Performance of Multiple TCP Flows: an Analytical Approach

Cédric Adjih, Philippe Jacquet, Georgios Rodolakis, Nikita Vvendenskaya

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Performance of Multiple TCP Flows: an Analytical Approach

Cédric Adjih — Philippe Jacquet — Georgios Rodolakis — Nikita Vvendenskaya

N° 5417

Décembre 2004

Thème 1

ISSN 0249-6399 ISRN INRIA/RR--5417--FR+ENG

INRIA
ROCQUENCOURT

Performance of Multiple TCP Flows: an Analytical Approach

Cédric Adjih, Philippe Jacquet, Georgios Rodolakis, Nikita Vvendenskaya*

Thème 1 — Réseaux et systèmes
Projet Hipercom

Rapport de recherche n° 5417 — Décembre 2004 — 24 pages

Abstract: We study the performance of several TCP connections subjected to the bottleneck of a slow network, accessed via a single queue with high but finite capacity. Using the mean-field methodology and fluid approximation, we establish asymptotic results on the queue length distribution and the window size distribution. We prove that the difference between the buffer capacity and the actual queue length tends to be exponentially distributed. Based on these results, we analyze the autocorrelation function of a single TCP connection, and we show that several TCP connections with heavy tailed round trip delays generate traffic with long term dependencies.

Key-words: TCP, asymptotics, mean-field, long term dependencies, heavy tails

* Partially supported by RFFI grant 02-01-00068

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

Performance de plusieurs flots TCP : une approche analytique

Résumé : Nous étudions les performances de plusieurs connexions TCP soumises au goulot d'étranglement d'un réseau lent desservi par une file d'attente de grande capacité. En utilisant la méthode asymptotique du champ moyen nous établissons des résultats asymptotiques sur la distribution de la longueur de la file d'attente et des tailles de fenêtre quand le nombre d'utilisateurs croit proportionnellement à la capacité de la file d'attente. Nous montrons que la quantité de places libres dans le buffer d'attente suit une loi exponentielle. Nous évaluons l'autocorrélation d'une connexion TCP et nous montrons que plusieurs connexions peuvent générer de longues dépendances quand la distribution des délais a une queue lourde.

Mots-clés : TCP, comportement asymptotique, champ moyen, longues dépendances, queues lourdes

1 Introduction

The dynamic window based protocol TCP [1] is widely used over the Internet. The reason of the success of TCP is mainly based on its high dynamic that makes it able to adapt itself to any kind of network capacity, from a few bits to several gigabits per second.

TCP has received special attention since its formulation as an internet standard. In particular, its performance has been the research focus of several generations of researchers. The pioneer paper [2] analyzed the performance of TCP in the specific case of a single TCP connection on a single router connected to an infinite capacity channel with a fixed independent error rate p. More recently, Baccelli et al. [3] analyzed a single TCP connection flow travelling through a sequence of finite capacity routers. The analysis is interesting because it relies on an explicit formulation of the problem in (max,plus) algebra. The problem of several connections, even to a single router, has only been simulated, with the notable exception of [4], which is, however, limited to very specific phases of the protocol.

The aim of this paper is to investigate via analytic methods the multi-connection case where N TCP connections coexist on the same finite capacity router, connected to a finite capacity network. Extending the already difficult single connection case to the multi-connection case seems to be out of reach of the present multi-queueing toolbox for an exact analytic performance expression. However, analysis is made possible because we consider the asymptotic case where N is large. Indeed, the calculation of the steady states turns out to be much simpler compared to the single connection case.

Practically, we investigate the case where N users access N servers under TCP/IP, and the bottleneck is a router with a finite buffer and a slow network interface. The network is divided into two areas:

- a local loop with relatively low speed (telephone line, cable TV, ADSL)
- a backbone with high throughput (ATM, DWDM)

The users are located on the local loop and the servers are located on the fast backbone. Furthermore, there is a router at the border between the local loop and the fast backbone.

We assume that all users are downloading files of infinite size. We are interested in analyzing the steady state of every connection. We also assume that the round trip delay between server and user is large.

Our primary works on the topic are in [5, 6]. The remainder of the paper is organized as follows.

In Section 2, we give a short presentation of TCP and we describe the continuous model, in which the window and buffer sizes are real numbers.

Section 3 is devoted to the analysis of the continuous model, with interpretation in the case of large round trip delays. In particular, we show that the remaining space in the buffer tends to be exponentially distributed with Poisson parameter a, when $N \to \infty$. We denote T the buffer service time corresponding to each connection. When T and N tend both to infinity, the window size distribution rescaled by $T^{-1/2}$ tends to a continuous distribution

$g(y)$, given by Figure 1. In [7], the same limiting function has been extracted in the more general approach of TCP control with side control traffic as in the RED protocol.

In Section **??**, we present a discrete model of multiple TCP connections, which follows the continuous model. We then use the fact that the remaining space in the buffer tends to be exponentially distributed in order to derive a simplified model of a single TCP connection with fixed packet loss probability. In particular, we characterize the autocorrelation function of a single TCP connection via an estimate of the second eigenvalue of the Markov transition matrix.

In Section 5, we investigate the autocorrelation function of several independent TCP connections with heavy tailed round-trip delays. We argue that heavy tailed round trip delays result in heavy tailed traffic autocorrelation.

At the end of each section we present a comparison between analytic and simulated results.

2 The TCP connection protocol and its model

2.1 TCP overview

In TCP, packets are transmitted in sequence and must be acknowledged by the end-user. A packet is considered to be lost when the acknowledgement does not arrive within the estimated round trip delay. A packet loss is considered as a congestion event.

In order to cope with the round trip delay, several packets are transmitted in advance without waiting for acknowledgement. The set of unacknowledged packets is called the window and its size varies in order to handle congestion.

(i) when no packet is lost in a window (successful window), the next window size is incremented by 1 unit: $W \leftarrow W + 1$.

(ii) when a packet loss occurs (failed window), packet retransmission starts from this packet, but with a window size halved: $W \leftarrow \lfloor W/2 \rfloor$.

In the detailed TCP specification, there is also the description of self-clocking and slow start. The latter feature makes TCP more reactive in network condition changes.

Self-clocking The server synchronizes the transmission of its new packets with the acknowledgement returns, leading to a *self-clocking* of packet transmissions. Packets can be acknowledged in batch.

Slow start In the *slow start* phase the window size doubles at each full successful window, until it reaches a predefined threshold.

2.2 The continuous models

For the sake of analytic derivations, we simplify the TCP model. We denote B the finite capacity of the router, and T the buffer service time corresponding to each connection.

2.2.1 The continuous model with batch transmission

We call $R(t)$ the current available room in the buffer at time t. We assume that the buffer contains continuous data (fluid approximation), *i.e.*, $R(t)$ is a positive real number. The buffer service rate is $\mu > 0$.

The current window size at server i is denoted $W_i(t)$, which is also a real number. We consider the following simplified mode of operation:

- The server transmits all packets in the window at the same time;
- The backbone has an infinite speed and capacity, and all packets in the window arrive in batch in the buffer;
- The end-user sends one acknowledgement to the server for all packets received correctly in the batch;
- There is no slow start phase.

Consequently, after window transmission of server i:

(i) If the current window size $W_i(t) \leq R(t)$, then $R(t)$ becomes $R(t) - W_i(t)$ and $W_i(t)$ becomes $W_i(t) + 1$;

(ii) else, if $W_i(t) > R(t)$, then $R(t)$ becomes 0, and $W_i(t)$ becomes $W_i(t)/2$.

2.2.2 The exponential round trip delay

We make a further approximation by assuming that the round trip delays are random and i.i.d, following an exponential distribution of mean N/λ. Since the servers transmit a window after each round trip delay, the window batches arrive to the buffer according to a Poisson process of rate $\lambda > 0$. Due to the memoryless property of the Poisson process, the system state at time t can be fully characterized by its repartition function $W(y,t)$:

$$W(y,t) = (\text{number of servers at time } t \text{ with window size} \geq y) \times \frac{1}{N}$$

with $W(0,t) = 1$.

The exponential round trip model is convenient for a quick analysis but it is highly unrealistic. The propagation delay contains the buffer delay experienced by the last packet of the window, *i.e.,* exactly $\frac{B-R(t)}{\mu}$, which is not expected to vary much. In fact, it will be proven that the buffer delay is close to $\frac{B}{\mu}$.

2.2.3 A realistic model with fixed delay plus random processing time

In this model we assume that the propagation delay has two components:

1. A fixed buffer delay $NT = \frac{B}{\mu}$;
2. A random exponential delay of mean $\frac{N}{\lambda}$, assumed to be much smaller than NT.

For the convenience of the presentation, we call the small exponential delay the *processing time*, but it can be just a component of the propagation delay, for example some buffer time in the high speed part of the network.

In this model, the servers are either transmitting, in fixed propagation delay or in random processing delay. We keep the repartition function $W(y,t)$ but with a different meaning:

$$W(y,t) = (\text{number of servers in processing at time } t \text{ with window size} \geq y) \times \frac{1}{N}$$

In this case $W(0,t) < 1$. Indeed, quantity $W(0,t)$ is equal to the proportion of servers in processing at time t. The average value of $W(0,t)$ with respect to t is equal to $\frac{1}{\lambda T+1}$.

This model is much more realistic than the previous one and can be treated as well. However, we will first handle the unrealistic model, which will give the foundations of our framework.

3 Analysis of the continuous TCP model

In this section, we present our analysis of the continuous TCP model in the two cases of exponential round trip delay case and fixed delay plus exponential processing time. We present the mean field approach that justifies the asymptotic derivations. We then discuss the system's behavior in the case of large round trip delay.

3.1 Notations and system description

We denote $R(x,t) = P(R(t) > x)$ and $w(y,t) = -\frac{\partial}{\partial y}W(y,t)$, *i.e.,* the density function of window sizes. In other words, $w(y,t) = \frac{1}{N}\sum_{i=1}^{N}\delta(y - W_i(t))$, where δ is the Dirac function. In the model with processing time, quantity $w(y,t)$ is equal to the window size density of servers that are in processing state at time t. We shall immediately outline two important points:

1. The quantity $R(x,t)$ addresses a probability distribution.
2. The quantity $W(y,t)$ addresses a state function of the system and *a priori* is *not* a probability distribution.

Therefore, quantities $R(x,t)$ and $W(y,t)$ are not sufficient to describe the probabilistic behavior of the system. The complete probabilistic description of the system should be given by function $\rho(x,f,t) = P(R(t) > x, W(y,t) = f(y))$, where f is a positive function.

3.2 Fixed window distribution

At first, we consider the auxiliary system where the distribution of W is fixed and does not change with $R(t)$. This is not the real system since $R(t)$ and $W(t)$ are actually dependent. Therefore, in this auxiliary system, we denote $\tilde{R}(t)$ and $\tilde{w}$ the available room in the buffer and the fixed window size density, respectively. We call this model the Fixed Window Distribution (FWD) model.

Lemma 1 *In the FWD model the functional equation of* $\tilde{R}(x,t) = P(\tilde{R}(t) > x)$ *is*

$$\frac{\partial \tilde{R}(x,t)}{dt} = -\mu \frac{\partial \tilde{R}(x,t)}{\partial x} - \lambda \tilde{R}(x,t) + \lambda \int_0^\infty \tilde{R}(x+y,t)\tilde{w}(y,t)dy, \tag{1}$$

Proof: In absence of window transmissions, quantity $R(t)$ increases at rate μ. Window batches are received at rate λ. When a window of size y (which occurs according to density function $\tilde{w}(y)$) is received between t and $t+dt$, two cases are possible:

1. $\tilde{R}(t) > y$, then $\tilde{R}(t+dt) = \tilde{R}(t) - y$.
2. $\tilde{R}(t) \leq y$, then $\tilde{R}(t+dt) = 0$.

Notice that, in the above, we implicitly assume that $\tilde{R}(t)$ is not bounded, *i.e.,* the buffer is never empty. Translating formally into stochastic equations leads to (1). ∎

We assume now that that $E[e^{-s\tilde{W}}] < \infty$ for some $s > 0$.

Corollary 1 *In the FWD model, let us assume that* $E[\tilde{W}] > \mu/\lambda$. *The steady state distribution of* $\tilde{R}(t)$ *is exponential with parameter* $\tilde{a} > 0$*:* $\lim_{t\to\infty} P(\tilde{R}(t) > x) = \exp(-\tilde{a}x)$. *Parameter* $\tilde{a}$ *satisfies the balance equation between input and output in the buffer:*

$$\lambda(1 - \mathrm{E}[e^{-\tilde{a}\tilde{W}}]) = \mu\tilde{a} \tag{2}$$

and the convergence rate to steady state is at least $\max_\omega\{\mu\omega - \lambda(\mathrm{E}[e^{-\omega\tilde{W}}] - 1)\}$

Proof: Consider an initial distribution such that $\mathrm{E}[e^{-s\tilde{R}(0)}] < A$ for some $A > 0$.

We denote by $R^*(\omega)$ the Laplace transform $\mathrm{E}[e^{-\omega\tilde{R}(t)}]$. From (1), we get by Laplace transform:

$$\frac{\partial}{\partial t}R^*(\omega,t) = \mu\omega R^*(\omega,t) + \lambda R^*(\omega,t)(\mathrm{E}[e^{-\omega\tilde{W}}] - 1)\ . \tag{3}$$

Therefore the stationary distribution is Poisson of rate $\tilde{a} > 0$ such that

$$\lambda(1 - \mathrm{E}[e^{-\tilde{a}\tilde{W}}]) = \mu\tilde{a} \tag{4}$$

∎

3.3 Varying window and buffer sizes

The kernel of our results is in the following theorem.

Theorem 1 *In the system with exponential delay, when $N \to \infty$, let us assume that $\lim_{N\to\infty} W(y,0) = \tilde{W}(y)$ and $\lambda \int \tilde{W}(y)dy > \mu$. In this case $\lim_{N\to\infty} W(y,Nt) = \int_0^y w(x,t)dx$ such that:*

$$\frac{\partial}{\partial t}w(y,t) = \lambda e^{-a(t)(y-1)}w(y-1,t) + 2\lambda(1-e^{-2a(t)y})w(2y,t) - \lambda w(y), \quad (5)$$

$$w(y,0) = \tilde{W}'(y) \quad (6)$$

where $a(t)$ is the non-negative solution of

$$\lambda(\int (1-e^{-a(t)y})w(y,t)dy = \mu a(t). \quad (7)$$

Furthermore $R(t)$ is exponential of parameter $a(t/N)$.

Sketch of proof: The proof of theorem 1 proceeds in three steps:

(i) $\partial W(y,t)/\partial t = O(\lambda/N)$.

(ii) $R(t)$ is exponential of rate $a(t/N)$.

(iii) $\partial W(y,t)/\partial y$ tends to be equal to $\frac{1}{N}\sum_i w_i(y,t)$.

Point (i) is given by the fact that, during any time interval of length Δt, an average of $\lambda\Delta t$ servers transmit, leading to a modification of order $O(\lambda\Delta t/N)$ of the repartition function $W(y,t)$.

Point (ii) comes from the fact that, when N increases, $R(t)$ will converge to the exponential steady state of parameter $a(t/N)$ (as in the FWD model) quicker than $W(y,t)$ would significantly change. For example, when $a(t/N)$ is given by (7) at time t, the function $R(x,t)$ has time to converge to $\exp(-a(t/N)x)$ in the interval $[t, t+\Delta t\sqrt{N}]$, during which $W(y,t)$ remains unchanged at $\pm O(\lambda\Delta t/\sqrt{N})$, for Δt fixed.

Point (iii) comes from point (ii), namely that at any time $R(t)$ is exponential of parameter $a(t/N)$, when N tends to infinity. For the same reasons, quantities $R(t)$ tend to be independent between each transmission time of a given server. Therefore, for an isolated server, the probability density $w_i(y,t)$ of window size distribution varies according to (5), if at first order we ignore its impact on quantity $a(t)$.

The mean-field methodology (see Section 4.1 for the discrete setting) justifies this first order approximation when $N \to \infty$ and leads to: $W(y,t) = 1 - \frac{1}{N}\sum_i \int_0^y w_i(x,t)dx + O(\frac{1}{\sqrt{N}})$. ■

Corollary 2 *When $N \to \infty$ the steady state of the window size distribution is the function $w(y)$ which satisfies the equation:*

$$w(y) = e^{-a(y-1)}w(y-1) + 2(1-e^{-2ay})w(2y) \ , \tag{8}$$

and

$$\int w(y)dy = 1 \tag{9}$$

where a is the non-negative solution of

$$\frac{\int (1-e^{-ay})w(y)dy}{a} = \frac{\mu}{\lambda} \tag{10}$$

In the realistic model, the round trip delay of each server is a constant delay NT plus a processing time exponentially distributed with parameter λ/N (therefore an average round trip delay equal to $NT+N/\lambda$). This introduces a twist in the differential equation (5) which makes it a retarded equation:

$$\begin{aligned} \frac{\partial}{\partial t}w(y,t) &= \lambda e^{-a(t-T)(y-1)}w(y-1,t-T) \\ &\quad +2\lambda(1-e^{-2a(t-T)y})w(2y,t-T) - \lambda w(y,t), \end{aligned} \tag{11}$$

3.4 Large round trip delay

3.4.1 Window size stationary distribution

Our aim is to exploit the results obtained in the previous section in the case when the round trip delay is large (*i.e.*, when T or $1/\lambda$ is large). In this case, quantity a that characterizes the Poisson parameter of $R(t)$ (the buffer free space) is small.

We are interested in the stationary unconditional distribution of the window sizes. It is interesting to consider the case where $a \to 0$.

Theorem 2 *When $a \to 0$ the window size distribution $w(y)$ satisfies:*

$$\lim_{a\to 0} w(y/\sqrt{a})/\sqrt{a} = g(y) \tag{12}$$

where

$$g(y) = \sqrt{\frac{2}{\pi}} \prod_{k\geq 0}(1-4^{-k}2^{-1})^{-1} \sum_{n\geq 0} a_n \exp(-4^n y^2/2) \ , \tag{13}$$

with a_n satisfying the Taylor identity: $\sum_{n\geq n} a_n x^n = \prod_{k\geq 0}(1-4^{-k}x)$.

And

$$\sqrt{a} = (1+O(\sqrt{a}))\frac{\lambda}{\mu}g^*(2) \tag{14}$$

with $g^*(2) = \sqrt{\frac{2}{\pi}} \prod_{k\geq 0} \frac{1-4^{-k-1}}{1-4^{-k}2^{-1}} \approx 1.309833$.

Proof: Let us make the change of variable to consider $w(y) = \sqrt{a}g(y\sqrt{a})$. Equation (8) can be rewritten as

$$g(y) = e^{-y\sqrt{a}+a}g(y-\sqrt{a}) + 2(1-e^{-2y\sqrt{a}})g(2y) \tag{15}$$

When $a \to 0$ the equation expanded to first order in $\sqrt{a}$ becomes

$$(1-y\sqrt{a})(g(y) - \sqrt{a}g'(y)) + 4y\sqrt{a}g(2y) = g(y) \tag{16}$$

where $g'(y)$ is the first derivative of function $g(y)$ at point y.

Simplifying we obtain a new differential equation:

$$yg(y) + g'(y) = 4yg(2y) \tag{17}$$

This equation is easy to solve via Mellin transform $g^*(s) = \int_0^\infty g(y)y^{s-1}dy$:

$$g^*(s+1) - (s-1)g^*(s-1) = 2^{1-s}g^*(s+1) \tag{18}$$

Let $p(s) = \prod_{k\geq 0}(1-2^{-s-2k})$. Then, by fixing $g^*(s) = v(s)p(s)2^{s/2}\Gamma(s/2)$, we get

$$v(s) = v(s+2)\ . \tag{19}$$

Therefore $v(s)$ is a periodic function of period 2, defined everywhere in the complex plan. In other words $v(s) = f(e^{i\pi s})$, for some analytical function. In fact, $v(s)$ must be a constant function equal to some α (proof skipped). Therefore

$$g^*(s) = \alpha 2^{s/2}\Gamma(s/2)p(s) \tag{20}$$

The value of α is extracted from the identity $g^*(1) = \int w(y)dy = 1$.

Notice that function $g^*(s)$ has no singularity since poles of $\Gamma(s/2)$ at $s = -2k$, k natural integer, are exactly cancelled by being roots of $1 - 4^{-k}2^{-s}$. Therefore the function $g(y)$ will have all its derivatives at $y = 0$, and decays faster than any power function of y when $y \to \infty$.

It comes that the average window size is

$$\mathrm{E}[W] = \frac{g^*(2)}{\sqrt{a}} = \frac{2}{\sqrt{2\pi a}}\prod_{k\geq 0}\frac{1-4^{-k-1}}{1-4^{-k}2^{-1}} \tag{21}$$

and (10) becomes

$$\frac{\mu}{\lambda} = \frac{1-\int_0^\infty e^{-ay}w(y)dy}{a} \approx \mathrm{E}[W] \tag{22}$$

which leads to $\lambda = \mu\sqrt{2\pi a}\prod_{k\geq 1}\frac{1-4^{-k}2^{-1}}{1-4^{-k-1}}$.

By reverse Mellin transform it comes that

$$g(y) = \sqrt{\frac{2}{\pi}}\prod_{k\geq 0}(1-4^{-k}2^{-1})^{-1}\sum_{n\geq 0}a_n\exp(-4^n y^2/2)\ , \tag{23}$$

with a_n satisfying the Taylor identity: $\sum_{n\geq n}a_n x^n = \prod_{k\geq 0}(1-4^{-k}x)$. ∎

Corollary 3 *When $a \to 0$ the average window size is*

$$\mathrm{E}[W] = \frac{2}{\sqrt{2\pi a}} \prod_{k\geq 0} \frac{1-4^{-k-1}}{1-4^{-k}2^{-1}} \tag{24}$$

The average packet retransmission $\mathrm{E}[W]\frac{\lambda}{\mu} - 1 \to 0$, and the average number of dropped packets per window $\frac{1}{2}(\mathrm{E}[W^2] \times a)$ tends to 1.

In the case of the fixed delay plus exponential processing time model we have the following theorem.

Theorem 3 *When $T + 1/\lambda \to \infty$,*

$$\lim(\lambda T + 1) w(y/\sqrt{a})/\sqrt{a} = g(y) \tag{25}$$

and

$$\sqrt{a} = (1 + O(\sqrt{a}))\frac{\lambda}{\lambda T + 1}\frac{g^*(2)}{\mu} \tag{26}$$

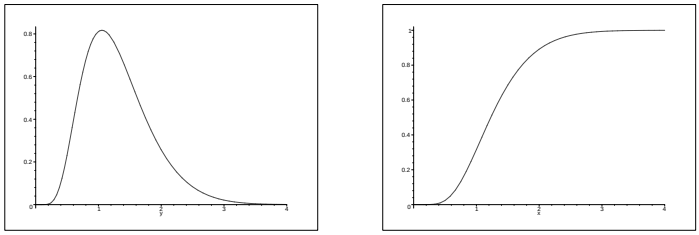

Figure 1: Limiting function $g(x)$ and its primitive for window size distribution.

3.4.2 Log-normal distribution of small windows

We show that the distribution is *log-normal* for small windows, namely that $\log \Pr\{W < x\} = -\Theta(\log^2 x)$.

Theorem 4 *In the large propagation delay condition we have $-\log \Pr\{W < x\} \sim (\log_2 x)^2$ when $x \to 0$.*

Proof: We use the fact that for any c

$$g(x) = \frac{\alpha}{2i\pi} \int_{c-i\infty}^{c+i\infty} x^{-s} 2^{s/2} \Gamma(s/2) \prod_{k\geq 0} (1 - 2^{-s-2k}) \tag{27}$$

We set $\Delta(x) = \min_k\{-\log x - k\log 2 - \frac{1}{2}\log k\}$, and $k(x)$ the value of k that minimizes the previous right hand side. We get $g(x) = \alpha\exp(2k(x)\log x + k(x)^2\log 2 - \log k(x))G(\Delta(x))$, where $G(.)$ is continuous. The fact that $k(x) \sim \log_2 x$ and that $\Delta(x)$ is bounded and tends to be periodic in $\log x$ completes the proof. ∎

It must be noted that the fluctuations $g(y)\exp((\log_2 y)^2)$ tend to be periodic in $\log y$, with period $\log 2$.

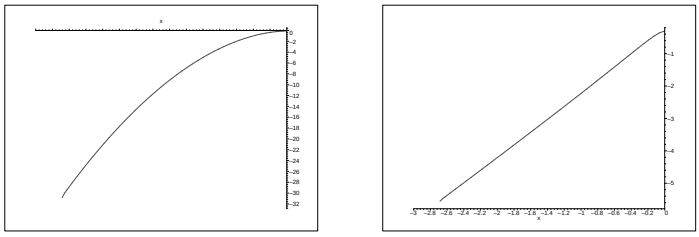

Figure 2: Limiting function $\log_{10} g(10^x)$ and $-\sqrt{\log_{10} g(10^x)}$ as function of x.

3.5 Simulation results

Simplified TCP

We have simulated simplified TCP with different numbers of connections, buffer sizes and random delays. By simplified TCP we mean the approximated version of TCP we have analyzed in previous sections, *i.e.*,

- windows are transmitted and acknowledged by batches;
- window sizes are real numbers.

Each of the connections starts at a random time after $t = 0$. The connection starts in slow start mode. A connection leaves the slow start mode when it meets its first congestion. The connections never come back to slow start mode (simulation runs are long enough so that no connection is in slow start mode). In all simulations of simplified TCP, the buffer service rate μ is 1 packet per time unit.

Real TCP

We have also simulated TCP Reno using the ns2 simulation tool [8]. The link from the buffer to the clients is 8 Mbps. The packet size is 1KB. Each server has a private link to the buffer at 1 Gbps.

Figures 3 (left) display the buffer occupancy distribution. We have simulated TCP Reno (top left) and simplified TCP (bottom left) On each plot the scale is logarithmic and the

straight line shows the theoretical exponential repartition function with the rate computed from limiting formula (26).

Figures 3 (right) display the window distribution, for TCP Reno (top right) and simplified TCP (bottom right). The distribution is obtained after freezing the simulation at a given time.

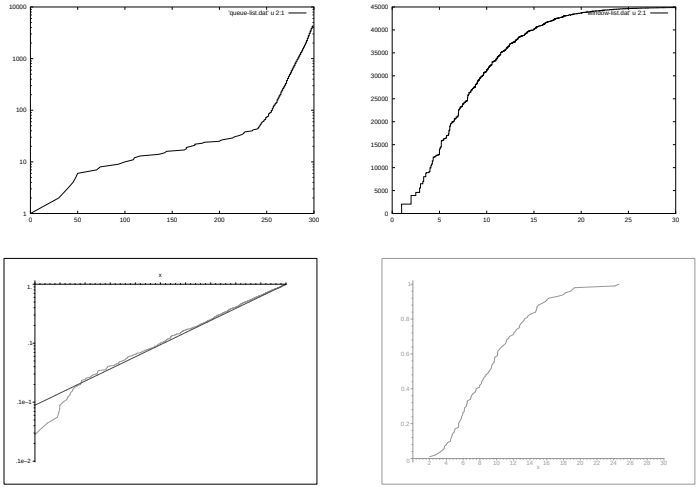

Figure 3: Repartition function of buffer occupancy (left), window size distribution (right), for 100 connections, buffer size 300, buffer service rate $\mu = 1$, average random processing time 10, simplified TCP (bottom), TCP Reno (top).

4 Discrete models

4.1 Discrete window and buffer sizes

Here we present a simple Markov model of a multi-connection TCP system that follows the continuous model of Section 3. Observe that, in the discrete Markov model with large number N of users, it is easier to derive the asymptotic equations and justify the mean field approach than in the continuous case (though we do not present here rigorous proofs).

We suppose now that the buffer and window sizes are integers (each size unit is associated with one packet). We are interested in the case where the number of users N, the maximum size of the window M_W and the size of the buffer M_R ($M_R > M_W$) are finite but large

and want to consider the asymptotic situation where $N \to \infty$, $M_W \gg 1$, $M_R \gg 1$. In particular, the case $M_W = O(N)$, $M_R = O(N)$ is of interest.

At moment t, let $W_n(t)$ be the size of the n-th window, $n = 1, 2, ..., N$, $r(t)$ be the size of the available (unoccupied) part of the buffer. Denote $R_j(t) = P(r(t) \geq j)$, $w_k(t)$ the ratio of window of size k, and $W(t) = \sum_k k w_k(t)$ the mean size of the window at the moment t. By $\bar{w}_k(t)$ we denote the mean values of random variables $w_k(t)$.

The users are addressed (and unload their packets) at the time moments t_i, $i = 1, 2, ...$, where the intervals $t_{i+1} - t_i$ are i.i.d. and distributed exponentially with parameter λ, thus the time points of window addressing form a Poisson flow. At the moments t_i a user is selected randomly, all users are i.i.d. The size W_n of a window that is addressed is changed at the same moment when it is addressed. The buffer length r changes at moments t_i and at the moments t'_k, $k = 1, 2, ...$, (when a packet in the buffer is processed). The intervals $t'_{k+1} - t'_k$ are i.i.d. and distributed exponentially with parameter μ, and the time points of change of buffer size also form a Poisson flow.

The window and buffer length changes follow the rules (see also Sections 2.2.1 and 3.2):

$$W_n(t_i + 0) = \begin{cases} \min[M_W, W_n(t_i - 0) + 1] & \text{if} \quad r(t_i - 0) \geq W_n(t_i - 0), \\ W_n(t_i - 0)/2 & \text{if} \quad r(t_i - 0) < W_n(t_i - 0). \end{cases} \tag{28}$$

$$r(t_i + 0) = \max\{0, r(t_i - 0) - W_n(t_i - 0)\}. \tag{29}$$

$$r(t'_k + 0) = \min[r(t'_k - 0) + 1, M_R]. \tag{30}$$

Here, to simplify notation, we define $(2k - 1)/2$ as k, $k = 1, 2, ...$.

For finite N, M_W and M_R the system performance is guided by a Markov process with finite number of states.

Our aim is to show that a theorem similar to theorem 1 is valid

Theorem 5 *If N is large, $N > N_0$, then $\forall t$, $t < T$ ($T \to \infty$ as $N_0 \to \infty$), the values $\bar{w}_k(t)$, $R_j(t)$ are close to the solution to an initial value problem for nonlinear ODE*

$$\begin{aligned} \frac{\mathrm{d}\bar{w}_k(t)}{\mathrm{d}t} &= \frac{\lambda}{N}\Big(\bar{w}_{k-1}(t)R_{k-1}(t) - \bar{w}_k(t) \\ &+ \bar{w}_{2k}(t)(1 - R_{2k}(t)) + \bar{w}_{2k-1}(t)(1 - R_{2k-1}(t))\Big), \quad 0 < k \leq N, \end{aligned} \tag{31}$$

$$\frac{\bar{w}_1(t)}{\mathrm{d}t} = \frac{\lambda}{N}\Big(- \bar{w}_1(t)R_1(t) + \bar{w}_2(t)(1 - R_2(t))\Big),$$

$$\frac{\mathrm{d}\bar{w}_N(t)}{\mathrm{d}t} = \frac{\lambda}{N}\Big(\bar{w}_{N-1}(t)R_{N-1}(t) - \bar{w}_N(t)(1 - R_N)\Big),$$

$$w_k(t) = 0, \; k > N.$$

$$\frac{\mathrm{d}R_j(t)}{\mathrm{d}t} = \mu\Big(R_{j-1}(t) - R_j(t)\Big) - \lambda R_j(t) - \lambda \sum_{k=0} \Big(R_j(t) - R_{j+k}(t)\Big)\bar{w}_k(t),$$

$$0 < j \le M_R, \tag{32}$$
$$R_0 \equiv 1, \;\; R_j = 0, \; j \ge M_R,$$

$$w_k(0) = w_k^{(0)}, \;\; \sum_k w_k^{(0)} = 1, \tag{33}$$

$$R_j(0) = R_j^{(0)}, \;\; R_0^{(0)} = 1, \;\; R_j^{(0)} \ge R_{j+1}^{(0)}, \;\; j \ge 0. \tag{34}$$

Further, as $N \to \infty$ *for any* $T < \infty$ *and* $\forall t, \; t < T, \; \bar{w}_k(t)$ *and* $R_j(t)$ *tend to the solution to equations* (31), (33) *and equation*

$$\mu\Big(R_{j-1}(t) - R_j(t)\Big) - \lambda R_j(t) - \lambda \sum_{k=0} \Big(R_j(t) - R_{j+k}(t)\Big)\bar{w}_k(t) = 0, \;\; 0 < j < M_R, \tag{35}$$

$$R_0 = 1, \; R_{j+1} \ge R_j, \;\; 0 < j \le M_R.$$

We use the following notation for the sequences: $\bar{\mathbf{w}} = \{\bar{w}_k\}_{k=1}^{M_W}$, $\bar{W} = \sum_k k\bar{w}_k$, $\mathbf{R} = \{R_j\}_{j=0}^{M_R}$. By convergence of a sequence we mean the convergence of its coordinates (see below).

At first, we have to know the global existence and the properties of the solution to (31),(32),(33),(34) and to (31),(33),(35) and the existence of the stationary solution to (31),(35). Below, we point out the problems we meet here and list the needed statements.

4.1.1 Differential equations with small parameter

In this subsection we deal with differential equations only (the connections with the Markov process are not considered here). We concentrate on the case where $M_W \gg 1, \;\; M_R \gg 1$. As $N \to \infty$, the systems (31),(32),(33),(34) become nontraditional systems with small parameter, because the number N of equations (31) is growing with the decline of the small parameter $1/N$. To investigate this system, we propose to consider first the systems with infinite number of equations of the same type as (31),(32),(33),(34), and (31),(33),(35), infinite number of variables $\bar{w}_k, \; R_j, \; k, j < \infty, \; M_W, \; M_R = \infty$, and with a small parameter $1/N = \varepsilon$. The solution to these systems is denoted by $\hat{\mathbf{w}} = \{\hat{w}_k\}_{k=1}^{\infty}$, $\hat{\mathbf{R}} = \{\hat{R}_j\}_{j=0}^{\infty}$, also $\hat{W} = \mathbb{E}k\hat{w}_k$. The global existence of a solution to such infinite nonlinear systems can be shown by the usual way. To go further, we present the conditions for the uniform in ε estimates for the decrease of $\hat{w_k}$ and $\hat{R}_j$ as $k, j \to \infty$. Below we describe informally the character of the solutions to the above ODE systems without giving the proofs.

We start with the initial value problem (31),(32),(33),(34), $N, M_W, M_R = \infty$. Consider first two auxiliary systems. In the first one R_j are fixed in (31) and $\hat{\mathbf{w}}(t; \mathbf{R})$ forms a solution to (31),(33). In the second one w_k are fixed in (32) and $\hat{\mathbf{R}}(t; \mathbf{w})$ forms a solution to (32),(34) (compare with Sections 3.2, 3.3).

Lemma 2 *Let* $\mathbf{R} = \{R_j\}$, $j = 1, 2, ...,$ *be fixed,* $1 = R_0 \geq R_1 \geq ...$, $\lim_{j\to\infty} R_j = 0$. *Then there exist a global solution* $\hat{\mathbf{w}}(t; \mathbf{R})$ *to* (31), (33) *and a unique stationary solution* $\hat{\mathbf{w}}^{st}(\mathbf{R})$ *to* (31), $\lim_{t\to\infty} \hat{\mathbf{w}}(t; \mathbf{R}) \to \hat{\mathbf{w}}^{st}(\mathbf{R})$. *The convergence rate is not faster than* $O\left(\frac{1}{N}\right)$.

If there are two sequences $R_j^{(i)}$, $i = 1, 2$, *and* $R_j^{(1)} \geq R_j^{(2)}$ *then* $\hat{W}^{st}(\mathbf{R}^{(1)}) \leq \hat{W}^{st}(\mathbf{R}^{(2)})$.

Lemma 3 *Let* $\bar{\mathbf{w}} = \bar{w}_k$, $k = 1, 2, ...,$ *be fixed,* $\bar{W} > \frac{\mu}{\lambda}$. *Then there exist a global solution* $\hat{\mathbf{R}}(t; \bar{\mathbf{w}})$ *to* (32), (34) *and a unique stationary solution* $\hat{\mathbf{R}}^{st}(\bar{\mathbf{w}})$ *to* (32), $\lim_{t\to\infty} \hat{\mathbf{R}}(t; \bar{\mathbf{w}}) = \hat{\mathbf{R}}^{st}(\bar{\mathbf{w}})$, $\hat{R}_j^{st} = e^{-aj}$ *where* $a = a(\bar{\mathbf{w}})$ *is a solution to equation*

$$\mu(e^a - 1) = \lambda(1 - \sum_k e^{-ak}\bar{w}_k). \tag{36}$$

If $R_j > e^{-a_1 j}$, $a_1 > a$ *or* $R_j < e^{-a_1 j}$, $a_1 < a$, *then* $R_j(t)$ *converge to* R_j^{st} *with a rate that is* $O\Big(|a - a_1|\Big)$.

If there are two sequences $\bar{w}_k^{(i)}$, $i = 1, 2$, *and* $\bar{W}^{(1)} > \bar{W}^{(2)}$ *then* $a(\bar{\mathbf{w}}^{(1)}) < a(\bar{\mathbf{w}}^{(2)})$.

Return to the problem (31),(32),(33),(34) with $N, M_W, M_R = \infty$, $\varepsilon > 0$. Here as $\varepsilon \to 0$ we have two time scales and two types of variables: 'fast' $\hat{R}_j(t)$ and 'slow' $\hat{w}_k$. A standard asymptotic approach to investigate the solution is to separate the fast and slow variables in the following way (compare with the first and second auxiliary systems).

First the slow variables $\hat{\mathbf{w}}(t_0)$ are fixed and fast variables follow (32),(34) during some time interval $(t_0, t_0 + T_w)$, where T_w, $1 \ll T_w \ll N$, is chosen so that, during time T_w, $\hat{\mathbf{R}}(t)$ becomes close to $\hat{\mathbf{R}}^{st}(\hat{\mathbf{w}}(t_0))$. After that the slow variables $\hat{\mathbf{w}}(t)$ follow (31),(33) with fixed values $\hat{\mathbf{R}}^{st}(\hat{\mathbf{w}}(t_0 + T_w))$ during some time interval $(t_0 + T_w, t_0 + T_w + T_R)$, etc (thus the solutions to auxiliary systems are considered). We will refer to this process as the *separation* process. Rescaling t: $t^{(N)} = t/N$ one has to prove that, as $N = \frac{1}{\varepsilon} \to \infty$, the sequences $\hat{\mathbf{w}}(t^{(N)})$, $\hat{\mathbf{R}}(t^{(N)})$ tend to solution to (31),(35),(33) $\forall T$ and $t < T$.

For the proof one has to show that the following statements are valid.

Lemma 4 *Let* $N, M_W, M_R = \infty$, *the initial values* $\hat{w}_k^{(0)}$ $\hat{R}_j^{(0)}$ *decrease exponentially as* $k, j \to \infty$. *If* $\hat{R}_j(t) < e^{-a_0 j}$, $a_0 > 0$, $t < T$, *then the terms* $\hat{w}_k(t)$ *of the solution to* $(31) - (34)$ *decrease at least exponentially as* $k \to \infty$, $t < T$.

It follows from this lemma that for $N \geq N_0 \gg 1$, $M_W \gg 1$, $M_R \gg 1$ and $t < T$, $T = T(N_0)$ the solution to the finite system $(31) - (34)$ with these large N, M_W, M_R and 'truncated' initial values is close to the solution to the system with $N, M_W, M_R = \infty$.

Lemma 5 *If* $\hat{W}(0) > \frac{\mu}{\lambda}$, $N, M_W, M_R = \infty$ *then* $\hat{R}_j(t) < e^{-a_0 j}$, $a_0 > 0$.

Idea of proof for the separated systems. It follows from lemmas 2 and 3 that for given a, $\hat{R}_j = e^{-aj}$, the mean value $\hat{W}^{st}(\hat{\mathbf{R}}) = \hat{W}^{st}(a)$ of solution to (31) decreases when a increases and, on the contrary, for fixed $\hat{\mathbf{w}}$ the value of $a = a(\hat{\mathbf{w}})$ (with $\hat{R}_j^{st}(\hat{\mathbf{w}}) = e^{-aj}$) increases as $\hat{W}$

increases. Thus we have on the $(\hat{W}, a)$ plain a decreasing in a curve $\hat{W}^{st}(a)$ and increasing in $\hat{W}$ curve $a(\hat{W})$.

As $\hat{\mathbf{R}}$ is a sequence of fast variables, during the separated process, the distance between the point $a = a(t)$ and the point $a(\hat{\mathbf{w}}(t))$ is of order $O(N^{-\nu})$, $0 < \nu < 1$, (ν depending on $\frac{T_W}{N}$), and on the plain $(\hat{W}, a)$ the point $(\hat{W}(t), a(t))$ moves in the direction of the intersection of curves $\hat{W}^{st}(a)$ and $a(\hat{W})$, $a(t)$ being bounded away from zero. ∎

Now we can summarize:

Lemma 6 *If* $\hat{W}^{(0)} = \sum_k k\hat{w}_k^{(0)} > \frac{\mu}{\lambda}$ *and* M_W, M_R *are sufficiently large then as* $N \to \infty$ *the solution* $\hat{\mathbf{w}}(t)$, $\hat{\mathbf{R}}(t)$ *to* (31), (32), (33), (34) *tends to solution to* (31), (33), (35) *for all* $t \le T$, $\forall T < \infty$. *The convergence is uniform on any interval* $0 < t \le T$.

Proof of Lemma 6 follows from Lemmas 2 - 5. ∎

4.1.2 Mean-field methodology

Return to the Markov process. To demonstrate the mean-field approach, consider an auxiliary process (28) with fixed transitional probabilities $R_j = P\{r > j\}$ (compare with Section 3.2). We want to show that as $N \to \infty$ the mean values $\mathbb{E}\bar{w}_k(t)$ tend in the weak sense to the solution to (31),(33),(29),(30).

Consider a set $\mathcal{U}$ of sequences $\mathbf{w} = \{w_i\}_{i=0}^{\infty}$, $w_i \ge 0$, $\sum_i w_i = 1$. The metric in $\mathcal{U}$ is defined by $\rho(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}) = \sup_i \dfrac{|w_i^{(1)} - w_i^{(2)}|}{i+1}$. Let the subset $\mathcal{U}_N$ be a set of $\mathbf{w}$, $w_i = n_i/N$, where $n_i = 0, 1, ..., N$.

Consider a space of smooth continuous function $f(\mathbf{w})$ on $\mathcal{U}$.

For $N \gg 1$ and fixed values R_j, the generating operator of the Markov process has the form

$$\begin{aligned}\mathbf{A}_N f(\mathbf{w}) \quad = \quad & \frac{\lambda}{N}\Big(\sum_{k=1} \frac{\partial f(\mathbf{w})}{\partial w_k}\Big(w_{k-1}R_{k-1} + w_{2k}(1-R_{2k}) + w_{2k-1}(1-R_{2k-1}) - w_k\Big) \\ & + \frac{1}{N} O\Big(\frac{\partial 2 f(\mathbf{w})}{\partial w_{k_1} \partial w_{k_2}}\Big)\Big). \end{aligned} \tag{37}$$

That makes it possible to compare the generating operator $\mathbf{A}_N f(\mathbf{w})$ with the differential operator

$$\frac{\mathrm{d}f(\hat{\mathbf{w}})}{\mathrm{d}t} = \sum_{k=1} \frac{\partial f(\hat{\mathbf{w}})}{\partial \hat{w}_k}\Big(\hat{w}_{k-1}R_{k-1} + \hat{w}_{2k}(1-R_{2k}) + \hat{w}_{2k-1}(1-R_{2k-1}) - \hat{w}_k\Big). \tag{38}$$

Define $f(\hat{\mathbf{w}}(t))$ as the solution to (38) with the initial condition $\hat{w}_k(0) = \mathbb{E}\bar{w}_k^{(0)}$. Consider a semigroup acting on functions defined on $\mathcal{U}_N$:

$$\mathbf{T}_N f(\mathbf{w})(t) = \mathbb{E}(f(\mathbf{w}(t)) \mid \mathbf{w}(0) = \mathbf{w}^{(0)}), \quad \mathbf{w} \in \mathcal{U}_N.$$

Using the technique of [9] (see also [10]) one can prove the weak convergence of the mean values $\bar{w}_k$ and of R_j to the solutions to the differential equation (38) :

Lemma 7 *For any* $0 < t < T_R$

$$\lim_{N\to\infty} \sup_{\mathbf{w}\in\mathcal{U}_N} \mathbf{T}_N f(\mathbf{w})(t) = f(\hat{\mathbf{w}}(t))$$

where $\mathbf{w}$ *is the random sequence guided by the auxiliary Markov process and* $\hat{\mathbf{w}}$ *is a solution to boundary value problem* (31), (35), (33).

We do not give the proof of the theorem presented above but make only a short remark. As N is large, the transition probabilities $R_j(t)$ of the Markov process (28),(29),(30) are close to the semi-stationary values $R_j^{st}(\mathbf{w})$ (of the auxiliary process), which in their turn are close to the stationary solution $\hat{\mathbf{R}}^{st}(\hat{\mathbf{w}})$ to (31),(32),(36). This permits us to compare the solution to ODE (31)−(36) with the Markov process (28)−(30) and to prove Theorem 5.

4.2 Fixed loss probability model

In this section, we derive a simplified model of a single TCP connection, in order to analyze the autocorrelation function of TCP traffic. We assume that the packet loss probability is constant and equal to a, and that the losses are independent. This is justified by the discrete model of the previous section (and by section 3.3) where a is constant. We assume that we collect all buffer changes within one round trip time (RTT) and, for simplicity, we turn to a discrete time model, where the unit is the RTT. Hence, we can model the window size adaptation with a Markov chain. A similar model has been studied via simulation in [11].

We deduce the following probabilities for $a \ll 1$:

$$Pr(\text{no loss in a window}) = (1-a)^W \approx e^{-aW} \quad (39)$$

$$Pr(\text{loss in a window}) = 1-(1-a)^W \approx 1-e^{-aW} \quad (40)$$

We now define the TCP discrete time Markov chain. The states are the window sizes and the transition probabilities can be calculated from (39) and (40). In reality, the receiver announces a maximum window size, which means that the Markov chain is finite. As the tendency in the Internet is towards increasing this value, we will ignore it in the analysis. We suppose that it corresponds to a very large window size, which is never achieved in practice.

Let $\mathbf{P}$ be the transition matrix of the TCP Markov chain. It is easy to see that $\mathbf{P}$ is stochastic, irreducible and aperiodic. Thus, according to the Perron-Frobenius theorem, its dominant eigenvalue is 1 with corresponding right eigenvector $\boldsymbol{\pi}$, which gives the stationary distribution of the Markov chain. We can also deduce that $\lim_{n\to\infty} \mathbf{P}^n = \boldsymbol{\pi}\cdot\mathbf{1}$, where $\mathbf{1}$ is a line vector of ones.

The matrix $\mathbf{P}$ is of the following form:

$$\mathbf{P} = \begin{pmatrix} 1-e^{-a} & 1-e^{-2a} & 1-e^{-3a} & 0 & \cdots \\ e^{-a} & 0 & 0 & 1-e^{-4a} & \\ 0 & e^{-2a} & 0 & 0 & \\ & 0 & e^{-3a} & 0 & \\ & & 0 & e^{-4a} & \cdots \\ \vdots & & & \vdots & \ddots \end{pmatrix} \tag{41}$$

If the initial distribution of windows is $\boldsymbol{\pi}(0)$, then after n RTT's: $\boldsymbol{\pi}(n) = \mathbf{P}^n\boldsymbol{\pi}(0)$. From the spectral decomposition of the matrix, we get that the convergence rate is exponential:

$$\|\boldsymbol{\pi}(n) - \boldsymbol{\pi}\| = O(\rho^n) \tag{42}$$

for every ρ such that $|\lambda_2| < \rho < 1$, where λ_2 is the second largest eigenvalue of $\mathbf{P}$.

4.2.1 Autocorrelation function of a single TCP connection

The traffic intensity $I(t)$ at time t is a function of the present window size $W(t)$ and the round trip time, which we denote as D:

$$I(t) = \frac{W(t)}{D} \text{ packets/second} \tag{43}$$

We calculate the traffic auto-covariance function $C(x)$ at time t:

$$C(x) = \text{cov}[I(t), I(t+x)] = E[I(t)I(t+x)] - (E[I(t)])^2 \tag{44}$$

Since we are dealing with a stationary process, $C(x)$ does not depend on time t, so we can fix $t = 0$. We express time in RTT multiples n, and compute the first term of (44) by supposing that the initial window size is k and averaging on all k. We want to calculate the auto-covariance of a system that has reached equilibrium, so we can assume that the initial distribution is the stationary distribution $\boldsymbol{\pi}$:

$$\begin{aligned} \text{E}[I(0)I(nD)] &= \frac{1}{D^2}E[W(0)W(nD)] = \frac{1}{D^2}\sum_k kE(W(nD)|W(0)=k)\boldsymbol{\pi}_k \\ &= \frac{1}{D^2}\sum_k k\sum_l l\,(\mathbf{P}^n\mathbf{1}_k)_l\boldsymbol{\pi}_k \end{aligned} \tag{45}$$

where $\mathbf{1}_k$ is a column vector with all entries being 0 except for entry k which is 1.

If we define $\mathbf{p}$ as a column vector such that $\mathbf{p}_i = i\boldsymbol{\pi}_i$, and $\mathbf{u}$ as a line vector such that $\mathbf{u}_i = i$, then:

$$E[I(0)I(nD)] = \frac{1}{D^2}\sum_l l\,(\mathbf{P}^n\mathbf{p})_l = \frac{1}{D^2}\mathbf{u}(\mathbf{P}^n\mathbf{p}) \tag{46}$$

For $n \to \infty$ we have:

$$\frac{1}{D^2}\mathbf{u}(\lim_{n\to\infty}\mathbf{P}^n\mathbf{p}) = \frac{1}{D^2}\sum_l l\,(\boldsymbol{\pi}\cdot\mathbf{1}\cdot\mathbf{p})_l = (E[I(0)])^2 \tag{47}$$

Combining (44), (46) and (47), we get the auto-covariance function:

$$C(nD) = \frac{1}{D^2}\mathbf{u}\left(\mathbf{P}^n\mathbf{p} - \lim_{n\to\infty}\mathbf{P}^n\mathbf{p}\right) \tag{48}$$

From the spectral decomposition of $\mathbf{P}$, as in (42), we conclude that the auto-covariance function decreases exponentially with rate $O(\rho^n)$ for all ρ such that $|\lambda_2| < \rho < 1$. By normalizing we obtain the autocorrelation function, which is also $O(\rho^n)$.

The next step in our analysis is to calculate the eigenvalues of $\mathbf{P}$ for a large scale of error rates. The calculations are made by fixing a maximal window size of 1000 packets, thus truncating the matrix $\mathbf{P}$. The probability of reaching the maximum window size is close to 0 for the error rates we consider, so ignoring larger values does not affect significantly our results.

In the equivalent continuous model of TCP, we scaled window sizes by a factor $\sqrt{a}$ to obtain the limit distribution. We expect to find a similar factor in the auto-covariance function. This observation leads us to approximate the spectral gap of the TCP Markov chain by expression $C\sqrt{a}$, where C is a constant. In figure 4, we compare the calculated spectral gap values, for different error rates a, and the values corresponding to the proposed approximation for $C = 1.6$. In all our calculations, the second eigenvalue is real, of multiplicity 1, which means that the auto-covariance is $O({\lambda_2}^n)$.

More precisely, we have the following first order approximation:

$$C(nD) = \frac{A}{D^2}{\lambda_2}^n \approx \frac{A}{D^2}e^{-C\sqrt{a}n} \tag{49}$$

where A is another constant.

Observe that a very small error rate results in a second eigenvalue close to 1, meaning that the autocorrelation function decreases very slowly (although exponentially).

4.2.2 Simulations of a single TCP connection

To verify that our simplified model can predict the autocorrelation of a real TCP connection, we conducted a number of simulations with the network simulator (ns2). In the simulations, the RTT is mainly caused by the link delays and the error rate is constant and due to a loss agent. By measuring the number of packets transmitted during an interval equal to the RTT, we obtain the traffic intensity $I(t)$. We then calculate the covariance of $I(t)$ and $I(t+x)$ when t varies and we normalize to obtain the traffic autocorrelation.

We present the results for two models with different error rates, compared to the calculations of the previous section. The duration of these simulations is 1000 seconds of simulated time, and we start making measurements after waiting for the system to stabilize for 100

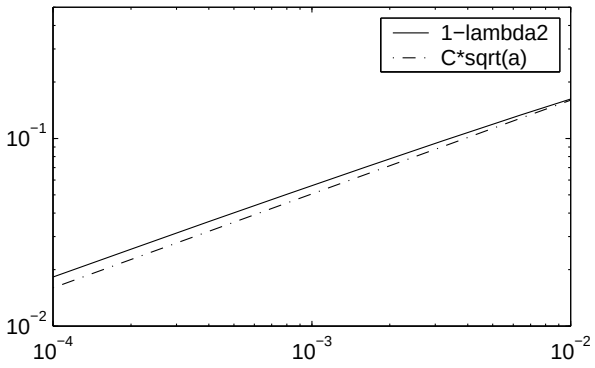

Figure 4: Spectral gap of the TCP Markov chain for different error rates a.

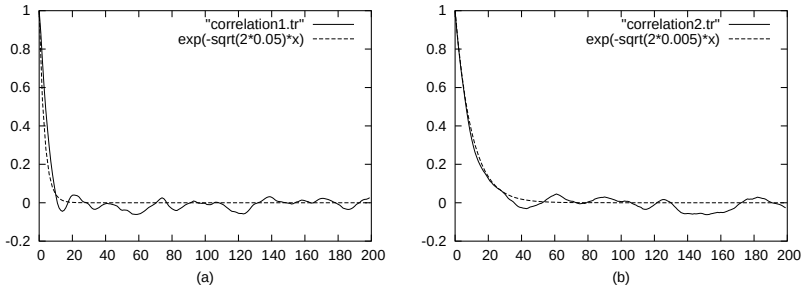

Figure 5: TCP traffic autocorrelations for error rates (a) $a = 0.01$, (b) $a = 0.005$. The time unit is the RTT.

seconds. In figure 5 we draw the autocorrelation for error rates $a = 0.01$ and $a = 0.005$. The time unit is the round trip delay, which in this case is equal to 100ms.

The oscillations are due to the finite duration of the simulations.

5 Long term dependencies

In this section we will show that long term dependencies can arise from heavy tailed round trip delays. One of the many undesirable effects of long term dependencies stands in the loss rates in the buffers. Long term dependent traffics generate loss rates that are inverse power functions of the buffer size, contrary to the exponential function of the buffer size in Poisson models. Consequently, the increase in buffer size is prohibitively expensive to significantly reduce the packet loss rate.

Recently, it has been discovered that the internet topology contains numerous heavy tailed features. Among them are the router degree, router reachability degree and the multicast tree size [13, 12, 14]. In [15] there is evidence that the RTT distribution is also heavy tailed, *i.e.*, the RTT complementary cumulative density function $P(RTT > x)$ corresponds to a power law with exponent approximately 1.5.

In the first subsection we show that a link shared by several TCP connections with Round Trip Times with a heavy tailed distribution generates long term dependence. In the second subsection we provide simulations to compare with the theoretical results.

5.1 Autocorrelation of several TCP connections with heavy tailed round trip delays

We consider that the link is shared by several TCP connections with different round trip delays, so that the round trip delays distribution is heavy tailed. To simplify, we assume there is an infinite sequence of TCP connections and sequence number i has round trip delay $D_i = Di^\beta$ for $\beta > 0$ and D fixed, and the error rate a remains the same.

Since the TCP connections are assumed to be independent, the auto-covariance function of the aggregated traffic is equal to the sum of the auto-covariance functions of the individual connections.

It comes from (49) that:

$$C(x) \approx \frac{A}{D^2} \sum_{i=1}^{\infty} i^{-2\beta} \exp(-C\frac{x}{D} i^{-\beta}) \tag{50}$$

Function $C(x)$ is a harmonic sum generated from function $\exp(-C\frac{x}{D})$. The Mellin transform of function $C(x)$: $C^*(s) = \int_0^\infty x^{s-1} C(x) dx$ has the expression:

$$C^*(s) = \frac{A}{D^2} \left(\frac{C}{D}\right)^{-s} \zeta((2-s)\beta)\ \Gamma(s) \tag{51}$$

where $\zeta(s) = \sum_{i=1}^{\infty} i^{-s}$ is Euler's *zeta* function and $\Gamma(s)$ is Euler's *Gamma* function ($\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$).

Quantity $\zeta((2-s)\beta)$ has a simple pole at $s = 2 - \frac{1}{\beta}$. Therefore $C^*(s)$ is defined on the strip $0 < \Re(s) < 2 - \frac{1}{\beta}$. The classical results on the Mellin transform state that there exist B and $\varepsilon > 0$ such that $C(x) = Bx^{\frac{1}{\beta}-2}(1 + O(x^{-\varepsilon}))$ when $x \to \infty$ [16]. This implies that the traffic has long term dependencies when $\frac{1}{2} \leq \beta \leq 1$. However, if the number of TCP connections is finite, we expect to observe a heavy tailed behavior for a finite time scale, which is a multiple of the largest RTT in the system. In the next section, we will see that, even for a small number of connections, this upper bound can be quite large.

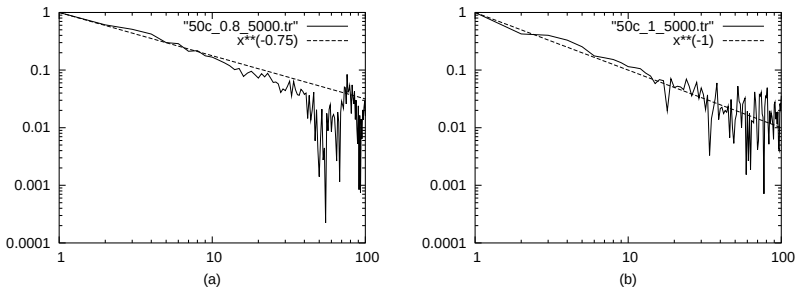

Figure 6: Traffic autocorrelations of 50 TCP connections with heavy tailed RTT distributions (a) $D_i = 40i^{0.8}$, (b) $D_i = 40i$. The time unit is 500ms.

5.2 Simulations of several TCP connections

We have simulated with ns2 the case of many TCP connections with heavy tailed RTT's sharing the same link. The simulation models 50 clients downloading very large files from different servers. The traffic of all connections traverses a shared link of capacity 100 Mbps. Each client is connected to the shared link via a slow link at 10 Mbps and, at the other end, each server is connected via a private link at 1 Gbps. The fixed propagation delays in these private links are chosen to follow a power law. The traffic measurements are made on the shared link in intervals of 500 ms. The packet loss rate is $a = 0.001$ for each connection. The packet size is 1KB and the maximum window size is 1000 packets.

In figure 6 we draw the autocorrelation functions in log-log scale. The dashed line shows the theoretical power law decrease. The RTT distributions are of the form $D_i = 40i^{\beta}$, with $\beta = 0.8$ in (a) and $\beta = 1$ in (b). According to the theoretical analysis, the expected power law exponents for the traffic autocorrelations are $\frac{1}{\beta} - 2$, that is -0.75 and -1, respectively.

The duration of the simulations is 5000s and the measurements start after a stabilization period of 100s.

The fluctuations in figure 6 are due to the finite duration of the simulations and the fact that there are RTT's which are larger than the time unit. The exact characterization of TCP autocorrelation in a finer time scale is a subject of further research.

References

[1] V. Jacobson, Congestion avoidance and control, *Proceedings of* ACM SIGCOMM '88, August 1988.

[2] T. J. Ott, J. H. B. Kemperman and M. Mathis, The Stationary Behavior of Ideal TCP Congestion Avoidance, *Proc. of* IEEE INFOCOM'99, New York, NY, March 1999.

[3] F. Baccelli and D. Hong, TCP is max-plus linear and what it tells us on its throughput, *Computer Communication Review*, vol.30, no.4 p. 219-30, Oct. 2000.

[4] L. Qiu, Y. Zhang and S. Keshav, On individual and aggregate TCP performance, *Proceedings of* ICNP'99: 7th International Conference on Network Protocols, Toronto, Canada; 31 Oct.-3 Nov. 1999.

[5] C. Adjih, P. Jacquet and N. Vvedenskaya, Performance evaluation of a single queue under multi-user TCP/IP, INRIA Research report RR-4141, 2001.

[6] C. Adjih, P. Jacquet and N. Vvedenskaya, Performance evaluation of a single queue under multi-user TCP/IP version 2, INRIA Research report RR-4478, 2002.

[7] F. Baccelli, D. McDonald and J. Reynier, A mean field model for multiple TCP connections through a buffer implementing RED, *Perform. Eval.* 49(1/4): 77-97 (2002).

[8] UCB/LBNL/VINT Network Simulator - ns2, http://www.isi.edu/nsnam/ns, 2001.

[9] S.N. Ethier and T.G. Kurtz, Markov Processes Characterization and Convergence, *John Willey and Sons* (1986).

[10] N.D. Vvedenskaya, R.L Dobrushin and F.I. Karpelevich, A queueing system with selection of the shortest of two queues: an asymptotical approach, *Problems of Information Transmission*, **32** (1996), 15–27

[11] D. R. Figueiredo, B. Liu, V. Misra, and D. Towsley, On the autocorrelation structure of TCP traffic, *Computer Networks Journal*, Special Issue on Advances in Modeling and Engineering of Long-Range Dependent Traffic, 2002.

[12] J. Chuang, M. Sirbu, Pricing multicast communication: a cost based approach, INET 1998.

[13] T. Bu, D. Towsley, On distinguishing between internet power law topology generator, INFOCOM 2002.

[14] C. Adjih, L. Georgiadis, P. Jacquet, W. Szpankowski, Is the internet fractal: the multicast power law revisited, SODA 2002.

[15] A. Broido, E. Basic and K.C. Claffy, Invariance of the Internet RTT spectrum. Global RTT analysis, ICIR, August 2002, http://www.caida.org/ broido/rtt/rtt.html

[16] P. Flajolet, X. Gourdon, and P. Dumas, Mellin transforms and asymptotics: Harmonic sums, *Theoretical Computer Science* 144, 1–2 (June 1995), 3–58.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay 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 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