Performance Evaluation of a Single Queue under Multi-User TCP/IP Connections

Cédric Adjih, Philippe Jacquet, Nikita Vvedenskaya

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Performance evaluation of a single queue under multi-user TCP/IP connections

Cédric Adjih, Philippe Jacquet, Nikita Vvedenskaya

N° 4141

THÈME 1

ISSN 0249-6399 ISRN INRIA/RR--4141--FR+ENG

Performance evaluation of a single queue under multi-user TCP/IP connections

Cédric Adjih, Philippe Jacquet, Nikita Vvedenskaya

Thème 1 — Réseaux et systèmes
Projet Hipercom

Rapport de recherche n° 4141 — — 40 pages

Abstract: We study the performance of several TCP connections through the bottleneck of a slow network accessed via a single queue with high capacity. Using mean-field approximation methology, we establish some asymptotical results about queue length distribution and windo size distribution when the number of user increases proportionally to buffer capacity. We also give an evaluation of TCP fairness under these traffic conditions.

Key-words: Internet, TCP, asymptotics, mean-field approximation, fairness

IST BRAIN project IST-1999-10050

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

Évaluation de performance d'une file d'attente unique soumise à des connexions TCP multiples

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 méthodologie de l'approximation par le 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 croît proportionnellement 'a la capacité de la file d'attente. Nous donnons aussi une évaluation de l'équité du protocole TCP dans ce genre de situation.

Mots-clés : Internet, TCP, évaluation de performances, comportement asymptotique, approximation par le champ moyen, équité.

1 Introduction

The protocol TCP [1] is widely used in the Internet (Web, FTP). The proportion of connections made on this protocol is overwhelming (close to 99.9%). The protocol TCP is an end-to-end flow control protocol based on dynamic transmission windows. The protocol does not make any assumption on the underlying network and on how heterogeneous networks are connected. The reason of the success of TCP is mainly based on its high dynamic that make it able to adapt itself to any kind of network capacity from few bauds to several gigabit per second.

The protocol TCP has received special attention since its formulation as an internet standard. In particular its performance has been the research focuse of several researcher. The paper [3] is a pioneer paper where the performance of TCP is analyzed 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. [2] have analyzed a single TCP connection through a sequence of finite capacity routers. The analysis is interesting because it relyes on an explicit formulation of the problem in (max,plus) algebra. The problem of several connections even a single router has been only simulated with the notable exception of [4], but limited to very specific phases of the protocol.

The aim of this paper is to investigate via analytic methods the multi-connection cases 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 is absolutely out of reach of the present performance evaluation toolbox. However our analysis is made possible because we consider the asymptotic case where N is large and the analysis is much simplified. 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 server under TCP/IP and the bottleneck is a router with a finite buffer and a slow network interface. The servers are not necessarily different, as well as the users, but we assume that N connections are active. 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)

We assume that the users are located on the local loop and the servers are located on the fast backbone. We assume that there is a router at the border of the local loop and the fast backbone (head-end, etc).

We will consider that every user is downloading a file of infinite size and we are interested into analyzing the steady state of every connection. We also assume that the round trip delay between server and user is large.

The paper is divided into the following sections. A first section is devoted to a short presentation of TCP. A second section describes of the models. A third section presents the analysis of a simplified model and the result are used to the analysis of a more realistic

model. A last section is devoted to numerical example with interpretations about traffic fairness. In particular it will be shown that

2 The TCP connection protocol

2.1 TCP overview

The connections are done under a window protocol like TCP/IP. Packets are transmitted in order and must be acknowledged by the end-user. A packet is loss 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.

- when no packet is lost in a window (successful window), the next window size is incremented of 1 unit.
- when a packet loss occur (failed window), packet retransmission starts from this packet but with a window size halved.

2.2 Self-clocking

The server updates its window size on-line with each acknowledgement return. 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 *via* appropriate tuning of parameter k.

2.3 Slow start

There is also the *slow start* process where the window is doubled at each successful window as long the window size is below a pre-defined threshold. We will not discuss of this feature here, since we focus on steady state analysis.

3 The models

3.1 The continuous model and batch transmission

We consider that the router has a very large buffer but whatever its size the window protocol strives to almost fill it in steady state situation. The parameter of interest is the current available room in the buffer at time t called $R(t)$.

We assume that the buffer contains continuous data (fluid approximation), and $R(t)$ is a positive real number. The buffer is served at speed $\mu > 0$.

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

- the server transmits all the packets within its window at the same time;
- The backbone has an infinite speed so that all the packets of the same window arrives simultaneously in the buffer;
- the end-user acknowledges all the packets received from the same window in the same packet (aggregated acknowledgement).

When server i transmits its window, we assume

1. If the current window $W_i(t) < R(t)$, then $R(t)$ becomes $R(t) - W_i(t)$ and $W_i(t)$ becomes $W_i(t) + 1$;
2. else, if $W_i(t) \geq R(t)$, then $R(t)$ becomes 0, and $W_i(t)$ becomes $W_i(t)/2$.

3.2 Discrete model

In this model we suppose that the buffer length $r(t)$ and the window length $w(t)$ are descrete, $r = 1, 2, ...,$ $w_i = 1, 2, ..N$ are integers.

It is supposed that the length of the window is increased by 1, or $w_i \rightarrow w_i/2$ at any step the window is addressed (here we have to specify that when $w = 2i - 1$ it becomes i); the length of free buffer can increase by 1 or can be made 0, following the model proposed above.

We will mainly focuse our presentation on the continuous model, but we will outline some sketches from the discrete model as often as possible.

3.3 The round trip model

3.3.1 The non-realistic exponential round trip delay

We make a further approximation by assuming that windows arrive on buffer according to a Poisson process of rate $\lambda > 0$. Within this model the round trip delays per server is i.i.d. and is Poisson with mean N/λ. Windows arrive on buffer according to a Poisson process of rate λ. A further level of abstraction in the model consists into monitoring the arrival point as Poisson events and at each Poisson event to randomly select the transmitter server i, uniformly in $(1, N)$. Within this model there is no need to keep track the exact matching between server and window size. It suffices to keep track of the repartition function $W(y, t)$:

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

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

The exponential round trip model is convenient for a first analysis but it is highly non realistic. First, it varies in too large proportion: a difference between two consecutive round

trip delay will be interpreted as a packet loss by the server and cause a window halving. Second 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 have an exponential distribution. In fact it will be proven that the buffer delay will be close to $\frac{B}{\mu}$.

3.3.2 A realistic model with fixed delay plus random processing time

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

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

We can see the factor N as the nominal buffer capacity per server. For the convenience of the presentation we will 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 server are either (i) transmitting, (ii) in fixed propagation delay or (iii) in random processing delay. We keep the repartition function $W(y,t)$ but with a different meaning:

$$W(y,t) = \frac{\text{number of servers in processing at time } t \text{ with window size} \geq y}{N} \quad (2)$$

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

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

4 Resolution of the exponential delay model

4.1 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 of window size distribution.In other words, $w(y,t) = \frac{1}{N}\sum_{i=1}^{i=N}\delta(y - W_i(t))$ where $\delta()$ is the Dirac function.

In the model with processing time, this is equal to the window size density of server in processing state. 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 is given by function $\rho(x,f,t) = P(R(t) > x, W(y,t) = f(y))$, where $f(.)$ is a positive function.

4.2 The equations and the asymptotic approximation

Quantity $R(x,t)$ satisfyes the differential equation

$$\frac{\partial}{\partial t}R(x,t) = -\mu\frac{\partial}{\partial x}R(x,t) + \frac{\lambda}{N}\sum_{i=1}^{i=N} R(x+W_i(t),t) - R(x,t)\ . \tag{3}$$

Similarly equation for $w(x,t)$ is

$$\frac{\partial}{\partial t}w(y,t) = \frac{\lambda}{N}\left(R(y-1,t)w(y-1,t) + (1-R(2y,t))w(2y,t) - w(y,t)\right)\ . \tag{4}$$

But these two equations are not sufficient to describe the system since it does not contain the joint distribution of $R(t)$ and $W(t)$. For this end we should use quantity $\rho(x,W,t)$. The equation in $\rho()$ is for $x > 0$:

$$\begin{array}{rcl} \frac{\partial}{\partial t}\rho(x,W,t) & + & \lambda\rho(x,W,t) + \mu\frac{\partial}{\partial x} = \frac{\lambda}{N}\times \\ & & \times\sum_{i=1}^{i=N}\rho(x+W_i-1, W+\frac{1}{N}(Y_{W_i+1}-Y_{W_i})) \end{array} \tag{5}$$

where function Y_b is Heaviside function translated by a real number b:

$$\left\{\begin{array}{ll} Y_a(x) = 0 & \text{when } x < a \\ Y_a(x) = 1 & \text{when } x \geq a \end{array}\right.$$

When $x = 0$ we obtain the last equation:

$$\begin{array}{rcl} \frac{\partial}{\partial t}\rho(0,W,t) + \lambda\rho(0,W,t) & = & \frac{\lambda}{N}\sum_{i=1}^{i=N}\rho(W_i-1, W+\frac{1}{N}(Y_{W_i+1}-Y_{W_i}))+ \\ & & +1-\rho(2W_i, W+\frac{1}{N}(Y_{2W_i}-Y_{W_i})) \end{array} \tag{6}$$

The resolution of such equation is very intricate if not impossible with existing toolboxes. However if N tends to infinity (*i.e.* for sufficiently large N), then the limiting distribution of $W()$ and R tend to simplify as we wil decribe below.

One must observe on one hand in (3) the expression of quantity $\frac{\partial}{\partial t}R(x,t)$ contains factor μ and λ. Therefore $R(t)$ sustains a variation rate of order μ and λ. On the other hand, in (4), quantity $w(y,t)$ obviously varies at a rate N times smaller, namely $\frac{\lambda}{N}$. Therefore when N is large quantity $w(y,t)$ tends to be so slowly varying that one can assume that $R(t)$ matches exactly the steady state distribution $\tilde{R}(x,t)$ when $w(x,t)$ is fixed and does not change in time.

It will be shown in the next sub-section that the steady state $\tilde{R}(x,t) = exp(-a(t)x)$, that means exponential, and how fast the $\tilde{R}(t)$ converges to its stationary state.

4.3 Transient behaviour of $R(t)$ with fixed W distribution

We consider the side system where the distribution of W is fixed and does not change with the $R(t)$. This is not the real system since $R(t)$ and $W(t)$ are actually dependent. Therefore we denote this fake system $\tilde{R}(t)$.

Let consider an initial distribution such that $\mathrm{E}[e^{aR(0)}] < A$ for some $a, A > 0$. We also consider that $\mathrm{E}[e^{aW}] < \infty$ and $\mathrm{E}[W] > \mu/\lambda$.

We denote by $R^*(\omega)$ the Laplace transform $\mathrm{E}[e^{-\omega \tilde{R}(t)}]$. From equation 3 we get by Laplace:

$$\frac{\partial}{\partial t} R^*(\omega, t) = -\mu - \mu\omega R^*(\omega, t) + \lambda R^*(\omega, t)(\mathrm{E}[e^{\omega W}] - 1) \ . \tag{7}$$

Therefore the stationary distribution is Poisson of rate $a > 0$ such that

$$\lambda(1 - \mathrm{E}[e^{-aW}]) = \mu a \tag{8}$$

Moreover if two initial distributions $\tilde{R}_1(x, 0)$ and $\tilde{R}_2(x, 0)$ satisfies $\mathrm{E}[e^{aR(0)}] < \infty$. Therefore their transient distribution, characterized by their Laplace transform $R_1^*(\omega, t)$ and $R_2^*(\omega, t)$ tend to converge.

Namely it comes from 7 that

$$R_1^*(\omega, t) - R_2^*(\omega, t) = \exp(-(\mu\omega - \lambda(\mathrm{E}[e^{\omega W}] - 1))t)(R_1^*(\omega, 0) - R_2^*(\omega, 0)) \tag{9}$$

and the convergence is exponential of rate $\max_\omega \{\mu\omega - \lambda(\mathrm{E}[e^{\omega W}] - 1)\} > 0$.

4.4 Joint transient behaviour of $R(t)$ and $W(t)$

We assume that the distribution of $R(t)$ is exponential of parameter $a(t)$. Indeed with $R(x, t) = \exp(-a(t)x)$, equation (3) writes

$$0 = \mu a(t) \exp(-a(t)x) + \lambda(\mathrm{E}[e^{-a(t)W(t)}] - 1) \exp(-a(t)x) \ , \tag{10}$$

which implies the identity

$$\lambda(1 - \mathrm{E}[e^{-a(t)W(t)}]) = \mu a(t) \tag{11}$$

We have $\mathrm{E}[e^{-a(t)W(t)}] = \int_0^\infty w(y, t) e^{-a(t)y} dy$. Quantity $a(t)$ is therefore of function of function $w(., t)$.

The final equation is therefore, changing the scaling of variable t by a factor N: $w^N(y, t) = w(y, Nt)$:

$$\frac{\partial}{\partial t} w^N(y, t) = \lambda e^{-a(t)(y-1)} w^N(y - 1, t) + \lambda(1 - e^{-2a(t)y}) w(2y, t) - \lambda w(y) \tag{12}$$

At steady state we have therefore $w^N(y, t) = w(y)$:

$$w(y) = e^{-a(y-1)} w(y - 1) + (1 - e^{-2ay}) w(2y) \ . \tag{13}$$

with the identity

$$\frac{1 - \int_0^\infty e^{-ay} w(y) dy}{a} = \frac{\mu}{\lambda} \tag{14}$$

It is clear that the best way to solve the above equations is to use a as a parameter and then to express λ/μ as a function of a.

4.5 Window steady state distribution

Our aim is to assume that $R(t)$ is Poisson of fixed parameter a, therefore $W(t)$ behaves like a semi Markov process satisfying 12. At steady state we have 14.

We are interested into the unconditional distribution of the window size. It is interesting to consider case when $1 \gg a$. In this asymptotic case we can consider $w(y) = \sqrt{a}g(y\sqrt{a})$. Equation (13) rewrites

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

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

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

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

Simplifying we obtain the differential equation:

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

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

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

By fixing $g^*(s) = v(s)2^{s/2}\Gamma(s/2)$ we get

$$v(s) = v(s+2)(1-2^{-s-1}) \ . \tag{19}$$

The above formula is easy to solve with $v(s) = \alpha\prod_{k\geq 1}(1-2^{-s-2k+1})$. Therefore

$$g^*(s) = \alpha 2^{s/2}\Gamma(s/2)\prod_{k\geq 1}(1-4^{-k}2^{1-s}) \tag{20}$$

The value of α is extracted from the identities $g^*(1) = \int w(y)dy = 1$, and it comes

$$g^*(s) = \sqrt{\frac{1}{2\pi}}2^{s/2}\Gamma(s/2)\prod_{k\geq 1}\frac{1-4^{-k}2^{1-s}}{1-4^{-k}} \tag{21}$$

It comes that

$$\mathrm{E}[W] = \frac{g^*(2)}{\sqrt{a}} = \frac{2}{\sqrt{2\pi a}}\prod_{k\geq 1}\frac{1-4^{-k}2^{-1}}{1-4^{-k}} \tag{22}$$

and (14) becomes to

$$\frac{\mu}{\lambda} = \frac{1-\int_0^\infty e^{-ay}w(y)dy}{a} \approx \mathrm{E}[W] \tag{23}$$

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

Or by reverse Mellin it comes that

$$g(y) = \sqrt{\frac{2}{\pi}} \prod_{k\geq 1} (1-4^{-k})^{-1} \sum_{n\geq 0} a_n 2^n \exp(-4^n y^2/2) \ , \tag{24}$$

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

From the window distribution we obtain that the average packet retransmission numbers is $\frac{\mathrm{E}[W]\lambda}{\mu}$, which tends to 1 when λ tends to zero. The average number of packets dropped per window is $\frac{1}{2}(\mathrm{E}[W^2] \times a)$ which tends to $2/3$.

4.6 Discrete model

In this model we suppose that the buffer length $r = r(t)$ and the window lengths $w_i = w_i(t)$ are integers, $r = 0, 1, 2, ...,\ i = 1, 2, ..N,\ w_i = 1, 2, ...$ After addressing a window w_{i_1} at some time t the system addresses a next window w_{i_2} at time $t + \Delta t$. Each window is selected randomly, all windows are i.i.d. The time points of window addressing form a Poisson flow of intensity λ. The value r is increased by 1 at the time points that form a Poisson flow of intensity μ, and r is also changed at a time point when a window is addressed. The changes of window lengths $w_i(t)$ depend on $r(t)$:

$$w_j(t+\Delta t) = \begin{cases} w_j(t)+1 & \text{if} \quad r(t) \geq w_j(t), \\ w_{j/2} & \text{if} \quad r(t) < w_j(t). \end{cases}$$

Here we define $(2k-1)/2$ as k. At the same time moment $(t + \Delta t)$

$$r(t+\Delta t) = \max\{0, r(t) - w_j(t)\}.$$

For finite number of windows N the performance of the system is guided by a Markov chain. The factor-chain we will investigate is defined on the set of sequences $\mathbf{W} = \{W_1, W_2, ...\}$ and a variable r, where W_k is the ratio of windows of length $\geq k$, $\quad k = 0, 1, 2, ...,$ $1 = W_1 \geq W_2 \geq ...$ The distance between two sequences is defined as

$$\delta(\mathbf{W}^1, \mathbf{W}^2) = \sup_i |W_i^1, W_i^2|/i.$$

The state space of Markov chain is denumerable and as $t \to \infty$ the distribution of $(r, \mathbf{W})$ tends to the stationary distribution for any finite N.

For a smooth function $f(r, \mathbf{W})$ the generating operator $\mathbf{A}_N f(r, \mathbf{W})$ of factor-chain has the form:

$$\begin{aligned} \mathbf{A}_N f(r, \mathbf{W}) &= \mu(f(r+1, \mathbf{W}) - f(r, \mathbf{W})) \\ &+\lambda\Big(\sum_{i, i-1\leq r} (f(r, \mathbf{W} + e_i/N) - f(r, \mathbf{W}))(\mathbf{W}_{i-1} - \mathbf{W}_i) \\ &+ \sum_{i,\ i>r} f(0, \mathbf{W} - (e_{i/2+1} + ... + e_i)/N) - f(r, \mathbf{W}))(\mathbf{W}_i - \mathbf{W}_{i+1})\Big), \end{aligned} \tag{25}$$

where e_i is a vector $(0,...0,1,0,...)$ with '1' on i-th place.

Let $\tilde{w}_i$ be the ratio of windows of length i, $i = 1,2,...$ and let Δt be small. The equation (25) suggests that for $R_i = \Pr\{r = i\}$, for $\tilde{w}_k$ and for mean values of there differences Δr and $\Delta\tilde{w}_k$ we have

$$\Delta R_i = \mu(R_{i-1} - R_i) + \lambda\sum_{j=1}^{\infty}(R_{i+j} - R_i)\tilde{w}_j,\ i > 0, \tag{26}$$

$$\Delta R_0 = -\mu R_0 + \lambda\sum_{i=1}^{\infty}(\tilde{w}_i \sum_{j,\ j\le i} R_j),$$

$$\Delta\tilde{w}_k = \frac{\lambda}{N}\Big[\tilde{w}_{k-1}R_{k-1} - \tilde{w}_k + \tilde{w}_{2k}(1 - R_{2k}) + \tilde{w}_{2k-1}(1 - R_{2k-1})\Big],\ k > 1, \tag{27}$$

$$\Delta\tilde{w}_1 = \frac{\lambda}{N}\Big[-\tilde{w}_1R_1 + \tilde{w}_2(1 - R_2)\Big],$$

$$R_k = \sum_{i=k}^{\infty} r_i.$$

For finite N random variables r and $\mathbf{W}$ are dependent and we have to fined there joint distribution. But r is changing N times faster than $\mathbf{W}$, therefore for large N, $N \to \infty$, we meet two time scales and can consider fast variable r and slow variables W_i, so to say, separately. A standard approach for such kind of variable 'separation' is to fix slow variables during some time interval Δ_N. In our case Δ_N has to be small and coordinated with N so that during Δ_N the distribution of r becomes close to the stationary one (for $\mathbf{W}$ distribution fixed during time Δ_N.) After that the slow variables are changed following the new distribution of r. The limit $\Delta_N \to 0$ gives two sets of 'separeted' equations: quasi-stationar ones for R (where the coefficients do not depend on t) and non stasionary ones for W_i (where coefficiens depend on t).

In case of random variables r and W_i the validity of such approuch has to be proved, the limit equations for the mean values of variables as $N \to \infty$ have to be justified. We do not give any proves here and consider only the *limit differential equations.*

It follows from (26), (27) that for fixed *values* $\mathbf{W}$

$$\frac{dR_i(t)}{dt} = \mu\Big(R_{i-1}(t) - R_i(t)\Big) + \lambda\sum_{k=1}^{\infty}\Big(R_{i+k}(t) - R_i(t)\Big)W_k, \tag{28}$$

$$R_0(t) = 1\ R_i \to 0\ \text{ as }\ i \to \infty,\ \ \forall t \ge 0,$$

where $R_i(0)$, $i > 0$, is a known distribution. (If the number of possible values of r if finite, $r < I$, then $R_i(t) = 0$ for $i \ge I$.)

Now, for fixed *values* of R_i the equations for mean values of W_k after rescaling t are (we use for mean values the same notation W_k)

$$\frac{dW_k(t)}{dt'} = \tilde{w}_{k-1}(t)R_{k-1} - \sum_{i=k}^{2k-2}(\tilde{w}_i(t)(1-R_i)$$

$$= (W_{k-1}(t) - W_k(t))R_{k-1} - \sum_{i=k}^{2k-2}(W_i(t) - W_{i+1}(t))(1-R_i), \tag{29}$$

$$W_1 = 1,\ W_k \to 0 \ \text{ as } \ k \to \infty, \ \ \forall t \geq 0,$$

Below we consider the differential-difference equations (28), (29). It is clear that for any bounded initial values the solution $W_i(t), R_i(t)(t)$ to (28), (29) exist $\forall t, t > 0$.

It easy to prove that

a). *If initial values are decreasing in k, i so are the values of* $W_k(t)$, $R_i(t)$:

$$1 = W_1 \geq W_2(t) \geq ..., \ \ 1 = R_0(t) \geq R_1(t) \geq .... \tag{30}$$

Really, it is sufficient to consider initial conditions where

$$R_i > R_{i+1}, \ \ W_i > W_{i+1}. \tag{31}$$

Suppose, for example, that for some $t_0 > 0$ and for some i an equality $R_i(t_0) = R_{i-1}(t_0)$ takes place for the first time. It follows from (28) that $dR_i(t_0)/dt > dR_{i+1}(t_0)/dt$, and that makes the equality $R_i(t_0) = R_{i+1}(t_0)$ impossible. The proof for W_k is similar.

It can also be shown that

b). *if* $\sum_i R_i(0) < \infty$ *then* $\sum_i R_i(t) < \infty \ \ \forall t, \ t > 0$;

c). *if* $\sum_k W_k(0) < \infty$ *then* $\sum_k W_k(t) < \infty \ \ \forall t, \ t > 0$.

Let us show that

d) *for fixed* $\mathbf{W}$ *and any initial conditions* $R_i(0)$ *the solution to differential equations* (28) *tends to stationary solution.*

Really, let $R_i^{(1)}(t)$ and $R_i^{(2)}(t)$ be two solutions to (28). The equations (28) being linear differences $\tilde{R}_i(t) = R_i^{(1)}(t) - R_i^{(2)}(t)$ satisfy the same equations with zero boundary values. Let i_0 be the smallest index for which $\tilde{R}_{i_0}(t) = \max_i |\tilde{R}_i(t)|$. Than by (28) $\frac{dR_{i_0}(t)}{dt} < 0$ if $\tilde{R}_{i_0}(t) > 0$ and $\frac{dR_{i_0}(t)}{dt} > 0$ if $\tilde{R}_{i_0}(t) < 0$. Thus $\max_i |\tilde{R}_i(t)| \to 0$ as $t \to \infty$.

For fixed $\mathbf{W}$ *there exist a stationary solution* $\mathbf{R}_{st}$ *to* (28). It has the form $R_i = b^i$, where b is defined by the equation

$$\mu b^{i-1}(1-b) = \lambda b^i(1-b)\sum_{k=1}^{\infty} b^k W_k, \tag{32}$$

$$\frac{\mu}{\lambda} = \frac{\sum_{j=1}^{\infty} b^j W_j}{b}. \tag{33}$$

(Compare with the continuous case !)

Because of the statement d) this stationary solution is unique and stable.

To show the existence and uniqueness of the stationary solution to (29) is more complicated.

Consider first equations (29) for $1 \leq k \leq K$, supposing that $W_k^{(K)} = 0$ for $k > K$. The stationary solution satisfies equations

$$(W_{k-1}^{(K)} - W_k^{(K)})R_{k-1} - \sum_{i=k}^{2k-2}(W_i^{(K)} - W_{i+1}^{(K)})(1 - R_i) = 0, \tag{34}$$

$$W_1^{(K)} = 1,\ W_k^{(K)} = 0 \ \text{ for } \ k > K \ \ \forall t \geq 0.$$

The value of $W_K^{(K)}$ determines all $W_k^{(K)}$, $k < K$. If $W_K^{(K)} = 0$ then all $W_k^{(K)} = 0$, if $W_K^{(K)} = 1$ then all $W_k^{(K)} > 1$, $k < K$. As $W_k^{(K)}$ depend continuously on $W_K^{(K)}$ there exist such $W_K^{(K)}$ for which $W_1^{(K)} = 1$. Let $K \to \infty$. There exist a subsequence $\mathbf{W}^{K_j}$ that converges to the stationary solution to (29)

Let us give the *sufficient* conditions for uniqueness of solution to (34).

It is easy to see that the solution to (29) is majorized by solution of equations

$$\frac{du_k(t)}{dt'} = (u_{k-1}(t) - u_k(t)R_{k-1} - (u_k(t) - u_{k+1}k(t)(1 - R_k)$$

with the same initial conditions. R_k decrease exponentially (for exponentially decreasing initial values of $R_k(0)$), therefore u_k decreases with k superexponentially (at least if $u_k(0)$ decreases superexponentially.) Thus W_i decrease superexponentially.

For fixed $R_k = b^k$ consider $|W_k^{(1)}(t) - W_k^{(2)}(t)|$, where $W_k^{(1)}(t), W_k^{(2)}(t)$ are solutions for (29) with different initial values.

The differences $\tilde{W}_i(t) = W_i^{(1)}(t) - W_i^{(2)}(t)$ satisfy equations (29) with zero boundary values. The sum $\sum_k \nu^k W_k$, $\nu > 1$, converges. Does it tend to 0 as $t \to \infty$? We have

$$\frac{d\sum_k \nu^k \tilde{W}_k(t)}{dt'}$$

$$= -\nu^2 \tilde{W}_2 + \sum_{k=3}^{\infty} \nu^k((\tilde{W}_{k-1}(t) - \tilde{W}_k(t))R_{k-1} - \sum_{i=k}^{2k-2}(\tilde{W}_i(t) - \tilde{W}_{i+1}(t))(1 - R_i)).$$

Remembering that from $du/dt = -u + a$ follows $d|u|/dt \leq -|u| + |a|$ we can rewrite the last sum (turning to the adjoint equations for $\nu_k = \nu^k$):

$$\frac{d\sum_k \nu^k |\tilde{W}_k(t)|}{dt'}$$

$$\leq -\nu^2|\ \tilde{W}_2(t)| + \sum_{k=3}^{\infty} |\tilde{W}_k(t)|[\nu^k(-b^{k-1} + \nu b^k - (1 - b^k)$$

$$+(1-b^{k-1})/\nu) + (\nu^{k-1} + \nu k - 2 + ...)(b^{k-1} - b^k) + \nu^{k/2}(1-b^{k/2-1}]$$
$$\leq -\nu^2|\tilde{W}_2(t)| + \sum_k |\tilde{W}_k(t)|\Big[\nu^k b^{k-1}(-1+\nu b - (1-b^k)$$
$$+(1-b^{k-1})/\nu + (b^{k-1}-b^k)/(\nu-1)) + \nu^{k/2}(1-b^{k/2-1}\Big] \quad (35)$$

If for given b there exist such ν that the square brackets in (35) are negative then $\tilde{W}_k(t)| \to 0$ as $t \to \infty$.

Apparently, as $N \to \infty$ the performance of Markov chain tends to performance of solutions of differential equations with 'separated' variables. Note, that the dynamical system (29) describes the behavior of probability distributions concentrated on a single sequence - the solution of boundary-value problem. Compare, for example, with similar asymptotic behavior of a system presented in [6] The behavior of an individual window length that may oscillate is not visible in our consideration

Note also that the random selection of the windows is essential here. If the windows are selected in sequential order the dependence of neighboring window lengths may bring different results.

5 The realistic model with random processing time

In the previous model we assumed that sources had to wait for a random exponential delay between windows transmissions. This is not realistic because the sources have to wait for the queueing delay in buffer before receiving feedback. The queueing delay is not exponential. In our asymptotic model the queueing delay is always close to $(B - R(t))/\mu \approx B/\mu$, *i.e.* the delay is approximately equal to the time needed to empty a full buffer. To this delay we can add a random propagation delay and a random processing delay. To simplify the presentation we call the random component the processing time.

A convenient way to model this is to consider that the source wait for an incompressible delay NT and then waits for a random exponential delay of parameter λ/N.

The equations about $R(x,t)$ remains the same:

$$\frac{\partial}{\partial t}R(x,t) + \mu\frac{\partial}{\partial x}R(x,t) = \lambda\int_0^\infty (R(x+y,t) - R(x,t))w(y,t)dy \ . \quad (36)$$

Quantity $Nw(y,t)$ is still the number of sources waiting for exponential delay with window size equal to y. But since the number of sources waiting for exponential delay is now smaller than N, we now have $\int_0^\infty w(y,t)dy < 1$.

The equation in $w(y,t)$ is now different and:

$$\begin{aligned}\tfrac{\partial}{\partial t}w(y,t) + \tfrac{\lambda}{N}w(y,t) &= \tfrac{\lambda}{N}(R(y-1,t)w(y-1,t-NT)+ \\ &\quad +(1-R(2y,t-NT))w(2y,t-NT)) \ .\end{aligned} \quad (37)$$

We now notice that we have a delayed differential equation.

With the same line of reasoning as in the previous section, $R(x,t)$ varies N times faster than $w(y,t)$ and therefore we can assume that $R(x,t)$ is on stationary distribution: $R(x,t) = \exp(-a(t)x)$. The quantity $a(t)$ satisfies the identity:

$$\frac{\lambda}{a(t)} \int_0^\infty w(y,t)(1-e^{-a(t)y})dy = \mu \ . \tag{38}$$

Going further we also have $\lambda \mathrm{E}[\int_0^\infty w(y,t)dy] = \frac{\lambda}{\lambda T+1}$, expressing that the average windows arrival rate in buffer is $\lambda/(\lambda T+1)$.

Equation (37) after rescaling $w^N(y,t) = w(y,Nt)$ becomes now:

$$\begin{aligned} \tfrac{\partial}{\partial t} w^N(y,t) + \lambda w^N(y,t) &= \lambda e^{-a(t-T)(y-1)} w^N(y-1,t-T) + \\ &\quad + \lambda(1-e^{-2a(t-T)y}) w(2y,t-T) \end{aligned} \tag{39}$$

In the asymptotic case where $a(t) << 1$ we can assume that $w^N(y) \approx \frac{\mu}{\lambda} \frac{a(t)}{g^*(2)} g(y\sqrt{a(t)})$, function $g(y)$ and $g^*(s)$ being described in the previous section. The reason for this assumption is that $\lambda \int \frac{1-e^{-a(t)y}}{a(t)} w(y,t)dy \approx \lambda \int y w(y,t)dy = \mu$. Keeping first order we get

$$\frac{\partial a(t) g(y\sqrt{a(t)})}{\partial t} + \lambda a(t) g(y\sqrt{a(t)}) = \lambda a(t-T) g(y\sqrt{a(t-T)}) \tag{40}$$

The solution of retarded differential equation of the kind $f'(t) + \lambda f(t) = \lambda f(t-T)$ is known to converge to a constant when t tends to infinity. Therefore the limit of $w^N(y,t)$ should be $a g(y\sqrt{a})$ where a can be computed via the identity: $\mathrm{E}[\int_0^\infty w(y,t)dy] = \frac{1}{\lambda T+1}$. It comes the expression for a:

$$\sqrt{a} = \frac{\lambda}{\lambda T+1} \frac{g^*(2)}{\mu} \tag{41}$$

We remind that $g^*(2) \approx 0.97206$.

Taking the system parameter, this expression becomes:

$$\sqrt{a} = \frac{N}{E[\mathrm{RTT}]} \frac{g^*(2)}{\mu} \ . \tag{42}$$

In fact formula (42) is not enough accurate when $\sqrt{a}$ is not enough small. In this case the implicit equation

$$\int_0^\infty \frac{1-e^{-\sqrt{a}y}}{\sqrt{a}} g(y)dy = \mu \frac{\lambda T+1}{\lambda} \sqrt{a} \ , \tag{43}$$

will more accurate. Or in system parameter:

$$\int_0^\infty \frac{1-e^{-\sqrt{a}y}}{\sqrt{a}} g(y)dy = \mu \frac{E[\mathrm{RTT}]}{N} \sqrt{a} \ , \tag{44}$$

An interesting behaviour of the delayed equation is that $f(t)$ may experience periodic oscillations before converging to its seady state. The amplitude of the oscillation exponentially decays. The rate at which the amplitude decays is equal to $1/T$ multiplied by the real

part of the smallest non-zero root of $\frac{1-e^z}{z} = 1/(\lambda T)$. The period of the oscillation is equal to $2\pi T$ divided by the imaginary part of this root.

When λT is large the real part of the root is $1/(2(\lambda T)^2)+O(1/(\lambda T)^3)$, and its imaginary part is $2\pi+O(1/(\lambda T))$. This results makes the oscillation period of $a(t)$ close to TN and the number of oscillations before attenuation of order $2(\lambda T)^2$. If λT is large then the oscillations lasts for the duration of the connections. Quantity λT is the difference between coefficient of variation of the propagation delay minus 1. The coefficient of variation is equal to the mean divided by the standard deviation.

5.1 Model limitations

This model has several limitations. The main limitation is the mean field approximation in identity (37) which needs a large number of connection in random processing time at every time t to be valid. This number is $N\int_0^\infty w(y,t)dy \approx \frac{N}{\lambda T+1}$ or in system parameter $N \times C$ where C is the coefficient of variation of the round trip delay. We call *mean field parameter* the quantity NC.

If the mean field parameter condition is not satisfied, then two consequences can be expected:

1. the analytical derivations are not accurate;
2. the system may have large fluctuations if the law of large number cannot apply.

If the system experiences large fluctuations, then the hypothesis that at each moment the input rate is always larger than the buffer output rate μ, may also fail. This condition, *i.e.* $\lambda\int_0^\infty yw(y,t)dy > \mu$ always hold, is necessary in order that the equation

$$\lambda\int_0^\infty \frac{1-e^{-ay}}{a}w(y,t)dy \tag{45}$$

be satisfied by a non negative a. However it may happen that this condition cannot be fulfilled every time. Let call $\lambda(t) = \lambda\int_0^\infty yw(y,t)dy$, the instantaneous input rate. When the instantaneous input rate fails to be larger than μ at time t, we are in a kind of transient state where, as long $\int_{t_0}^t(\mu-\lambda(\theta))d\theta > 0$, we have

1. $E[R(t)] = \int_{t_0}^t(\mu-\lambda(\theta))d\theta$;
2. Factor $a(t)$ can be considered as equal to zero in the equation of evolution (*i.e.* the windows always increases).

Notice that the condition $\lambda(t) > \mu$ is not sufficient to leave the transient state, since $E[R(t)]$ can still be very large. Notice that $E[R(t)] = O(N)$ during the transient interval, since $\lambda(t)$ varies at a peace $O(1/N)$. The event which ends the transient interval is the first congestion event, or when $\int_{t_0}^t(\mu-\lambda(\theta))d\theta = 0$, in the present model.

Via very simple analysis it comes that when $t > t_0 + T$, quantity $\lambda'(t)$, the first derivative of λ, evolves like

$$\lambda'(t) + \frac{\lambda}{N}\lambda(t) = \frac{\lambda}{N}\lambda(t - NT) + \frac{\lambda^2}{N}\int_0^\infty w(y, t - NT)dy \tag{46}$$

Since the average of $\lambda \int_0^\infty w(y,t)dy$ is $\frac{\lambda}{\lambda T+1}$ we can approximate the above equation with

$$\lambda'(t) + \frac{\lambda}{N}\lambda(t) \approx \frac{\lambda}{N}\lambda(t - NT) + \frac{\lambda^2}{N}\frac{1}{\lambda T + 1} \tag{47}$$

it comes the obvious linear expression

$$\lambda(t) \approx \frac{t - t_0}{N}\frac{\lambda^2}{(\lambda T + 1)^2} \ . \tag{48}$$

Within this expression it comes that

$$\int_{t_0}^{t} (\mu - \lambda(\theta))d\theta \approx (\mu - \lambda(t_0))(t - t_0) - \frac{(t - t_0)^2}{2N}\frac{\lambda^2}{(\lambda T + 1)^2} \tag{49}$$

and the transient period stopping point t_0 has expression:

$$t_1 = t_0 + 2N\frac{(\lambda T + 1)^2}{\lambda^2}(\mu - \lambda(t_0)) \tag{50}$$

and $\lambda(t_1) = 2\mu - \lambda(t_0)$ which is larger than μ.

If T is large, the average window size will tend to be large. Therefore between t_1 and $t_1 + T$ all windows will be very likely be halved, leading to $\lambda(t_1 + T) \approx \lambda(t_1)/2$. Since $\lambda(t_1)/2 < \mu$, $t_1 + T$ will be the beginning of a transient period. Notice that $\lambda(t_0) = \frac{2}{3}\mu$ is the fixed point of this operation.

The simulations show that the existence of oscillation in $R(t)$ in such a conditions well in agreement with the fact that the period of oscillation should be close to $\frac{2N}{3}\frac{(\lambda T+1)^2}{\lambda^2}\mu$. In term of the system parameters this period has expression

$$(E(\mathrm{RTT}))^2 * \frac{2\mu}{3N} \ . \tag{51}$$

The amplitude is close to

$$(E(\mathrm{RTT}))^2 * \frac{\mu^2}{9N} \ . \tag{52}$$

6 Simulation and interpretation

6.1 Simplified TCP

We have simulated the simplified TCP with different number of connections and different buffer size and random delay. By simplified TCP we mean the approximated version of TCP we have analysed in the paper, *i.e.*:

- windows are transmitted and acknowledged by batches;
- window sizes are real numbers;
- connection never come again in slow start.

Each of the connections start at a random time after time $t = 0$. 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. The simulation runs are enough long so that every connection has left its slow start mode and that the process has attained its stationary state to be compared with analytical results. In all simulations of simplified TCP, the buffer service rate μ is 1 packet per time unit.

6.2 Real TCP

We have simulated the real TCP using the ns2 simulation tool of [5]. The used TCP version is Reno. The link from the buffer to the clients is 8 Mbps. The packet size is 1K octets. Each server has a private link to the buffer at 1 Gbps (see figure 1).

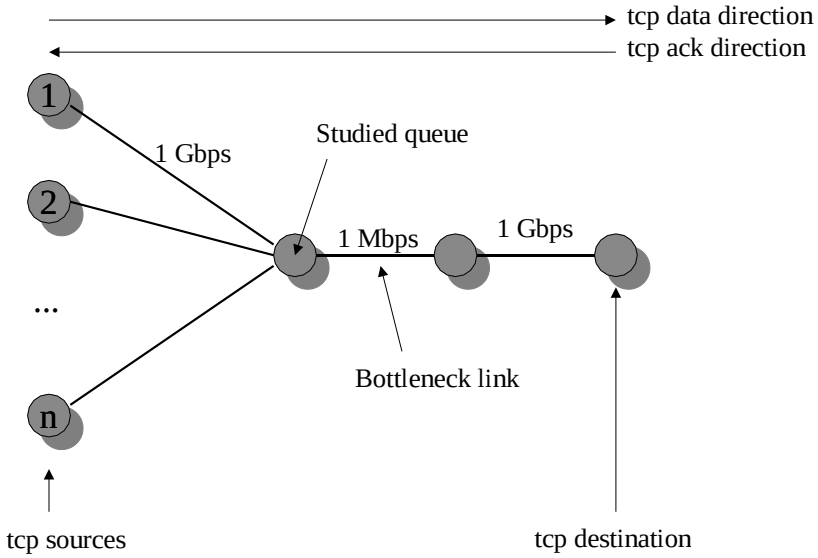

Figure 1: ns2 simulation diagram for TCP Reno

6.3 Histograms of TCP connections

6.3.1 Simplified TCP

Figure 2, 3 and 4 show the histogram of the buffer occupancy during a certain interval of time. Figure 2 shows the buffer occupancy when the number of connections $N = 100$ and the buffer size is $B = 100$. Figure 3 shows the buffer occupancy when the number of connections $N = 100$ and the buffer size is $B = 300$. Figure 4 shows the buffer occupancy when the number of connections $N = 300$ and the buffer size is $B = 800$. In all three pictures, the average processing time is 10. We also added a waterline which indicates the average buffer occupancy computed with the asymptotical formula of the section about window size distribution. For figure 2, 3 and 4 the mean field parameter (*i.e* the average number of servers in processing time) is respectively 10, 3.2 and 3.7. It seems that the two last figure are two small to have accurate matching with analytical results.

Figures 5 and 6 shows simulations results which are well outside the validity of the model. Indeed the mean field parameter is only 1.2 and 2.5, respectively.

Within this context oscillating behaviour is not surprising. The periods and amplitudes of these oscillation are in fair agreement with the tentative analysis and are inversely proportional to the number of connections. Figure 5 shows the buffer occupancy when the number of connections $N = 100$ and the buffer size is $B = 800$. Figure 5 shows the buffer occupancy when the number of connections $N = 200$ and the buffer size is $B = 800$. In both pictures, the average procesing time is 2.

6.3.2 TCP Reno

The time unit is the millisecond. Various connection numbers and buffer sizes are considered in figures 7, 8, 9 and 10.

6.4 Buffer ocupancy distribution

6.4.1 Simplified TCP

Figures 11 and 12 display the buffer occupancy distribution obtained via simulation. On each plot the scale is logarithmic and the straight line shows the theoretical exponential repartition function with the rate computed from limiting formula (44). In figure 11 there 100 connections with a buffer of size 100, with random processing time 10. In figure 12, the parameters are the same excepted that the average processing time is 100.

6.4.2 TCP Reno

Figures 14, 15 and 16 display the window distribution throughout the simulations. In each simulations, the buffer size has been sampled and the sample set fas been ranked from the smallest to the largest. The rank of size x divided by the number of samples gives the estimated probability $P(R(t) > x)$. Therefore a linear look in log scale outlines an exponential behavior. In every figure the number of samples is 35,000.

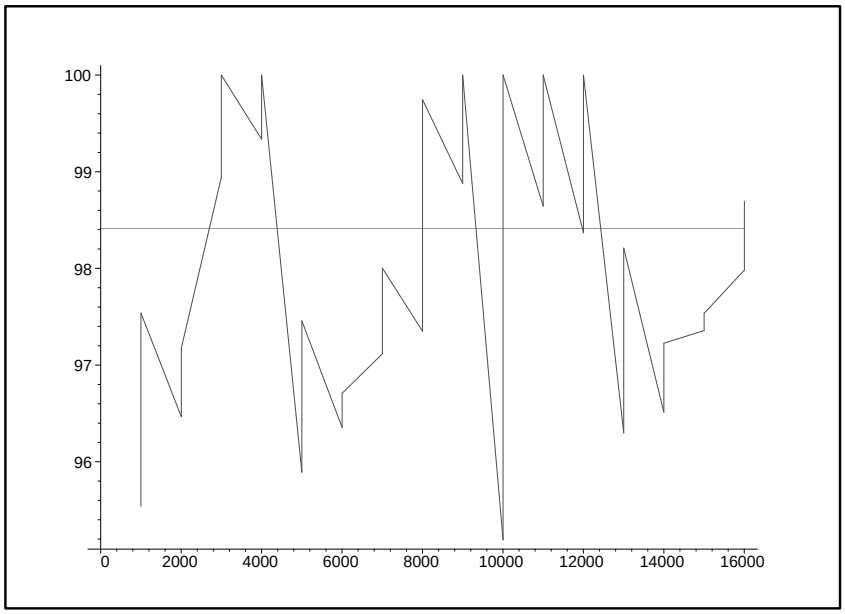

Figure 2: Histogram of buffer occupancy versus time for simplified TCP, for 100 connections, buffer size 100, buffer service rate $\mu = 1$, average random processing time 10. The dashed waterline is the theoretical average buffer size computed from the limiting approximation (44.

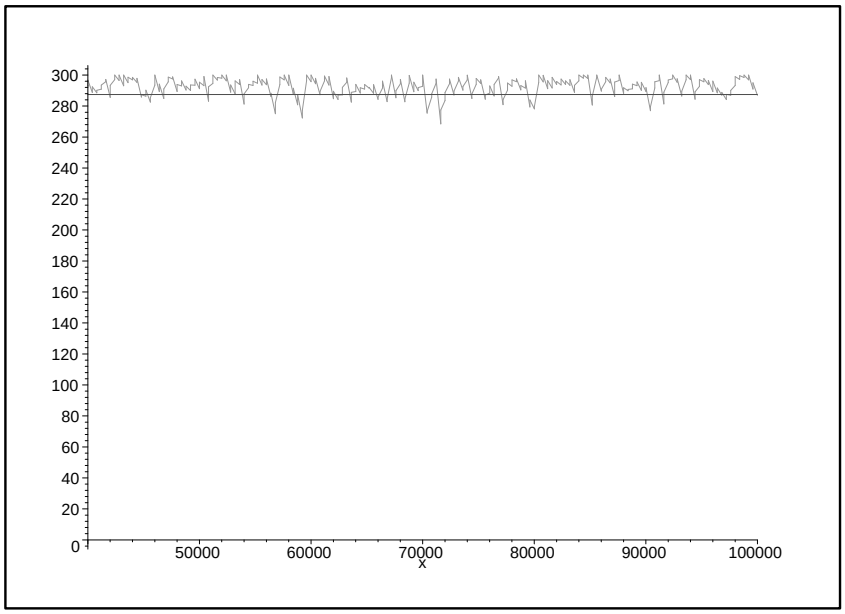

Figure 3: Histogram of buffer occupancy versus time for simplified TCP, for 100 connections, buffer size 300, buffer service rate $\mu = 1$, average random processing time 10. The dashed waterline is the theoretical average buffer size computed from the limiting approximation (44.

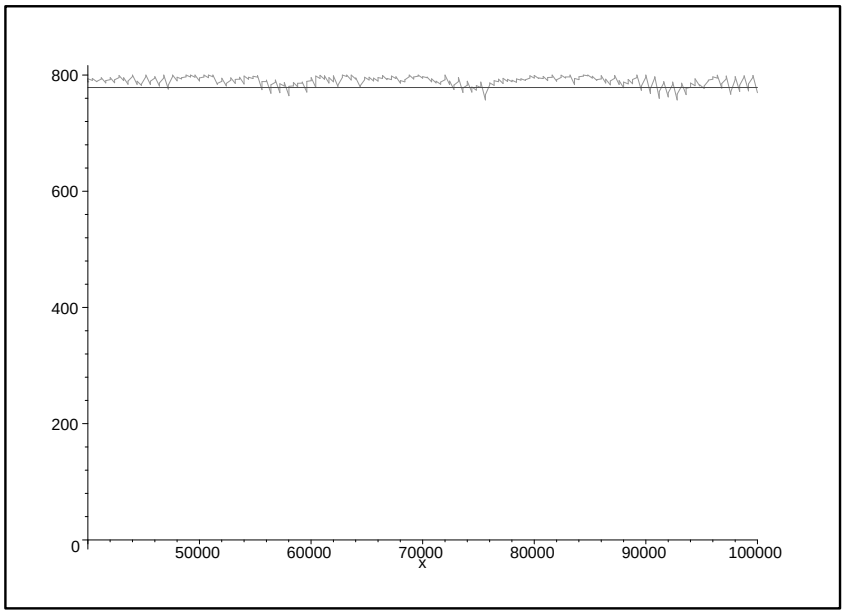

Figure 4: Histogram of buffer occupancy versus time for simplified TCP, for 200 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10. The dashed waterline is the theoretical average buffer size computed from the limiting approximation (44.

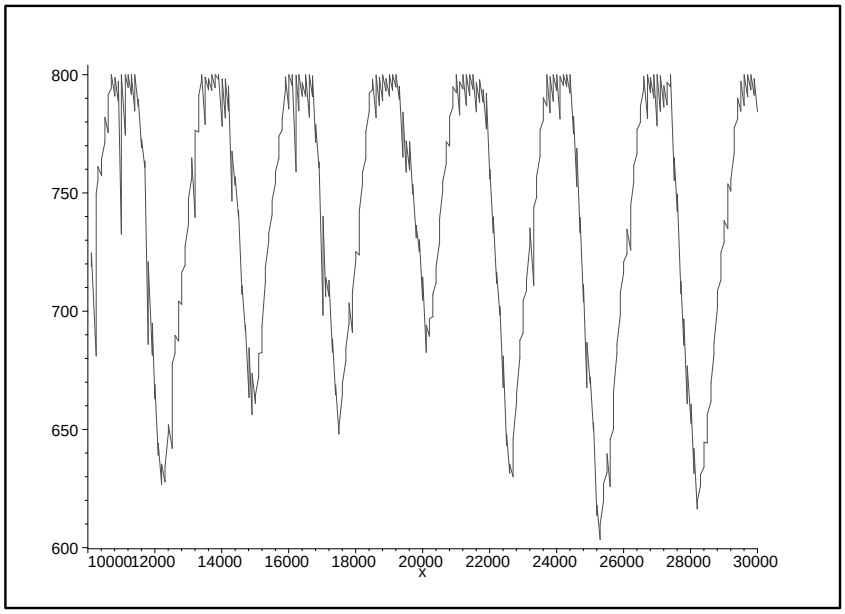

Figure 5: Histogram of buffer occupancy versus time for simplified TCP, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 2.

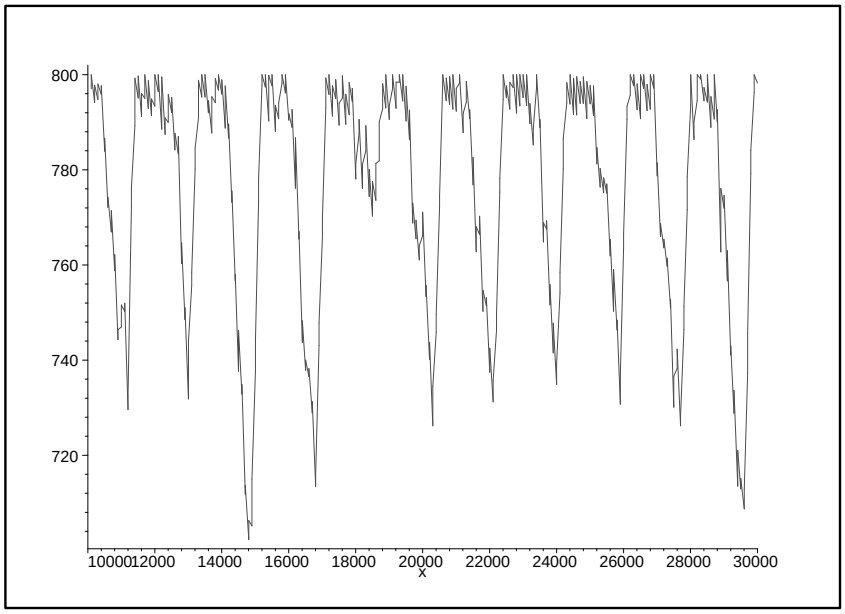

Figure 6: Histogram of buffer occupancy versus time for simplified TCP, for 200 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 2.

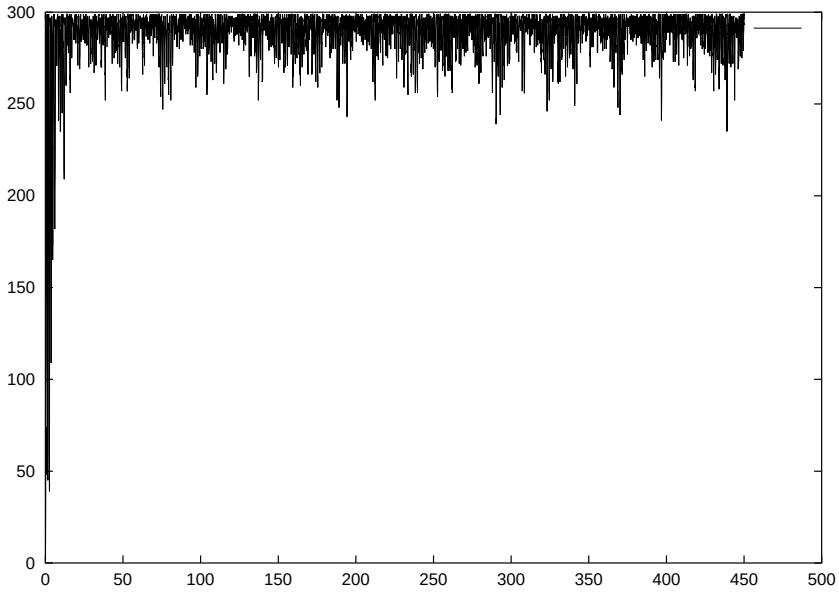

Figure 7: Histogram of buffer occupancy versus time for TCP Reno, for 100 connections, buffer size 300, buffer service rate $\mu = 1$, average random processing time 10.

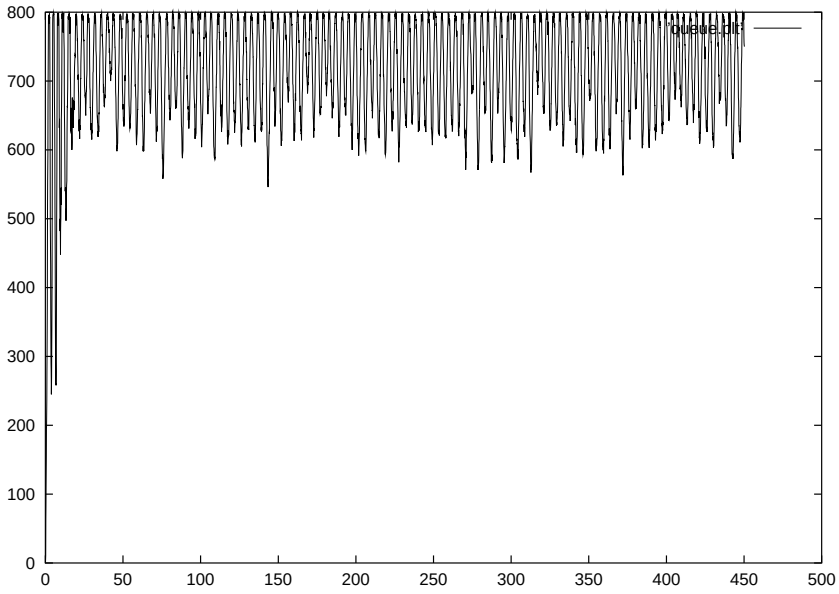

Figure 8: Histogram of buffer occupancy versus time for TCP Reno, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

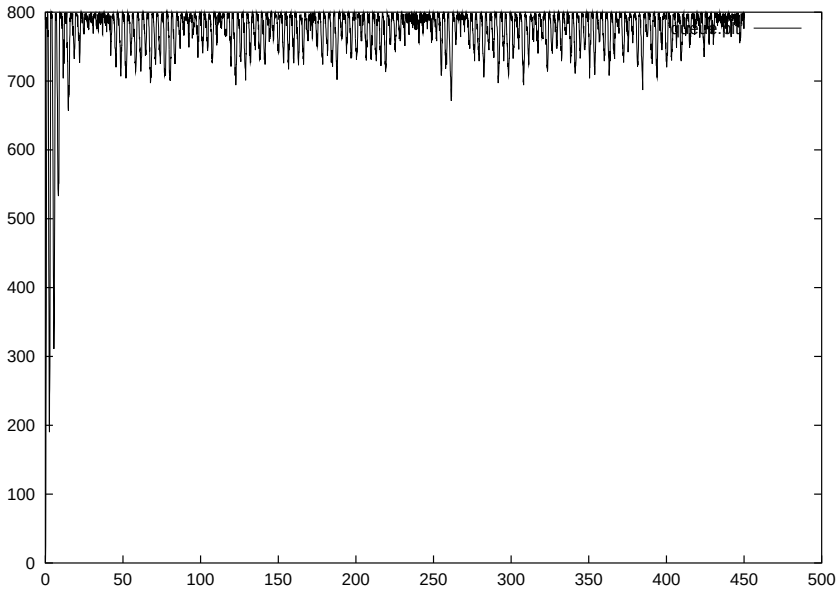

Figure 9: Histogram of buffer occupancy versus time for TCP Reno, for 800 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

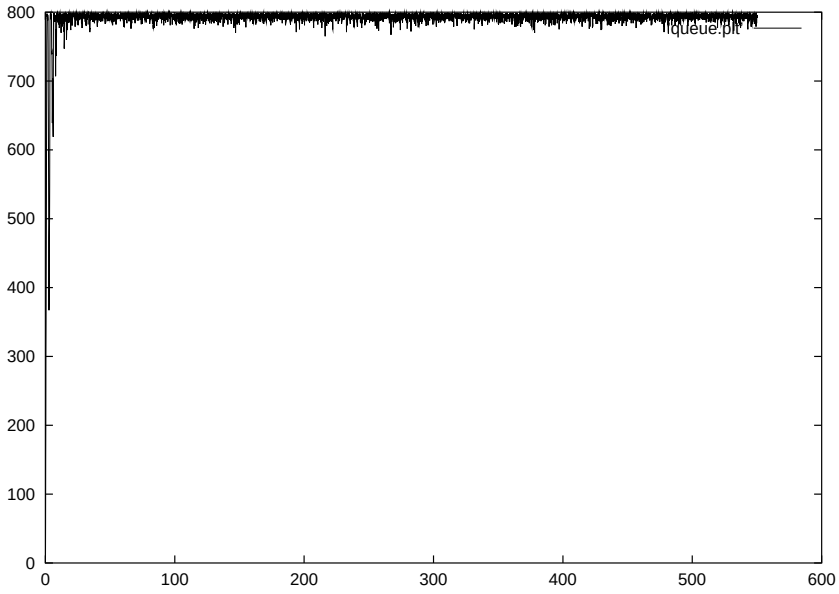

Figure 10: Histogram of buffer occupancy versus time for TCP Reno, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

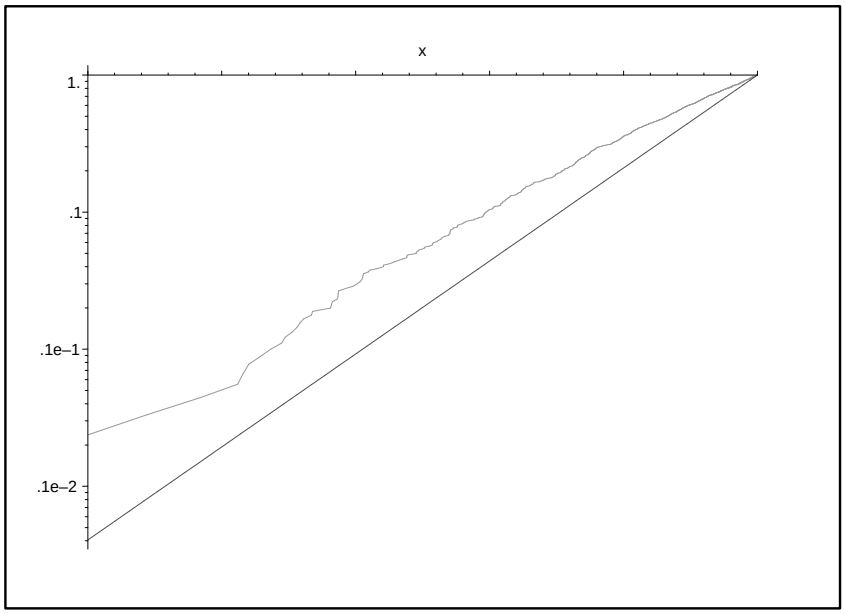

Figure 11: Repartition function of buffer occupancy, simplified TCP, for 100 connections, buffer size 100, buffer service rate $\mu = 1$, average random processing time 10. The straight line is the theoretical exponential distribution computed via the approximated formula (44).

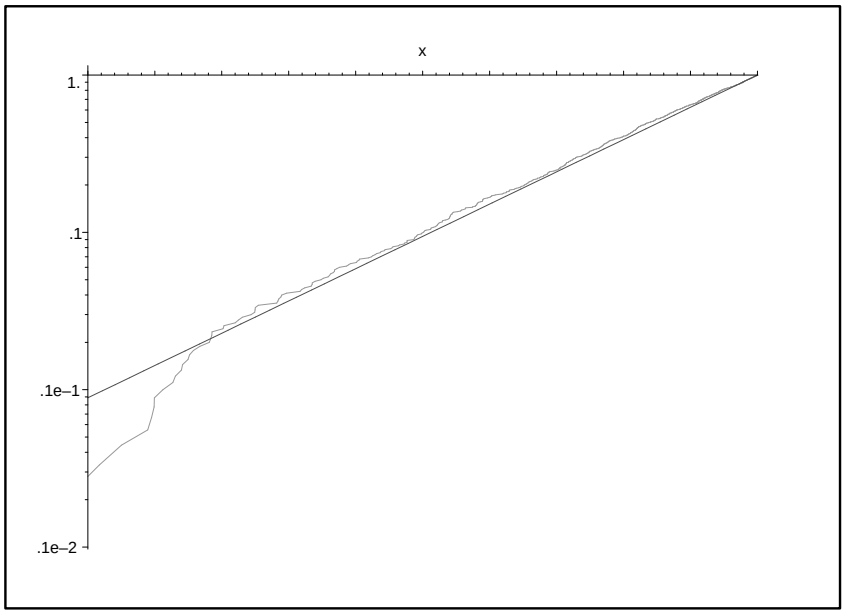

Figure 12: Repartition function of buffer occupancy, simplified TCP, for 100 connections, buffer size 100, buffer service rate $\mu = 1$, average random processing time 100. The straight line is the theoretical exponential distribution computed via the approximated formula (44).

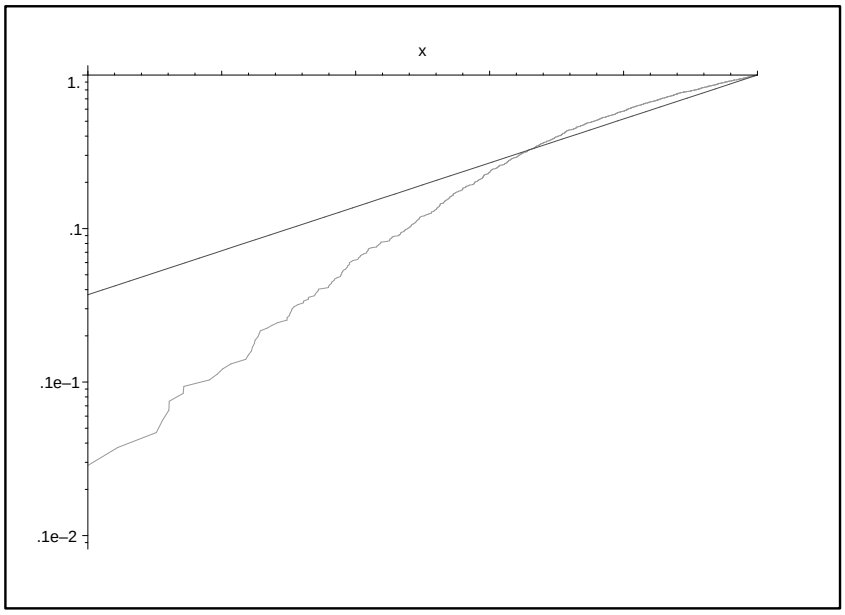

Figure 13: Repartition function of buffer occupancy from buffer location 700 to 800, simplified TCP, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 100. The straight line is the theoretical exponential distribution computed via the approximated formula (44).

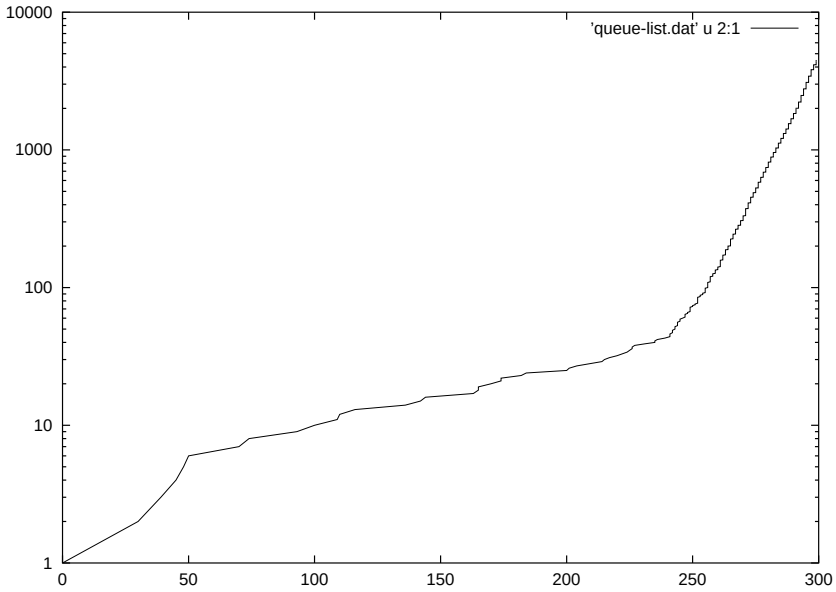

Figure 14: Repartition function of buffer occupancy, simplified TCP, for 100 connections, buffer size 300, buffer service rate $\mu = 1$, average random processing time 10.

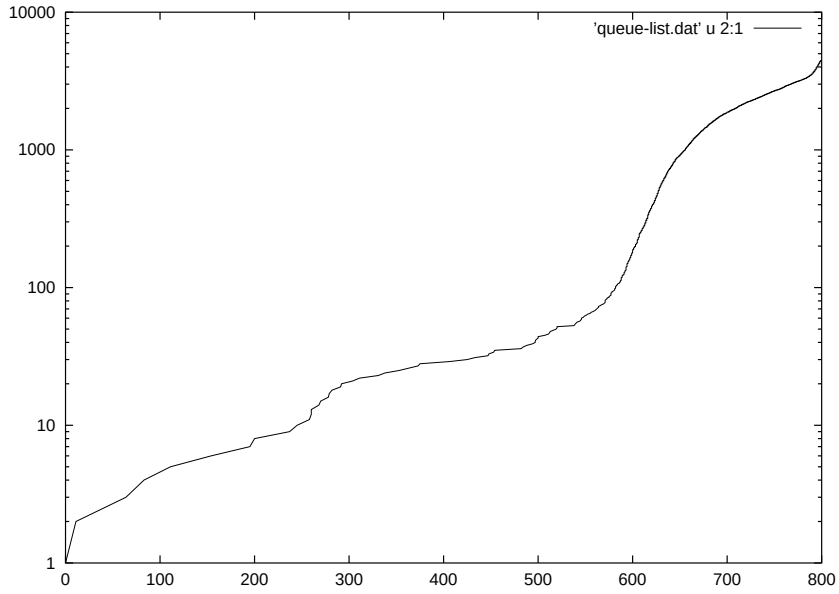

Figure 15: Repartition function of buffer occupancy, simplified TCP, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

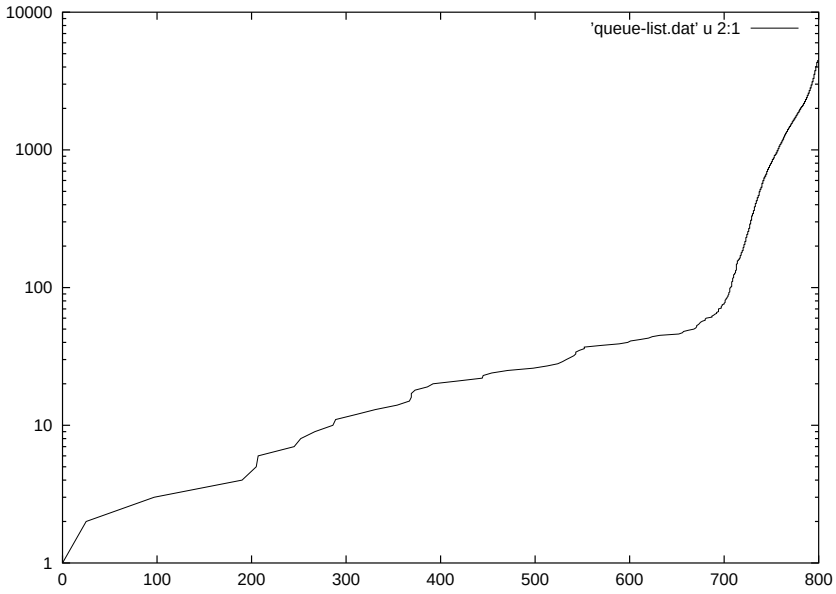

Figure 16: Repartition function of buffer occupancy, simplified TCP, for 200 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

1. $\frac{E[T_{\text{packet}}]}{N}$ tends to infinity when $N \to infty$;
2. the distribution of T_{packet}/N is heavy tailed.

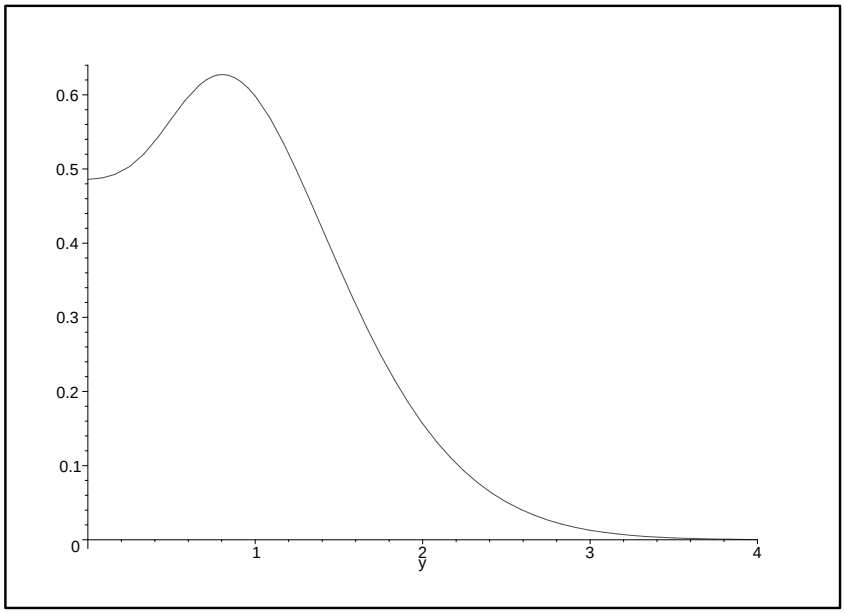

Figure 17: Theoretical function $g(x)$ for window size distribution.

6.5 Window size distribution

6.5.1 Theoretical plots

Figure 17 displays the theoretical function $g(x)$ and figure 18, the theoretical primitive of $g(x)$. Figure 19 gives the quantity $E[\text{RTT}]/N$ *versus* $\sqrt{a}$ according to (44).

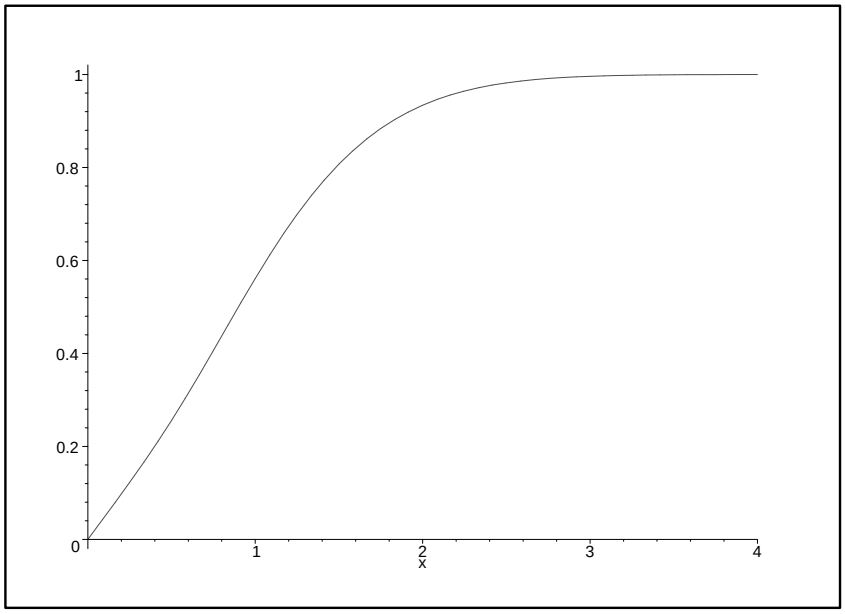

Figure 18: Primitive theoretical of function $g(x)$ for window size distribution.

6.5.2 Simplified TCP

Figures 20, 22 display the simulated window distribution. The distribution is obtained after having frozen the simulation at a certain time. Notice that non-integer values are attainable. Figure 22 also displays the theoretical distribution using function $g(x)$ (the smooth plot).

6.5.3 TCP Reno

6.6 Fairness analysis

The theoretical window size distribution and the simulated window size are in fair agreement, as we can see the theoretical distribution as the limiting distribution when N tends to infinity. When we observe the distribution of window, interesting stuff are within small windows. Indeed the distribution of small windows is a good unfairness indicator since the

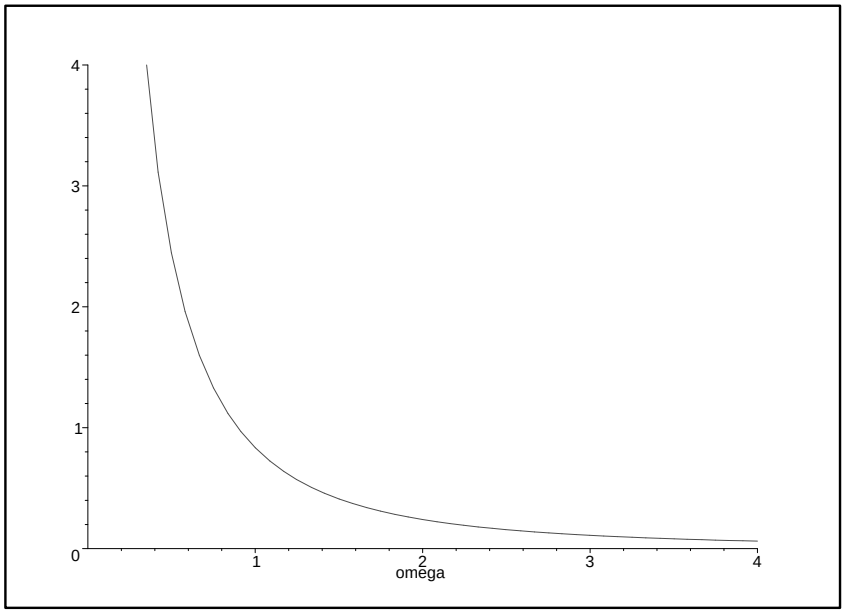

Figure 19: Theoretical $E[\mathrm{RTT}]/N$ versus $\sqrt{a}$.

instantaneous throughput of a server at time t is $\frac{W_i(t)}{E[\mathrm{RTT}]}$, where $W_i(t)$ is the size of the window of the server at time t. When a server has a small window it will very likely stay a long time with a small window. A good indicator of unfairness is also the distribution of $T_{\mathrm{packet}}(t)$, the time needed to transmit one packet at time t. Clearly $T_{\mathrm{packet}}(t) = \frac{E[\mathrm{RTT}]}{W_i(t)]}$ for the server with window size $W_i(t)$.

From the analysis of window size distribution we have $P(W_i(t) < y) \approx \sqrt{a}g(0)y$. Since $g(0) > 0$ (equal to $g^*(2)/2$ approximately 0.48) there is a non zero weight toward small windows. Using (44) we get

$$P(W_i(t) < y) \approx \frac{N}{E[\mathrm{RTT}]}(g^*(2))^2 y \ . \tag{53}$$

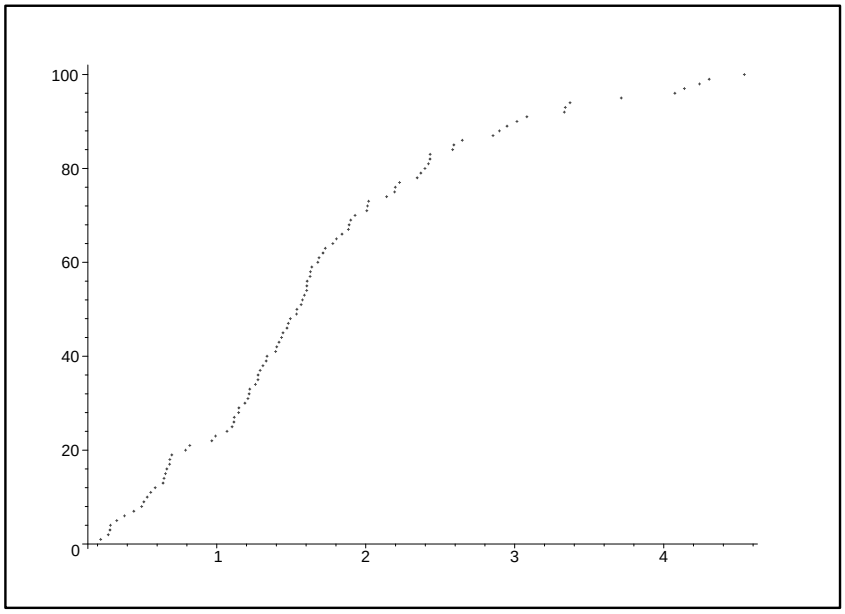

Figure 20: Window size distribution frozen at a random time for simplified TCP, for 100 connections, buffer size 100, buffer service rate $\mu = 1$, average random processing time 10.

Therefore

$$P(T_{\text{packet}} > y) \approx N \frac{(g^*(2))^2}{2} y^{-1} \ . \tag{54}$$

It comes that the time to transmit packet has an heavy tail distribution and an infinite mean. This is a fair indication that this model of TCP implies medium term unfairness between users.

References

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

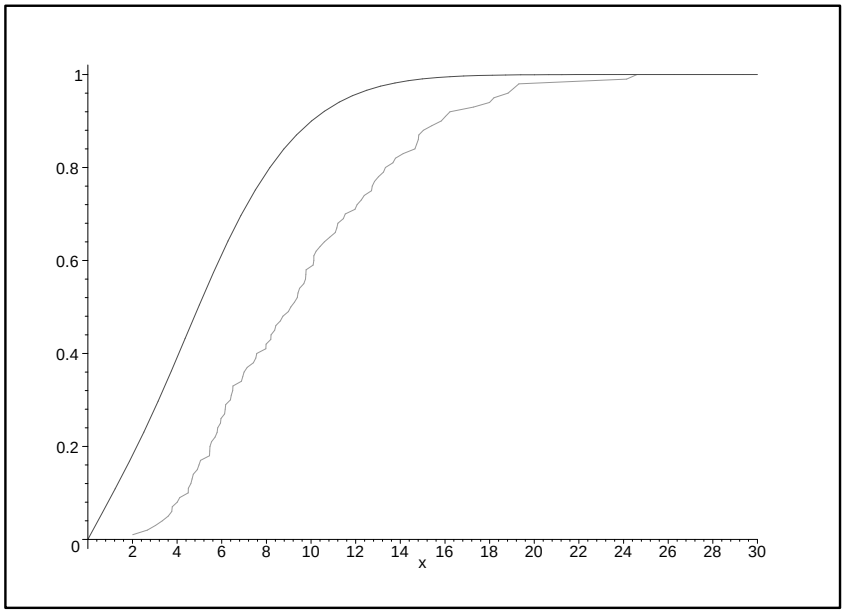

Figure 21: Window size distribution frozen at a random time for simplified TCP, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 100. Theoretical plot added

[2] F. Baccelli, Dohy 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

[3] T. Ott and J. Kemperman and M. Mathis, *The stationary behavior of ideal TCP congestion avoidance.* ftp://ftp.bellcore.com/pub/tjo/TCPwindow.ps

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

[5] UCB/LBNL/VINT Network Simulator - ns (version 2) http://www.isi.edu/nsnam/ns/, 2001

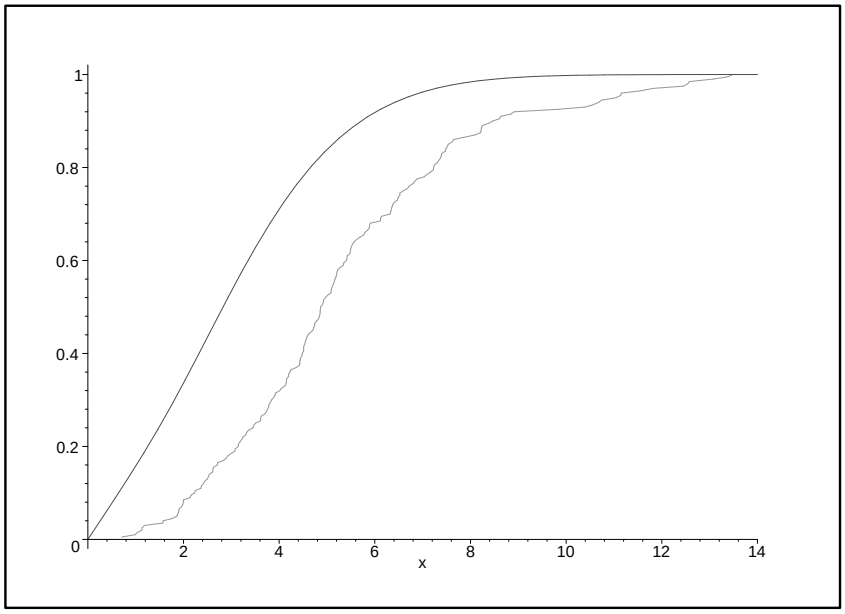

Figure 22: Window size distribution frozen at a random time for simplified TCP, for 200 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 100. Theoretical plot added

[6] 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.

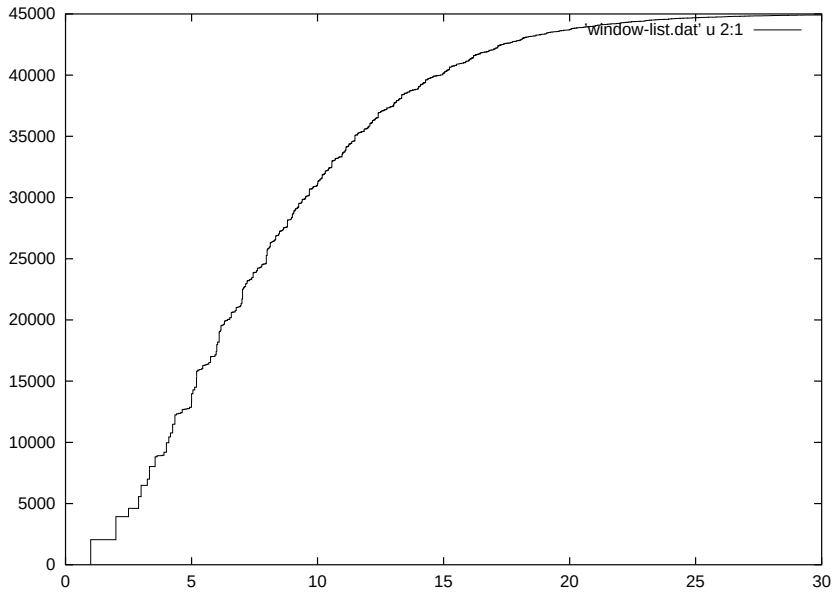

Figure 23: Window size distribution sampled at periodic time for TCP Reno, for 100 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.

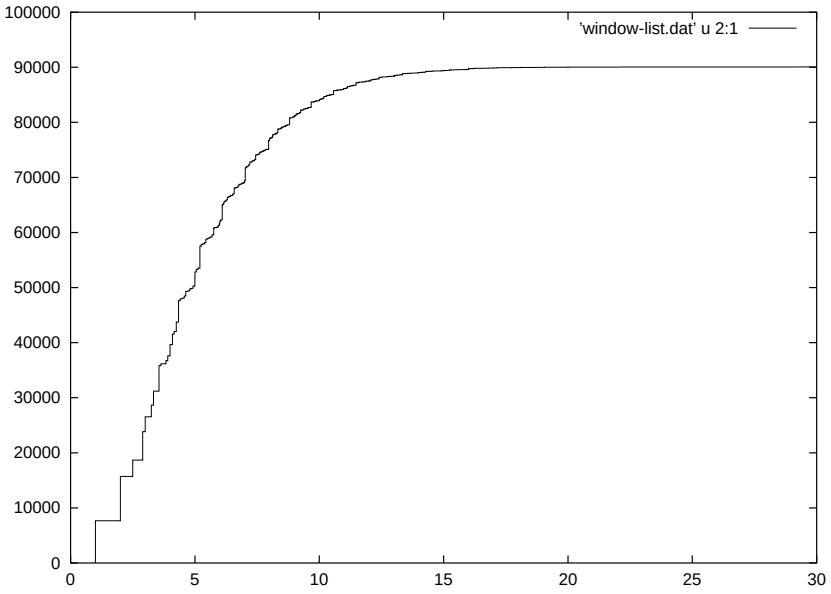

Figure 24: Window size distribution sampled at periodic time for TCP Reno, for 200 connections, buffer size 800, buffer service rate $\mu = 1$, average random processing time 10.