}.

6.2 Proofs of the Results of §3

From the expressions of the last section, we get that

$$\begin{aligned}\widehat{s}(u) &= \widehat{s}(2)\left(A\frac{\Gamma(\frac{u}{2})}{\left(\frac{(p+\mu)R^2}{2}\right)^{\frac{u}{2}}}\Pi_\infty(u) \right.\\ &\left. -\frac{\mu R^2}{2}\sum_{k=0}^{\infty}\Pi_k(u)\widehat{h}(u+2k+1)\left(\frac{(p+\mu)R^2}{2}\right)^k\frac{\Gamma(\frac{u}{2})}{\Gamma(\frac{u}{2}+k+1)}\right)\end{aligned}$$

with $\widehat{s}(2)$ given by (4) and with

$$A = \frac{(p+\mu)R^2}{2}\frac{1}{\Pi_\infty(2)} + \frac{\mu R^2}{2}\sum_{k=0}^{\infty}\frac{\Pi_k(2)}{\Pi_\infty(2)}\frac{\widehat{h}(2k+3)}{(k+1)!}\left(\frac{(p+\mu)R^2}{2}\right)^{k+1}$$

which gives (19).

In the case with no slow start,

$$\begin{aligned}\widehat{s}(u) &= \frac{\widehat{s}(2)}{\Pi_\infty(2)}\frac{(p+\mu)R^2}{2}\Gamma(\frac{u}{2})\Pi_\infty(u)\left(\frac{(p+\mu)R^2}{2}\right)^{-\frac{u}{2}}\\ &= \frac{\widehat{s}(2)}{\Pi_\infty(2)}\frac{(p+\mu)R^2}{2}\Gamma(\frac{u}{2})\sum_{n=0}^{\infty}a_n\left(2^{-u}\right)^n\left(\frac{(p+\mu)R^2}{2}\right)^{-\frac{u}{2}}\\ &= \frac{\widehat{s}(2)}{\Pi_\infty(2)}\frac{(p+\mu)R^2}{2}\sum_{n=0}^{\infty}a_n\Gamma(\frac{u}{2})\left(4^n\frac{(p+\mu)R^2}{2}\right)^{-\frac{u}{2}}. \qquad (40)\end{aligned}$$

Using now the relation

$$2\int_0^\infty e^{-az^2}z^{u-1}dz = \int_0^\infty e^{-at}t^{u/2-1}dt = a^{-u/2}\Gamma(u/2)$$

we obtain that the density $s(z)$ with Mellin transform given in (20) is that given in (21).

References

[1] `http://lists.w3.org/Archives/Public/ietf-http-wg-old/2000SepDec/0078.html`

[2] `http://www.faqs.org/rfcs/rfc2581.html`

[3] ADJIH, C., JACQUET, P., VVEDENSKAYA, N. (2001) Performance evaluation of a single queue under multi-user TCP/IP connections. *INRIA Research report #4141.*

[4] AWEYA J., OUELLETTE, M., MONTUNO, D.Y. (2002) DRED: a random early detection algorithm for TCP/IP networks *International Journal of Communication Systems,* **15***, p. 287 - 307.*

[5] BACCELLI, F., MCDONALD, D. R., REYNIER, J. (2002) A mean-field model for multiple TCP connections through a buffer implementing RED. *Performance Evaluation Vol. 11, (2002) pp. 77-97.* Elsevier Science.

[6] BACCELLI, F., CHAINTREAU, A., DE VLEESCHAUWER, D., MCDONALD, D. (2004). A mean-field analysis of short lived interacting TCP flows. *In Proceedings of ACM Sigmetrics'04, June 2004, New York.*

[7] CHANG, C.-S., LIU, Z. (2002) A Bandwidth Sharing Theory for a Large Number of HTTP-like Connections. *In Proceedings of the IEEE Infocom 2002 Conference, New York, June 2002.*

[8] FLOYD, S., JACOBSON, V. (1993) Random early detection gateways for congestion avoidance *IEEE/ACM Trans. Networking, No.4, 397-413.*

[9] FRED, S., BONALD, T., PROUTIÈRE, A., RÉGNIÉ, G., ROBERTS, J. (2001) Statistical bandwidth sharing: a study of congestion at flow level. *in Proceedings of ACM SIGCOMM 2001: pp111-122*

[10] HEYMAN, D., LAKSHMAN, T., NEIDHARDT, A. (1997) A new method for analyzing feedback-based protocols with applications to engineering web traffic over the Internet. *in ACM Sigmetrics, pp. 24–38.*

[11] KHERANI, A.A., KUMAR, A. (2000) Performance Analysis of TCP with Non-persistent Sessions. *Workshop on Modeling of Flow and Congestion Control, INRIA, Ecole Normale Supérieure, Paris, September 4-6, 2000.*

[12] KNUTH, D. (1997) The Art of Computer Programming, Addison Wesley.

[13] OTT, T., KEMPERMAN, J.H.B., MATHIS, M. (1992) The Stationary Behavior of Ideal TCP Congestion Avoidance. *Internetworking: Research and Experience, V.11, p.115-156.*

[14] ROBERTS, J., MASSOULIÉ, L. (1998) Bandwidth sharing and admission control for elastic traffic. *ITC Specialist Seminar, Yokohama, October.*

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
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 - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

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