A MARKOVIAN ANALYSIS OF ADDITIVE-INCREASE MULTIPLICATIVE-DECREASE (AIMD) ALGORITHMS

VINCENT DUMAS, FABRICE GUILLEMIN, AND PHILIPPE ROBERT

Abstract. The Additive-Increase Multiplicative-Decrease (AIMD) schemes designed to control congestion in communication networks are investigated from a probabilistic point of view. Functional limit theorems for a general class of Markov processes that describe these algorithms are obtained. The asymptotic behavior of the corresponding invariant measures is described in terms of the limiting Markov processes. For some special important cases, including TCP congestion avoidance, an important AR (Autoregressive) property is proved. As a consequence, the explicit expression of the related invariant probabilities is derived. The transient behavior of these algorithms is also analyzed.

Contents

1. Introduction

This paper investigates the mathematical structure underlying the so-called additive -increase multiplicative-decrease (AIMD) window based flow control schemes used in data transmission. With the emergence of TCP (Transmission Control Protocol) as the ubiquitous data transfer protocol, the study of these algorithms is crucial to understand the complex behavior of modern communication networks. Keeping in mind that TCP is one of the AIMD algorithms, using the language of communication networks, these algorithms can be described as follows. When a packet is sent, it is acknowledged by the destination when received. To control the reception of packets by the destination, there is a packet loss detection mechanism used in the AIMD scheme. This mechanism is able to detect lightweight loss (the loss of a single packet from time to time) as well as heavy loss (for instance in the

Date: November, 21 2001.

1991 *Mathematics Subject Classification.* Primary 60K30, 90B18; Secondary 68M12.

Key words and phrases. TCP. Convergence of invariant probabilities. Auto-Regressive processes. Hitting times. Network protocols.

Partially supported by the contract 00-1B320 with France Telecom and the Future and Emerging Technologies programme of the EU under contract number IST-1999-14186 (ALCOM-FT).

case of severe congestion in the network). In the case of TCP, lightweight loss is detected through duplicated acknowledgements while heavy loss detection is performed through a time-out mechanism. In the following, we consider the case of lightweight loss only.

To simplify the description, we shall assume that the round trip time (RTT) between the source and the destination is constant. The source maintains a variable W referred to as the congestion window size which controls the transmission of packets over an RTT interval. When W packets are sent to the destination during an RTT interval, the source detects if one of them has been lost. If there is a loss, the variable W is changed to $\lfloor \delta W \rfloor$ for the next RTT interval, $\delta \in]0,1[$ is the multiplicative constant which decrements the window size; in this case the number of packets sent during the next cycle is drastically reduced. If all the packets are successfully transmitted, W is just incremented by some value $\beta > 0$ if it does not exceed some maximal value $w_{\max}$, the maximal congestion window size. This is of course a simplification of the real algorithms involved, but the basic mechanism of reducing the congestion (called congestion avoidance) is captured by this model. See Jacobson [15], Allman *et al.* [2] and Stevens [24] for more details.

Roughly speaking, the motivation of this algorithm can be described as follows: the loss of packets in the network is mainly due to buffer overflow in the nodes of the network. If the congestion window of many sources in the network is large, these sources will send many packets at approximately the same time and, very likely, they will create more and more overflow and thus more and more retransmissions, if their respective window sizes are not reduced very quickly. Multiplicative decrease is a way of rapidly cooling down congestion. The additive increase can be seen as a very progressive test of the congestion of the network.

A Markovian representation. The model considered in the paper describes the exchange of packets between the source and the destination. Each packet has some probability of being lost and the probability that packets are lost are independent. (In section 2 we consider a more general loss model.) The influence of the network is described through this loss process. With this assumption, the sizes of the congestion windows over the successive RTT intervals is a Markov chain (W_n).

The transitions of the Markov chain (W_n) are described by: if $W_0 = x \geq 1$,

$$W_1 = \begin{cases} \min\left(\lfloor x+\beta \rfloor, w_{\max}\right) & \text{with probability } \exp(-\alpha x) \\ \max\left(\lfloor \delta x \rfloor, 1\right) & \text{otherwise,} \end{cases} \tag{1}$$

where $\beta > 0$, $0 < \delta < 1$ and $\alpha > 0$; $\lfloor x \rfloor$ is the integer part of x. The quantity $\exp(-\alpha)$ is the probability that a packet is not lost in the network and, assuming independence of the losses, $\exp(-\alpha x)$ is the probability that all the x packets of the window are successfully transmitted. In this model the interaction between the network and the data transfer is represented only through losses of packets. The constant $w_{\max}$ is the maximal congestion window size, only determined by the destination.

When W_0 is very large, the drift $\mathbb{E}(W_1 - W_0)$ of the Markov chain is equivalent to $-(1-\delta)W_0$; it implies that the Markov chain cannot travel very far from the origin, in particular it is ergodic. (See the details below.) Thus, the long term behavior of the source is mainly driven by the invariant measure (π_n) of (W_n). For a fixed α, very little is known about this invariant probability, Dumas *et al.* [9] gives stochastic bounds for (π_n) which are accurate when the loss rate is not small.

In practice, hopefully, the loss rate will be small in the network. This suggests to look at the limiting behavior of this Markov chain when α tends to 0. This is the main topic of this paper. This topic is not new and has been studied in great details in the technical literature (see the section on related works below). The contribution of the present paper to the modeling of TCP is in that it takes benefit as far as possible of the Markovian structure of the congestion avoidance regime in order to derive rigorous convergence results and to derive closed formulæ for the mean throughput of a TCP connection with or without upper bound for the congestion window size. Moreover, still owing to the Markovian structure of the system, it is possible to carry out a transient analysis of the congestion avoidance regime. In particular, we compute the distribution of the duration of time necessary to reach the maximal congestion window size (Section 5).

Related work. Floyd [13], Floyd *et al.* [14] and Madhavi and Floyd [17] analyze the impact of AIMD algorithms through simulations and some approximated models.

Padhye *et al.* [21] considers a detailed model of the evolution of TCP. Using a finite Markov chain taking into account the key features of TCP (window size reduction, time out, etc.), they express some of the stationary characteristics of the protocol. In particular, they obtain a closed formula for the throughput of a TCP connection, which has become central in the field of TCP modeling. While the results obtained in [21] rely on an approximation of the different characteristics of a finite state Markov chain (in particular its steady state distribution), we exploit as far as possible in this paper the Markovian structure of the congestion avoidance regime to establish rigorous convergence results when the loss probability tends to 0. Moreover, we take into account a possible upper bound for the maximum window size. With regard to the throughput, this does not simply translate into a truncation of the throughput formula obtained when there is no upper bound. In fact, the presence of the upper bound for the maximum window size affects the whole distribution of the steady state probability distribution of the congestion window size in an intricate way.

Via a different approach, Ott *et al.* [20] analyzes the evolution of the size of the congestion window as the perturbation of a deterministic differential equation by a Poisson process when the loss rate is small. In this setting they are able to give a detailed description of the invariant measure when α tends to 0. Altman *et al.* [4] extended some of these results to study the case of a finite maximal congestion window. The model considered by Ott *et al.* [20] can be seen as an approximation of the model considered here (when the window size is infinite) but at a different time scale (or, more accurately, packet granularity). It is not clear for us how one could justify the approximation of the invariant probabilities in this setting (see Section 4).

Along the same lines of investigations, Adjih *et al.* [1] gives an asymptotic expression of the invariant measure of the size of the congestion window. They specifically study a large number of TCP connections multiplexed in a single buffer and then perform asymptotic analysis via a mean field technique. Altman *et al.* [3] give some moments of the window size at equilibrium with the assumption that the evolution of the size of the congestion window is an autoregressive process. If we indeed *prove* in our paper that an autoregressive process indeed plays a role, it does not seem to be related to the process introduced in [3] (see Propositions 12 and 13).

Baccelli and Hong [6] considers an algebraic setting describing precisely TCP when the input is deterministic or periodic.

Finally, while most of studies (including this one) assume fixed round trip times, the case of several TCP connections with different round trip times multiplexed on a link has been analyzed by Brown [8] via a fluid-based model. In particular, this analysis shows how the different connections occupy the link and the critical impact of round trip times and the size of the buffer on the performance of the system (e.g., the link utilization and the connection throughputs).

So far, we have discussed studies on the performance of TCP connections. But TCP as well as any other procotol used for carrying elastic traffic poses more general and very interesting problems with regard to fairness at the network level. Different studies have addressed this issue in the recent past (see for instance Kelly [16], Roberts and Massoulié [22],Vojnović *et al.* [26]). The problem of fairness also appears in the design of transmission protocols, which roughly behave as TCP (TCP friendly protocols). See for instance Vojnović and Le Boudec [25] for a mathematical formulation of the problem.

The results of the paper. Two Markov chains are considered in this paper (W_n^α) describing the evolution of congestion window size over the successive RTT intervals and (V_n^α) which is the embedded Markov chain of (W_n^α) observed when a packet is lost. In a quite general framework we prove the convergence in distribution of their invariant probability measures properly rescaled when the loss rate α goes to 0, i.e. the following convergences in distribution $\lim_{\alpha\to 0}\sqrt{\alpha}W_\infty^\alpha = \overline{W}_\infty$ and $\lim_{\alpha\to 0}\sqrt{\alpha}V_\infty^\alpha = \overline{V}_\infty$ (Theorem 9 and Theorem 10).

When the loss probability per packet is constant, an interesting AR (Auto-Regressive) property of the Markov chain $\left((V_n^\alpha)^2)\right)$ holds when α is close to 0. This result and the convergence results give the key to most of the explicit calculations of distributions: the distribution of $\overline{V}_\infty$ (Propositions 13 and 18), of $\overline{W}_\infty$ (Proposition 16) and the asymptotic throughput of the algorithm (Propositions 15 and 19).

This AR property does not seem to have been earlier identified in the literature. This is a real benefit of the approach of this paper to consider the convergence of the complete dynamics of the system rather than analyzing only the convergence of the invariant probability distribution when α tends to 0. In some of the studies of TCP (e.g. [3]), the AR property is *assumed*, not for $((V_n)^2)$ but for (V_n). It turns out that this assumption seems to lead a different constant for the throughput, 1.22 instead of the constant derived here 1.3098, already observed by Floyd *et al.* More important, when the maximum congestion window is infinite, the equilibrium has an exponential tail distribution instead of quadratic exponential in our case (i.e. $\sim \exp(-Cx^2)$), the number of large congestion windows is thus significantly smaller for the model considered here.

To our knowledge the transient behavior of AIMD algorithms has not been investigated through analytical models. For example, the time to reach the maximal congestion window is clearly an important characteristic. This measure illustrates the performance of the AIMD scheme in the sense that it indicates how long it takes, after a perturbation causing lightweight packet loss, to recover the maximal throughput obtained when the congestion window size is equal to the maximal value. For the moment it has not received much attention in the stochastic models. Some results in this domain are derived with our approach.

Organization. In Section 2 we prove the main convergence results of this paper: convergence in finite time of the Markov processes. More important, convergence of the corresponding invariant distributions is shown in Section 3. In Section 4 the explicit expression of the corresponding invariant distributions is derived. The formulas for the asymptotic throughput are obtained and discussed. Section 5 gives some results concerning the transient behavior of TCP, in particular the hitting time of the maximal congestion window size is investigated.

2. A generalized Markovian model

In this section, with the terminology of communication networks, we consider the sequence (W_n^α) describing the size of the congestion window over the successive RTT intervals. It is assumed that the round trip time is large and the state of the network evolves sufficiently rapidly so that the packet loss events in one RTT interval do not depend on the contiguous RTT intervals.

Since we are interested in the asymptotic regime when the loss probability tends to 0 and thus when packet loss rarely occurs so that it is reasonable to assume that there is a single window reduction in an RTT, we directly consider the sequence (W_n^α) described as a Markov chain with the following transitions: if $W_0^\alpha = n \geq 1$,

$$\begin{cases} W_1^\alpha = \min(n+1, w_{\max}^\alpha), \text{ with probability } \prod_{i=1}^n \exp(-h_i^\alpha), \\ W_1^\alpha = \max\left(\lfloor \delta n \rfloor, 1\right), \text{ otherwise }. \end{cases} \tag{2}$$

The quantity $\exp(-h_i^\alpha)$ is the probability that during an RTT interval the ith packet is not lost when $i-1$ packets have been successfully transmitted. We shall assume that for $i \in \mathbb{N}$,

$$h_i^\alpha = \alpha \overline{h}\left(i\sqrt{\alpha}\right), \tag{3}$$

where $\overline{h}$ is a non identically 0, continuous, non decreasing function on $\mathbb{R}_+$. In particular, for $x \geq 0$,

$$\lim_{\alpha \to 0} \frac{h^\alpha_{\lfloor x/\sqrt{\alpha}\rfloor}}{\alpha} = \overline{h}(x).$$

The parameter α controls the loss rate of the algorithm. The original model (1) corresponds to the case where $\overline{h}$ is constant equal to 1.

With this model, it is implicitly assumed that the probability of loss of a packet is non decreasing with respect to its index in the current window. (As the number of packets in the network grows, the more likely they will be lost). The constant $w_{\max}^\alpha \in \mathbb{N} \cup \{+\infty\}$ is the maximal window size, it is assumed that it satisfies the following scaling relation with α,

$$\lim_{\alpha \to 0} \sqrt{\alpha}\, w_{\max}^\alpha = \overline{w}_{\max}. \tag{4}$$

Without loss of generality for the asymptotic results we are considering, the additive constant β of the transitions (1) can be taken equal to 1.

The embedded Markov chain

It is natural to consider the state of the Markov chain (W_n^α) just after a loss. The associated process is denoted by (V_n^α), this is clearly a Markov chain whose transitions are given by, if $V_0^\alpha = n \geq 1$

$$V_1^\alpha = \lfloor \delta \min\left(n + G_n^\alpha, w_{\max}^\alpha\right)\rfloor \vee 1 \tag{5}$$

where $x \vee 1 = \max(x, 1)$ and the random variable G_n^α is defined by

$$\mathbb{P}(G_n^\alpha \geq m) = \prod_{k=n}^{n+m-1} \prod_{j=1}^{k} \exp(-h_j^\alpha),$$

for $m \geq 1$ and $n \geq 1$, recall that $\prod_{j=1}^k \exp(-h_j^\alpha)$ is the probability that a congestion window of k packets is successfully transmitted over an RTT interval. The quantity G_n^α is the number of congestion windows sent successfully when the initial window size is n. With the above equation it is easy to check that the random variable G_n^α is stochastically decreasing with n because $\overline{h}$ is non decreasing. Roughly speaking the number of successful consecutive windows is smaller when the starting point is higher.

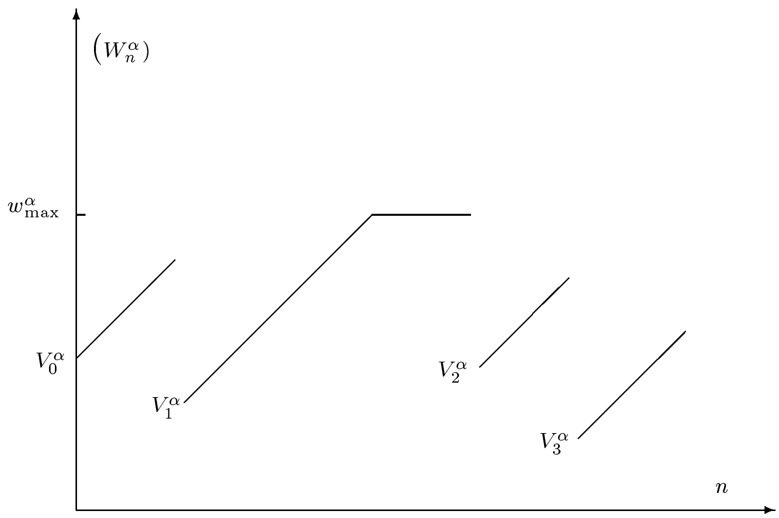

FIGURE 1. Evolution of the congestion window size

Convergence to a Markov process in finite time. We now show that the Markov chains described above are of the order $1/\sqrt{\alpha}$ when α tends to 0. The next proposition shows that $\sqrt{\alpha}$ is indeed the right scaling for the embedded Markov chain.

Proposition 1. *For $x > 0$, as α goes to 0, the random variable $\sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha} \rfloor}^\alpha$ converges in distribution to a non negative random variable $\overline{G}_x$ such that for $y \geq 0$,*

$$\mathbb{P}\left(\overline{G}_x \geq y\right) = \exp\left(-\int_x^{x+y} \overline{H}(u)\, du\right), \tag{6}$$

with $\overline{H}(u) = \int_0^u \overline{h}(v)\, dv$, and for any $K > 0$,

$$\lim_{\alpha \to 0} \sup_{\sqrt{\alpha} \leq x, y \leq K} \left| \mathbb{P}\left(\overline{G}_x \geq y\right) - \mathbb{P}\left(\sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha} \rfloor}^\alpha \geq y\right) \right| = 0. \tag{7}$$

Moreover, there exists $\lambda_0 > 0$ and $\alpha_0 > 0$ such that for $\lambda < \lambda_0$,

$$\sup_{0<\alpha<\alpha_0} \sup_{x \geq \sqrt{\alpha}} \mathbb{E}\left(e^{\lambda \sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha} \rfloor}^\alpha}\right) < +\infty. \tag{8}$$

When $\overline{h}\equiv 1$, the distribution of the random variable $\overline{G}_x$ is given by, for $y\geq 0$,

$$\mathbb{P}\left(\overline{G}_x \geq y\right) = \exp\left(-y^2/2 - xy\right).$$

Proof. For x, $y\geq\sqrt{\alpha}$ and $\alpha\in]0,1[$, we have

$$\mathbb{P}\left(\sqrt{\alpha}\,G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) = \mathbb{P}\left(G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > \lfloor y/\sqrt{\alpha}\,\rfloor\right) = \prod_{k=\lfloor x/\sqrt{\alpha}\rfloor}^{\lfloor x/\sqrt{\alpha}\rfloor+\lfloor y/\sqrt{\alpha}\rfloor}\prod_{j=1}^{k}\exp(-h_j^{\alpha}),$$

therefore

$$\begin{aligned}\log\mathbb{P}\left(\sqrt{\alpha}G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) &= -\sum_{k=\lfloor x/\sqrt{\alpha}\rfloor}^{\lfloor x/\sqrt{\alpha}\rfloor+\lfloor y/\sqrt{\alpha}\rfloor}\sum_{j=1}^{k}\alpha\overline{h}(j\sqrt{\alpha}\,)\\ &= -\sum_{k=\lfloor x/\sqrt{\alpha}\rfloor}^{\lfloor x/\sqrt{\alpha}\rfloor+\lfloor y/\sqrt{\alpha}\rfloor}\sqrt{\alpha}\int_0^{k\sqrt{\alpha}}\overline{h}(\sqrt{\alpha}\lceil v/\sqrt{\alpha}\,\rceil)\,dv\\ &= -\int_x^{\sqrt{\alpha}(\lfloor x/\sqrt{\alpha}\rfloor+\lfloor y/\sqrt{\alpha}\rfloor)}\int_0^{\sqrt{\alpha}\,\lfloor u/\sqrt{\alpha}\rfloor}\overline{h}(\sqrt{\alpha}\lceil v/\sqrt{\alpha}\,\rceil)\,dv\,du.\end{aligned}$$

This implies Relation (6). The uniform continuity of $\overline{h}$ over compact intervals implies that $\overline{h}(\sqrt{\alpha}\lceil\cdot/\sqrt{\alpha}\rceil)$ converges uniformly on $[0,K]$ to $\overline{h}$ as α tends to 0, using that for $u\in\mathbb{R}$, $|u-\sqrt{\alpha}\lfloor u/\sqrt{\alpha}\rfloor\leq\sqrt{\alpha}$, we get

$$\lim_{\alpha\to 0}\sup_{u\leq K}\left|\int_0^{\sqrt{\alpha}\,\lfloor u/\sqrt{\alpha}\rfloor}\overline{h}(\sqrt{\alpha}\lceil v/\sqrt{\alpha}\,\rceil)\,dv - \int_0^u\overline{h}(v)\,dv\right| = 0,$$

consequently,

$$\lim_{\alpha\to 0}\sup_{\sqrt{\alpha}\leq x,y\leq K}\left|\log\mathbb{P}\left(\sqrt{\alpha}G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) - \log\mathbb{P}\left(\overline{G}_x\geq y\right)\right| = 0.$$

From the uniform continuity of the exponential function on $]-\infty,0]$, the uniform convergence (7) is then easily obtained.

Since the function $\overline{h}$ is non decreasing, the above identity shows that for $x\geq 0$, $K>0$ and $y\geq 1$,

$$\log\mathbb{P}\left(\sqrt{\alpha}G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) \leq -\int_1^{y-1}\int_0^{K\wedge\sqrt{\alpha}\,\lfloor u/\sqrt{\alpha}\rfloor}\overline{h}(\sqrt{\alpha}\lceil v/\sqrt{\alpha}\,\rceil)\,dv\,du.$$

If λ_0 is the integral of $\overline{h}$ on $\mathbb{R}_+$, then for $\lambda<\lambda_0$ there exists $K>0$ and $\varepsilon_0>0$ such that

$$\lambda < \int_0^K\overline{h}(u)\,du - \varepsilon_0, \tag{9}$$

using again the uniform convergence of $\overline{h}(\sqrt{\alpha}\lceil\cdot/\sqrt{\alpha}\rceil)$ on the interval $[0,K]$, one gets that there exists some $y_0>0$ and $\alpha_0>0$ such that for $y\geq y_0$,

$$\sup_{0<\alpha<\alpha_0}\sup_{x\geq 0}\frac{1}{y}\log\mathbb{P}\left(\sqrt{\alpha}G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) \leq -\int_0^K\overline{h}(v)\,dv + \varepsilon_0$$

consequently, for $y \geq y_0$,

$$\mathbb{P}\left(\sqrt{\alpha} G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right) \leq \exp\left(y\left(\varepsilon_0 - \int_0^K \overline{h}(v)\,dv\right)\right),$$

hence, Inequality (9) gives the relation

$$\sup_{0<\alpha<\alpha_0} \sup_{x\geq 0} \int_2^{+\infty} e^{\lambda y} \mathbb{P}\left(\sqrt{\alpha} G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right)\,dy < +\infty.$$

Since, by Fubini's Theorem,

$$\int_2^{+\infty} e^{\lambda y} \mathbb{P}\left(\sqrt{\alpha} G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor} > y\right)\,dy = \mathbb{E}\left(\int_2^{\sqrt{\alpha} G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor}} e^{\lambda y}\,dy\right)$$

the above inequality yields

$$\sup_{0<\alpha<\alpha_0} \sup_{x\geq 0} \mathbb{E}\left(e^{\lambda\sqrt{\alpha} G^{\alpha}_{\lfloor x/\sqrt{\alpha}\rfloor}}\right) < +\infty.$$

The proposition is proved. □

Notice that the scaling $\sqrt{\alpha}$ does not depend on the particular choice of $\overline{h}$. The scaling of $\overline{h}$ in $h_i^{\alpha} = \alpha\overline{h}\,(i\sqrt{\alpha})$ determines the correct profile for the space dependence of the loss probability. One can introduce the following Markov chain, as we shall see, this is the asymptotic embedded Markov chain when α tends to 0.

Definition 2. *The sequence $(\overline{V}_n)$ denotes a Markov chain whose transitions are given by*

$$\overline{V}_1 = \delta \min\left(\overline{V}_0 + \overline{G}_{\overline{V}_0}, \overline{w}_{max}\right), \tag{10}$$

where $\left(\overline{G}_x;\, x > 0\right)$ is a family of random variables independent of V_0 such that, for $x > 0$, the distribution of $\overline{G}_x$ is given by Relation (6).

Proposition 3. *For any continuous function f on $\mathbb{R}_+$ with compact support, the following convergence holds,*

$$\lim_{\alpha\to 0} \sup_{x\geq\sqrt{\alpha}} \left|\mathbb{E}\left(f\left(\sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha}\rfloor}\right)\right) - \mathbb{E}\left(f(\overline{G}_x)\right)\right| = 0. \tag{11}$$

If for $x > 0$, $\mathbb{P}_x$ denotes the probability such that $V_0^{\alpha} = \lfloor x/\sqrt{\alpha}\rfloor$ for all $\alpha > 0$ and $\overline{V}_0 = x$, then for $n \in \mathbb{N}$ and $a_1, \ldots a_n \in \mathbb{R}_+$,

$$\lim_{\alpha\to 0} \sup_x \left|\mathbb{P}_x\left(\sqrt{\alpha} V_1^{\alpha} \leq a_1, \ldots, \sqrt{\alpha} V_n^{\alpha} \leq a_n\right) - \mathbb{P}_x\left(\overline{V}_1 \leq a_1, \ldots, \overline{V}_n \leq a_n\right)\right| = 0$$

Proof. If f is a C^1 function on $\mathbb{R}_+$ with support in $[0, K]$, $K > 0$, the identities

$$\sum_{k=2}^{\lceil K/\sqrt{\alpha}\rceil} (f(k\sqrt{\alpha}) - f((k-1)\sqrt{\alpha}))\mathbb{P}(Y \geq k\sqrt{\alpha}) = \mathbb{E}\left(f(Y) - f(\sqrt{\alpha})\right),$$

for $Y = \sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha}\rfloor}$ and $Y = \overline{G}_x$ are easily verified. The uniform continuity of f' on $\mathbb{R}_+$ and the estimate (7) give the desired convergence (11). If f is a continuous function with compact support in $[0, K]$, it can be approximated uniformly by C^1 functions. For $\varepsilon > 0$ there exists a C^1 function g on $[0, K]$ such that $\sup(|f(x) - g(x)|; x \leq K) \leq \varepsilon$, hence

$$\left|\mathbb{E}\left(f\left(\sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha}\rfloor}\right)\right) - \mathbb{E}\left(f(\overline{G}_x)\right)\right| \leq 2\varepsilon + \left|\mathbb{E}\left(g\left(\sqrt{\alpha}\, G_{\lfloor x/\sqrt{\alpha}\rfloor}\right)\right) - \mathbb{E}\left(g(\overline{G}_x)\right)\right|.$$

The first part of the proposition is proved.

We shall prove the last part of the proposition for $n=2$, i.e.

$$\lim_{\alpha\to 0}\sup_x \left|\mathbb{P}_x\left(\sqrt{\alpha}V_1^\alpha \le a_1, \sqrt{\alpha}V_2^\alpha \le a_2\right) - \mathbb{P}_x\left(\overline{V}_1 \le a_1, \overline{V}_2 \le a_2\right)\right| = 0.$$

Because of the transitions of the Markov chains, it is sufficient to show that for any $a_1,\, a_2 \le \overline{w}_{\max}$, the following convergence holds

$$\lim_{\alpha\to 0}\sup_x \Big|\mathbb{P}_x\left(\sqrt{\alpha}V_1^\alpha \le a_1, \sqrt{\alpha}V_1^\alpha + \sqrt{\alpha}G^\alpha_{V_1^\alpha} \le a_2\right) - \mathbb{P}_x\left(\overline{V}_1 \le a_1, \overline{V}_1 + \overline{G}_{\overline{V}_1} \le a_2\right)\Big| = 0.$$

Since

$$\mathbb{P}_x\left(\sqrt{\alpha}V_1^\alpha \le a_1, \sqrt{\alpha}V_1^\alpha + \sqrt{\alpha}G^\alpha_{V_1^\alpha} \le a_2\right) = \int_0^{a_1} \mathbb{P}\left(y + \sqrt{\alpha}G^\alpha_{\lfloor y/\sqrt{\alpha}\rfloor} \le a_2\right)\mathbb{P}_x(\sqrt{\alpha}V_1^\alpha \in dy),$$

since the quantity under the integrand is uniformly close to $f(y) = \mathbb{P}(y + \overline{G}_y \le a_2)$ with respect to $y \in [0, a_1]$, one has to verify that $|\mathbb{E}_x(f(\sqrt{\alpha}V_1^\alpha)) - \mathbb{E}_x(f(\overline{V}_1))|$ is uniformly small, but this is precisely a consequence of what has just been proved. The proof for an arbitrary n is done by induction. This completes the proof. ☐

Corollary 4. *If $\lim_{\alpha\to 0}\sqrt{\alpha}V_0^\alpha = \overline{v}$, then, as α tends to 0, the Markov chain $(\sqrt{\alpha}V_n^\alpha)$ converges in distribution to the Markov chain $(\overline{V}_n)$ (See Definition 2).*

Proof. Since one has to prove the convergence of the finite dimensional distributions, the corollary is a direct consequence of Proposition 3 and Relation (5). ☐

Proposition 5. *If $\lim_{\alpha\to 0}\sqrt{\alpha}W_0^\alpha = \overline{w}$, then the Markov process*

$$(W^\alpha(t)) = \left(\sqrt{\alpha}W^\alpha_{\lfloor t/\sqrt{\alpha}\rfloor}\right)$$

converges in distribution to the Markov process $\left(\overline{W}(t)\right)$ on $[0, \overline{w}_{max}]$ such that $\overline{W}(0) = \overline{w}$ and with the infinitesimal generator given by

$$\Omega(f)(x) = f'(x)1_{\{x<\overline{w}_{max}\}} + \int_0^x \overline{h}(u)\,du\,\left(f(\delta x) - f(x)\right), \tag{12}$$

for any C^1 function f on $\mathbb{R}_+$.

The Markov process $(W^\alpha(t))$ is right continuous with left limits everywhere. The convergence mentioned in this proposition is the convergence of probability distributions on the space of right continuous functions on $\mathbb{R}_+$ with left limits endowed with Skorohod topology. (See Billingsley [7] or Ethier and Kurtz [12] for the definitions and results concerning this topology).

The distributions of the variables $(\overline{G}_x,\, x \ge 0)$ have the following remarkable property that for $x,\, y \ge 0$,

$$\mathbb{P}(\overline{G}_x \ge y) = \frac{\mathbb{P}(\overline{G}_0 \ge x+y)}{\mathbb{P}(\overline{G}_0 \ge x)}.$$

(See Definition (6).) This property can be also seen as a consequence of the Markov property of the asymptotic Markov process $\left(\overline{W}(t)\right)$ when the maximum window size is infinite.

Proof. We shall assume that $\overline{w}_{\max}$ is infinite, the proof is analogous (even simpler) when this quantity is finite. Basically we shall prove that the infinitesimal generators converge to the appropriate infinitesimal generator. If this gives an indication of the kind of results one might expect, to prove the convergence rigorously some uniform convergence has to be established. We shall use the criterion given by Condition (h) of Corollary 8.9 page 233 of Ethier and Kurtz [12] for the Markov processes obtained from Markov chains.

We denote and by $\mathcal{C}$ the set of C^2-functions f on $\mathbb{R}_+$ such that the convergence

$$\lim_{x\to+\infty}(1+\overline{H}(x)^2)\sup_{u\geq x-1}|g(u)|=0,$$

($\overline{H}$ is defined in Proposition 1), is true for $g=f'$, f'' and $g(x)=f(\delta x)$; $\mathcal{C}$ is an algebra that strongly separates the points, i.e. if $x\in\mathbb{R}_+$ and $\delta>0$ then there exists $f\in\mathcal{C}$ such that

$$\inf(|f(y)-f(x)|:|y-x|\geq\delta)>0.$$

Condition (h) of Ethier and Kurtz is applied by taking (in the notations of this corollary) $G_n=\mathbb{R}_+$, so that Equation (8.47) is automatically satisfied.

If P^α is the transition matrix of the Markov chain $(\sqrt{\alpha}W_n^\alpha)$ and

$$A_\alpha=(P^\alpha-I)/\sqrt{\alpha},$$

where I is the identity; if we prove that for any $f\in\mathcal{C}$,

$$\lim_{\alpha\to0}\|A_\alpha(f)-\Omega(f)\|_\infty=\lim_{\alpha\to0}\sup_{x\geq0}|A_\alpha(f)(x)-\Omega(f)(x)|=0, \tag{13}$$

then Equation (8.48) of Condition (h) of Ethier and Kurtz is established, hence the Corollary 8.9 can be applied and the convergence is then proved.

For $x\geq0$ and $f\in\mathcal{C}$,

$$A_\alpha(f)(x)=\frac{1}{\sqrt{\alpha}}(f(x+\sqrt{\alpha})-f(x))\mathbb{P}(G_x^\alpha\geq1)\\ +\left(f\left(\sqrt{\alpha}\lfloor\delta x/\sqrt{\alpha}\rfloor\right)-f(x)\right)\frac{1-\mathbb{P}(G_x^\alpha\geq1)}{\sqrt{\alpha}}.$$

For $x\geq0$ the difference $|A_\alpha(f)(x)-\Omega(f)(x)|$ can be bounded by the quantity $\Delta_1(x)+\Delta_2(x)$, with

$$\Delta_1(x)=\left|\frac{1}{\sqrt{\alpha}}(f(x+\sqrt{\alpha})-f(x))-f'(x)\right|\\ +|f'(x)|\left|1-\exp\left(-\sqrt{\alpha}\int_0^x\overline{h}(\lfloor u/\sqrt{\alpha}\rfloor\sqrt{\alpha})\,du\right)\right|$$

and

$$\Delta_2(x)=\left|f\left(\sqrt{\alpha}\lfloor\delta x/\sqrt{\alpha}\rfloor\right)-f(\delta x)\right|\int_0^x\overline{h}(u)\,du\\ +(|f(\delta x)|+|f(x)|)\left|\frac{1-\exp\left(-\sqrt{\alpha}\int_0^x\overline{h}(\lfloor u/\sqrt{\alpha}\rfloor\sqrt{\alpha})\right)\,du}{\sqrt{\alpha}}-\int_0^x\overline{h}(u)\,du\right|.$$

Taylor's formulas for f and $x \to \exp(-x)$, the fact that f is in $\mathcal{C}$ and straightforward calculations give the desired uniform convergence (13). The proposition is proved. $\square$

The limiting Markov process grows deterministically at rate 1 and jumps from x to δx with intensity $\int_0^x \overline{h}(u)\,du$. It is easy to check that, starting from $x>0$, the duration of time to jump downwards has the same distribution as $\overline{G}_x$.

3. Convergence of the invariant measures

We are now interested by the equilibrium behavior of the AIMD algorithm. Up to now, a closed form expression for the invariant probabilities of the Markov chains (W_n^α) and (V_n^α) is not known, see Dumas *et al.* [9] for stochastic bounds in some special cases. The main results of this part concern the convergence in distribution of these invariant probability measures when α tends to 0. As we shall see in Section 4, the limiting probabilities have an explicit expression in some cases of practical interest. For the moment we study the behavior of the embedded Markov chain (V_n^α).

Definition 6. *If $K \geq 0$, T_K^α [resp. T_K] is the hitting time of the interval $[0,K]$ by the Markov chain $(\sqrt{\alpha}\,V_n^\alpha)$ [resp. $(\overline{V}_n)$],*

$$T_K^\alpha = \inf\left\{n \geq 1 : \sqrt{\alpha}\,V_n^\alpha \leq K\right\} \text{ and } \overline{T}_K = \inf\left\{n \geq 1 : \overline{V}_n \leq K\right\}.$$

Proposition 7. *For $\alpha>0$, the Markov chain (V_n^α) is ergodic. When $\overline{w}_{max}^\alpha = +\infty$, there exist K, ξ and $\lambda > 0$ such that for $0<\alpha<1$,*

$$\mathbb{E}\left(e^{\xi T_K^\alpha - \lambda\sqrt{\alpha}V_0^\alpha} \,\middle|\, \sqrt{\alpha}V_0^\alpha > K\right) \leq 1. \tag{14}$$

Proof. The proof uses a Foster-like criterion to give an estimate of the exponential moment of T_K^α (see Meyn and Tweedie [18]). For $n \in \mathbb{N}$, $\mathcal{F}_n$ denotes the σ-field generated by the random variables $V_0^\alpha, \dots V_{n-1}^\alpha, V_n^\alpha$.

The Markov chain (V_n^α) is clearly irreducible and aperiodic on $\mathbb{N} - \{0\}$. For $\lambda > 0$ and $n \in \mathbb{N}$, define $Z_n = \exp(\lambda\sqrt{\alpha}V_n^\alpha)$. For $K>0$, the inequality

$$\mathbb{E}(Z_{n+1} - Z_n | \mathcal{F}_n) \leq Z_n \mathbb{E}\left(e^{\lambda(\delta-1)K + \lambda\delta\sqrt{\alpha}G^\alpha_{V_n^\alpha}} - 1\right),$$

holds on the $\mathcal{F}_n$-measurable event $E_n \stackrel{\text{def}}{=} \{\sqrt{\alpha}V_n^\alpha > K\}$, therefore if $\alpha < 1$,

$$\mathbb{E}(Z_{n+1} - Z_n | \mathcal{F}_n) \leq Z_n \left(e^{\lambda(\delta-1)K} \sup_{\substack{0<\alpha<1\\ y\geq 0}} \mathbb{E}\left(e^{\lambda G^\alpha_{\lfloor y/\sqrt{\alpha}\rfloor}}\right) - 1\right)$$

holds on E_n. According to Proposition 1, we can fix $\lambda < \lambda_0$ so that there exists a constant C satisfying

$$\mathbb{E}(Z_{n+1} - Z_n | \mathcal{F}_n) \leq Z_n \left(Ce^{\lambda(\delta-1)K} - 1\right),$$

on E_n for all $\alpha \in]0,1[$. Consequently, there exist some $K_0 > 0$ and $\eta < 1$ such that

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) \leq \eta Z_n, \quad \text{on the event } \{\sqrt{\alpha}V_n^\alpha > K\}, \tag{15}$$

for $K \geq K_0$. The inequality (15) implies that if $\sqrt{\alpha}V_0^\alpha > K$, the sequence

$$(U_n) = \left(\eta^{-n\wedge T_K^\alpha} Z_{n\wedge T_K^\alpha}\right)$$

is a super-martingale with respect to $(\mathcal{F}_n)$ (since the relation $T_K^\alpha > n$ implies that the inequality $\sqrt{\alpha}V_i^\alpha > K$ holds for $i \leq n$). Hence for $n \geq 0$,

$$\mathbb{E}\left(\eta^{-n\wedge T_K^\alpha} \,\middle|\, \mathcal{F}_0\right) \leq \mathbb{E}\left(\eta^{-n\wedge T_K^\alpha} Z_{n\wedge T_K^\alpha} \,\middle|\, \mathcal{F}_0\right) = \mathbb{E}(U_n|\mathcal{F}_0) \leq U_0 = e^{\lambda\sqrt{\alpha}V_0^\alpha}.$$

By letting n go to infinity we get the inequality

$$\mathbb{E}\left(\eta^{-T_K^\alpha} \,\middle|\, \mathcal{F}_0\right) \leq e^{\lambda\sqrt{\alpha}V_0^\alpha}$$

on the event $\{\sqrt{\alpha}V_0^\alpha > K\}$. The proposition is proved. □

Proposition 8. *The continuous state space Markov chain $(\overline{V}_n)$ is Harris ergodic. If $\overline{V}_\infty$ is some random variable distributed as the invariant probability of $(\overline{V}_n)$, it satisfies the following identity*

$$\overline{V}_\infty \stackrel{dist.}{=} \delta \min\left(\overline{V}_\infty + \overline{G}_{\overline{V}_\infty}, \overline{w}_{max}\right) \tag{16}$$

where $(\overline{G}_x;\, x \geq 0)$ are random variables independent of $\overline{V}_\infty$ whose distributions are given by the relation (6).

For the general definitions and results concerning Markov processes with a continuous state space, see Nummelin [19].

Proof. Since the transition of the Markov chain has a continuous density, it is a Harris chain (See Durrett [10] page 326). It is easily seen that for any $x \geq 0$, then $\mathbb{P}(\overline{G}_0 \geq y) \leq \mathbb{P}(\overline{G}_x \geq y)$ for $y \geq 0$; $\overline{G}_x$ is stochastically bounded by $\overline{G}_0$. Thus we can construct a sequence (Z_n) and an i.i.d sequence $(\overline{G}_{0,n})$ with the same distribution as $\overline{G}_0$ such that $Z_0 = \overline{V}_0$,

$$Z_{n+1} = \delta(Z_n + \overline{G}_{0,n}) \text{ and } \overline{V}_n \leq Z_n,$$

for all $n \in \mathbb{N}$. The sequence (Z_n) is an AR process (Autoregressive) which is Harris ergodic (see Durrett [*ibid*]). Thus we get that the sequence of probability distributions

$$\left(\frac{1}{n}\sum_{k=1}^{n} \mathbb{P}\left(\overline{V}_k \in \cdot\right)\right)$$

is tight; clearly any limit of this sequence is an invariant probability distribution for the Markov chain $(\overline{V}_n)$. To conclude, an Harris Markov chain with an invariant probability distribution is necessarily ergodic. (See Durrett [10] Exercise 6.11 page 330.)

According to Definition (10) of the transitions of the Markov chain $(\overline{V}_n)$, its invariant distribution satisfies the relation (16). □

Theorem 9. *When α tends to 0 the invariant distribution of the Markov chain $(\sqrt{\alpha}V_n^\alpha)$ converges to the invariant distribution of the Markov chain $(\overline{V}_n)$. Consequently, the following diagram commutes,*

$$\begin{array}{ccc}
(\sqrt{\alpha}V_n^\alpha) & \xrightarrow{n\to+\infty} & \sqrt{\alpha}V_\infty^\alpha \\
{\scriptstyle \alpha\to 0}\downarrow & & \downarrow \\
(\overline{V}_n) & \longrightarrow & \overline{V}_\infty.
\end{array}$$

Proof. We denote by π^α the invariant probability of $(\sqrt{\alpha}V_n^\alpha)$ and (Z_n^α) is the sequence of the successive elements of $[0,K]$ visited by this Markov chain. In the rest of the proof, for variables with index α, the notation $\mathbb{E}_\mu(\cdot)$ and $\mathbb{P}_\mu(\cdot)$ refer to the Markov chain $(\sqrt{\alpha}V_n^\alpha)$ when the distribution of $\sqrt{\alpha}V_0^\alpha$ is μ.

Notice that if $\sqrt{\alpha}V_0^\alpha \le K$, then $Z_0^\alpha = \sqrt{\alpha}V_0^\alpha$ and $Z_1^\alpha = \sqrt{\alpha}V_{T_K^\alpha}^\alpha$. With the same argument as in the proof of Proposition 8 it easily seen that (Z_n^α) is a Harris ergodic Markov chain; π_K^α denotes its invariant probability, in particular $\pi_K^\alpha[0,K]=1$. The probability π^α can be represented as

$$\mathbb{E}_{\pi^\alpha}(f) \stackrel{\text{def}}{=} \int_{\mathbb{R}+} f\,d\pi^\alpha = \frac{1}{\mathbb{E}_{\pi_K^\alpha}(T_K^\alpha)}\mathbb{E}_{\pi_K^\alpha}\left(\sum_{k=0}^{T_K^\alpha-1} f\left(\sqrt{\alpha}V_k^\alpha\right)\right), \tag{17}$$

for any bounded measurable function f on $\mathbb{R}_+$ (see Asmussen [5] for example). For the asymptotic Markov chain, with the corresponding notations ($\overline{\pi}$ is the invariant probability of $(\overline{V}_n)$), the following identity holds,

$$\mathbb{E}_{\overline{\pi}}(f) = \frac{1}{\mathbb{E}_{\overline{\pi}_K}(\overline{T}_K)}\mathbb{E}_{\overline{\pi}_K}\left(\sum_{k=0}^{\overline{T}_K-1} f(\overline{V}_k)\right), \tag{18}$$

where $\overline{\pi}_K$ is the invariant probability of $(\overline{Z}_n)$ which is the embedded Markov chain of the visits of $(\overline{V}_n)$ in the set K.

The proof of the Theorem consists in showing that, for some $K>0$, the left hand side of (17) converges to the left hand side of (18) when α tends to 0. To prove this convergence, the inequality (14) is used to truncate the sum under the expectation and the remaining terms are shown to converge with the help of relation (11).

The set of probability measures $\{\pi_K^\alpha;\,\alpha>0\}$ is obviously tight. Consequently, there exists a probability π_K on $[0,K]$ and a sequence (α_n) converging to 0 such that $(\pi_K^{\alpha_n})$ converges to π_K.

The probability π_K is invariant for $(\overline{Z}_n)$.
For $n\in\mathbb{N}$ and a, $x\le K$,

$$\mathbb{P}(T_K^\alpha = n, Z_1^\alpha \le a) = \mathbb{P}\left(V_1^\alpha > K,\dots,V_{n-1}^\alpha > K, V_n^\alpha \le a\right)$$

According to Proposition 3, when α goes to 0 this last quantity is converging to

$$\mathbb{P}\left(\overline{V}_1 > K,\dots,\overline{V}_{n-1} > K, \overline{V}_n \le a\right) = \mathbb{P}(\overline{T}_K = n, \overline{Z}_1 \le a)$$

uniformly on $\sqrt{\alpha}V_0^\alpha \in [0,K]$. In particular the variables $(T_K^{\alpha_n})$ [resp. $(Z_1^{\alpha_n})$] converge in distribution to the variable $\overline{T}_K$ [resp. $\overline{Z}_1$]. By invariance, for $\alpha>0$,

$$\pi_K^{\alpha_n}([a,+\infty[) = \mathbb{P}_{\pi_K^{\alpha_n}}(Z_0^{\alpha_n}\le a) = \mathbb{P}_{\pi_K^{\alpha_n}}(Z_1^{\alpha_n}\le a),$$

the uniform convergence gives the identity

$$\pi_K([a,+\infty[) = \lim_{n\to+\infty}\mathbb{P}_{\pi_K^{\alpha_n}}(Z_1^{\alpha_n}\le a) = \mathbb{P}_{\pi_K}(\overline{Z}_1\le a)$$

The probability π_K is therefore an invariant probability measure for the Markov chain $(\overline{Z}_n)$; by uniqueness this implies that $\pi_K = \overline{\pi}_K$. Thus, we have shown that the probabilities $(\pi_K^\alpha;\alpha>0)$ converge to $\overline{\pi}_K$ as α tends to 0.

Proposition 7 shows that there exist constants K, ξ, $\lambda>0$ such that

$$\mathbb{E}\left(e^{\xi T_K^\alpha - \lambda\sqrt{\alpha}V_0^\alpha}\,\middle|\,\sqrt{\alpha}V_0^\alpha \ge K\right) \le 1.$$

The Markov property gives the relation

$$\begin{aligned} e^{-\xi}\mathbb{E}_{\pi_K^\alpha}\left(e^{\xi T_K^\alpha}\right) &\le 1+\mathbb{E}_{\pi_K^\alpha}\left(e^{\lambda\sqrt{\alpha}V_1^\alpha}\ \mathbb{E}_{\sqrt{\alpha}V_1^\alpha}\left(e^{\xi T_K^\alpha-\lambda\sqrt{\alpha}V_1^\alpha}\,\middle|\,\sqrt{\alpha}V_1^\alpha>K\right)\right) \\ &\le 1+\mathbb{E}_{\pi_K^\alpha}\left(e^{\lambda\sqrt{\alpha}V_1^\alpha}\right)\le 1+\mathbb{E}_{\pi_K^\alpha}\left(e^{\lambda\delta\sqrt{\alpha}V_0^\alpha+\delta\lambda\sqrt{\alpha}G^\alpha_{V_0^\alpha}}\right)\end{aligned}$$

since $\sqrt{\alpha}V_0^\alpha\le K$ for the probability measure $\mathbb{P}_{\pi_K^\alpha}$,

$$e^{-\xi}\mathbb{E}_{\pi_K^\alpha}\left(e^{\xi T_K^\alpha}\right)\le 1+e^{\lambda\delta K}\mathbb{E}_{\pi_K^\alpha}\left(e^{\lambda\sqrt{\alpha}G^\alpha_{V_0^\alpha}}\right),$$

and this quantity is bounded for $0<\alpha<1$, according to the inequality (8) (by choosing λ sufficiently small). With the initial distributions (π_K^α), the variables $(T_K^\alpha;\alpha<1)$ are therefore uniformly integrable, in particular

$$\lim_{\alpha\to 0}\mathbb{E}_{\pi_K^\alpha}(T_K^\alpha)=\mathbb{E}_{\pi_K}(\overline{T}_K),$$

and for $\varepsilon>0$ there exists $C>0$ such that

$$\mathbb{E}_{\pi_K}\left(\overline{T}_K 1_{\{\overline{T}_K\ge C\}}\right)\le\varepsilon \quad\text{and}\quad \mathbb{E}_{\pi_K^\alpha}\left(T_K^\alpha 1_{\{T_K^\alpha\ge C\}}\right)\le\varepsilon,$$

for $0<\alpha<1$. From these estimates and the uniform convergence of Proposition 3, for a bounded measurable function f, we deduce that

$$\lim_{\alpha\to 0}\mathbb{E}_{\pi_K^\alpha}\left(\sum_{k=0}^{T_K^\alpha-1}f(V_k^\alpha)\right)=\mathbb{E}_{\overline{\pi}_K}\left(\sum_{k=0}^{\overline{T}_K-1}f(\overline{V}_k)\right),$$

as α goes to 0, hence according to the identities (17) and (18), the probabilities $(\pi^\alpha;\alpha>0)$ converge to $\overline{\pi}$ as α tends to 0. The theorem is proved. □

Theorem 10. *When α tends to 0 the invariant distribution of the Markov chain $(\sqrt{\alpha}W_n^\alpha)$ converges to the invariant distribution of the Markov process $(\overline{W}(t))$,*

$$\begin{array}{ccc} \left(\sqrt{\alpha}W^\alpha_{\lfloor t/\sqrt{\alpha}\rfloor}\right) & \xrightarrow{t\to+\infty} & \sqrt{\alpha}W^\alpha_\infty \\ {\scriptstyle\alpha\to 0}\downarrow & & \downarrow \\ (\overline{W}(t)) & \longrightarrow & \overline{W}(\infty). \end{array}$$

Proof. The proof is similar to the proof of Theorem 9. Basically the sums in (17) and (18) have to be replaced by integrals. □

4. The representation of the limiting invariant measures

In this section we shall consider the case when $\overline{h}\equiv 1$; the loss probability is a constant for all packets. In this case, for $x\ge 0$, the distribution of $\overline{G}_x$ is given by

$$\mathbb{P}(\overline{G}_x\ge y)=e^{-xy-y^2/2}, \tag{19}$$

for $y\ge 0$. The infinitesimal generator Ω given by (12) of the asymptotic Markov process is given by

$$\Omega(f)(x)=f'(x)1_{\{x<\overline{w}_{\max}\}}+x(f(\delta x)-f(x)). \tag{20}$$

The following simple proposition is crucial to get the stationary behavior of AIMD algorithms analyzed in this section.

Proposition 11. *For $x > 0$, if the distribution of a random variable $\overline{G}_x$ is given by (19), the following identity holds*

$$(x + \overline{G}_x)^2 \stackrel{dist.}{=} 2E_1 + x^2, \tag{21}$$

where E_1 an exponentially distributed random variable with parameter 1.

Proof. For $y \geq 0$, the relation (19) gives

$$\mathbb{P}((x + \overline{G}_x)^2 \geq y + x^2) = e^{-y/2},$$

and the result follows. □

We first study the case of the infinite maximum window size. The characteristics of the system at equilibrium (invariant probability, throughput) have rather simple closed form expressions. The corresponding expressions for the finite maximum window size case are still explicit but more intricate.

4.1. **Infinite maximum window size.**

The embedded Markov chain. The following proposition gives a probabilistic representation of $\overline{V}_\infty$ when the maximum window size is infinite. Its analytic counterpart is Proposition 13 below. It shows that the square of the limiting embedded Markov chain is an AR process. To our knowledge, this crucial property does not seem to have been remarked previously. The AR property is the key characteristic of the AIMD control scheme. As we shall see later the AR property has important consequences on the qualitative behavior of TCP. In particular it implies that the tail distribution of the size of the congestion window decays as $\exp(-\beta x^2)$. This indicates that the probability of having large windows is rapidly decreasing and that the steady state probability distribution is concentrated on relatively small values of the congestion window size.

Proposition 12. *When the maximum window size is infinite, $\overline{w}_{max} = +\infty$, the square of the Markov chain $(\overline{V}_n)$ is an AR process with the following representation: for $n \in \mathbb{N}$,*

$$\overline{V}_{n+1}^2 = \delta\left(\overline{V}_n^2 + 2E_n\right)$$

where (E_n) is an i.i.d. sequence of exponentially distributed random variables with parameter 1.

Its invariant probability can represented by the random variable $\overline{V}_\infty$ satisfying relation (16)

$$\overline{V}_\infty \stackrel{dist.}{=} \sqrt{2\sum_{n=1}^{+\infty} \delta^{2n} E_n}. \tag{22}$$

Proof. From relation (10) we get

$$\overline{V}_{n+1} \stackrel{\text{dist.}}{=} \delta\left(\overline{V}_n + \overline{G}_{\overline{V}_n}\right),$$

where the variables $(\overline{G}_g; x > 0)$ are independent of $\overline{V}_n$. Therefore Proposition 11 gives the identity

$$\overline{V}_{n+1}^2 \stackrel{\text{dist.}}{=} \delta^2\left(\overline{V}_n^2 + 2E_n\right),$$

where E_n is an exponential variable with parameter 1. The AR property is proved. At equilibrium, using relation(16), the corresponding equation is

$$\overline{V}_\infty^2 \stackrel{\text{dist.}}{=} \delta^2\left(\overline{V}_\infty^2 + 2E_1\right).$$

If this relation is iterated, one gets identity (22) (the residual term $\delta^{2n}\overline{V}_\infty^2$ converges in distribution to 0). □

From the above proposition one can derive the density of the limiting random variable $\overline{V}_\infty$.

Proposition 13. *The density function of $\overline{V}_\infty$ when the maximum window size is infinite is given by, for $x \geq 0$,*

$$h_\delta(x) = \frac{1}{\prod_{n=1}^{+\infty}(1-\delta^{2n})}\sum_{n=1}^{+\infty}\frac{1}{\prod_{k=1}^{n-1}(1-\delta^{-2k})}\delta^{-2n}xe^{-\delta^{-2n}x^2/2}. \tag{23}$$

Proof. Let $\widetilde{V}_\infty$ denote the Laplace transform of the random variable $\overline{V}_\infty^2/2$, which is defined for $\xi \in \mathbb{C}$ with $\Re(\xi) \geq 0$ by

$$\widetilde{V}_\infty(\xi) = \mathbb{E}\left(e^{-\xi\overline{V}_\infty^2/2}\right).$$

A direct consequence of the independence of the exponential random variables (E_n) in equation (22) is that

$$\widetilde{V}_\infty(\xi) = \prod_{k=1}^{\infty}\frac{1}{(1+\delta^{2k}\xi)}.$$

The Laplace transform $\widetilde{V}_\infty$ has simple poles located at the points $\{-1/\delta^{2n},\, n \geq 1\}$. For $n \geq 1$, the residue of $\widetilde{V}_\infty$ at $-1/\delta^{2n}$ is given by

$$\frac{1}{\delta^{2n}}\prod_{k=1}^{n-1}\frac{1}{(1-\delta^{2k}/\delta^{2n})}\prod_{k=n+1}^{\infty}\frac{1}{(1-\delta^{2k}/\delta^{2n})} = \frac{\delta^{-2n}}{\prod_{k=1}^{n-1}(1-\delta^{-2k})}\frac{1}{\prod_{k=1}^{\infty}(1-\delta^{2k})}.$$

It follows that

$$\widetilde{V}_\infty(\xi) = \frac{1}{\prod_{k=1}^{\infty}(1-\delta^{2k})}\sum_{n=1}^{+\infty}\frac{1}{\prod_{k=1}^{n-1}(1-\delta^{-2k})}\frac{\delta^{-2n}}{\xi+\delta^{-2n}},$$

hence the density of $\overline{V}_\infty^2/2$ is given by, for $x \geq 0$,

$$\frac{1}{\prod_{k=1}^{\infty}(1-\delta^{2k})}\sum_{n=1}^{+\infty}\frac{1}{\prod_{k=1}^{n-1}(1-\delta^{-2k})}\delta^{-2n}e^{-\delta^{-2n}x}.$$

This completes the proof of the proposition. □

As shown by the picture below the distribution of $\overline{V}_\infty$ is sharply concentrated near the origin. Its tail distribution is equivalent to $Cx\exp(-\delta^2x^2/2)$ as x gets large.

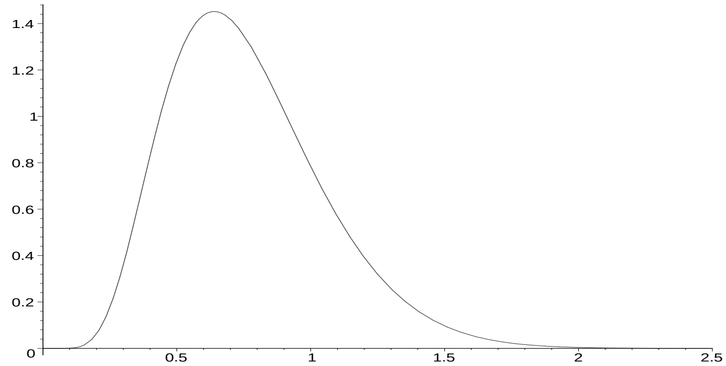

FIGURE 2. The density of $\overline{V}_\infty$ for $\delta = 1/2$

The throughput.

Definition 14. *For $\alpha > 0$, the throughput of the algorithm is defined as the limit*

$$\rho^\alpha(\delta) = \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} W_k^\alpha.$$

This definition assumes that the round trip time (RTT) is taken equal to 1. Thus, up to this factor, this is the definition of the literature. The ergodic theorem for the Markov chain (W_n^α) gives that $\rho^\alpha(\delta) = \mathbb{E}(W_\infty^\alpha)$. Using the embedded Markov chain (V_n^α) it is easily seen that the throughput can also be written as

$$\rho^\alpha(\delta) = \lim_{n\to+\infty} \frac{1}{\sum_1^n G^\alpha_{V_i^\alpha}} \sum_{i=1}^{n} \sum_{k=0}^{G^\alpha_{V_i^\alpha}-1} (V_i + k).$$

From the ergodic theorem applied to the Markov chain (V_n^α), we get finally

$$\rho^\alpha(\delta) = \frac{\mathbb{E}\left(\sum_{k=0}^{G^\alpha_{V_\infty^\alpha}-1} (V_\infty^\alpha + k)\right)}{\mathbb{E}\left(G^\alpha_{V_\infty^\alpha}\right)} = \frac{\mathbb{E}\left(2G^\alpha_{V_\infty^\alpha} V_\infty^\alpha + \left(G^\alpha_{V_\infty^\alpha}\right)^2\right)}{2\mathbb{E}\left(G^\alpha_{V_\infty^\alpha}\right)} - 1/2. \tag{24}$$

Proposition 15. *The asymptotic throughput of an AIMD algorithm with multiplicative decrease factor δ is given by*

$$\overline{\rho}(\delta) \stackrel{def}{=} \lim_{\alpha\to 0} \sqrt{\alpha}\rho^\alpha(\delta) = \frac{\delta}{(1-\delta)\mathbb{E}\left(\overline{V}_\infty\right)} = \sqrt{\frac{2}{\pi}}\, \frac{\prod_{n=1}^{+\infty}\left(1-\delta^{2n}\right)}{\prod_{n=0}^{+\infty}\left(1-\delta^{2n+1}\right)}. \tag{25}$$

Proof. By definition of the throughput, we have

$$\rho^\alpha(\delta) = -1/2 + \mathbb{E}\left(\left(V_\infty^\alpha + G^\alpha_{V_\infty^\alpha}\right)^2 - \left(V_\infty^\alpha\right)^2\right) \Big/ 2\mathbb{E}\left(G^\alpha_{V_\infty^\alpha}\right).$$

Identity (16) shows that

$$\begin{aligned}\rho^\alpha(\delta) &= -1/2 + \mathbb{E}\left((V_\infty^\alpha)^2(1/\delta^2-1)\right)/\, 2\mathbb{E}\left(G^\alpha_{V^\alpha_\infty}\right)\\ &= \frac{(1+\delta)\mathbb{E}((V_\infty^\alpha)^2)}{2\delta(1-\delta)\mathbb{E}(V_\infty^\alpha)} - 1/2.\end{aligned}$$

According to Theorem 9, the following convergence holds

$$\lim_{\alpha\to 0}\sqrt{\alpha}\rho^\alpha(\delta) = \frac{(1+\delta)\mathbb{E}\left(\overline{V}_\infty^2\right)}{2\delta(1-\delta)\mathbb{E}(\overline{V}_\infty)} = \frac{\delta}{(1-\delta)\mathbb{E}(\overline{V}_\infty)},$$

where the last relation is a consequence of representation (22). The quantity $\mathbb{E}(\overline{V}_\infty)$ is obtained with expression (23) of the density h_δ. The relation

$$\int_0^{+\infty} \delta^{-2n}x^2\exp(-\delta^{-2n}x^2/2)\,dx = \sqrt{\frac{\pi}{2}}\,\delta^n,$$

for $n\geq 1$ and Euler's identity (see Erdelyi [11] formula (25) page 261),

$$\sum_{n=0}^{+\infty}\frac{x^{n(n+1)/2}t^n}{\prod_{l=1}^n(1-x^l)} = \prod_{l=1}^{+\infty}\left(1+tx^l\right) \tag{26}$$

applied for $x=\delta^2$ and $t=-\delta$ give the final formula for the throughput. □

Note that Ott *et al.* [20] obtained expressions similar to (23) and (25) for a continuous system, described as the solution of a deterministic differential equation perturbated at the points of a Poisson process. For the finite time behavior, this system can be considered as a limit of the original model. If we observe the system at the level of the packets instead of the successive congestion windows, the points of the Poisson process are the instants, properly renormalized, when packets are lost. At equilibrium there is no convergence result for the invariant distributions supporting the fact that this system may be seen as a limit of the original model. Technically, the trouble comes from the deterministic differential equation

$$x'(t) = a/x(t)$$

considered in that paper. It has a singularity when $x(0)$ is close to 0. In finite time, this singularity can be controlled; for the equilibrium, i.e. at $t=+\infty$, it is less clear how one can prove that the invariant probability distributions converge to the invariant probability measure of the perturbated differential equation.

REMARKS.

a) *For the case of TCP, $\delta = 1/2$, the throughput is ~ 1.3098 which is the value observed in earlier simulations and experiments (see Floyd* [13], *Floyd* et al. [14] *and Madhavi and Floyd* [17]*).*

b) *Trite manipulations with the expression* (25) *of $\overline{\rho}(\delta)$ show that*

$$\overline{\rho}(\delta)\sim\sqrt{\frac{2}{\pi(1-\delta)}}$$

when $\delta\nearrow 1$. This does not mean that, in practice, the throughput really increases with δ. Indeed, the loss process is in fact also related to δ. The model we consider does not take into account this relation.

The continuous time process. The above results on the invariant distribution of the embedded Markov chain give the expression of the density of the invariant distribution of the continuous time process $(\overline{W}(t))$.

Proposition 16. *When the maximum window size is infinite, the rescaled asymptotic density function of the congestion window size at equilibrium $\overline{W}_\infty$ is given by, for $x \geq 0$,*

$$H_\delta(x) = \frac{\sqrt{2/\pi}}{\prod_{n=0}^{+\infty}(1-\delta^{2n+1})} \sum_{n=0}^{+\infty} \frac{\delta^{-2n}}{\prod_{k=1}^{n}(1-\delta^{-2k})} e^{-\delta^{-2n}x^2/2}. \tag{27}$$

Proof. The classical representation of the invariant measure of the continuous time process $(\overline{W}(t))$ with the invariant probability of $(\overline{V}_n)$ is given by

$$\mathbb{E}\left(f\left(\overline{W}_\infty\right)\right) = \frac{1}{\mathbb{E}\left(\overline{G}_{\overline{V}_\infty}\right)} \mathbb{E}\left(\int_0^{\overline{G}_{\overline{V}_\infty}} f\left(\overline{V}_\infty + s\right)\, ds\right), \tag{28}$$

for any non negative measurable function f on $\mathbb{R}_+$. The invariance relation (16) (with $\overline{w}_{\max} = +\infty$) gives the identity

$$\mathbb{E}\left(\overline{G}_{\overline{V}_\infty}\right) = \frac{1-\delta}{\delta} \mathbb{E}\left(\overline{V}_\infty\right),$$

the right hand side is precisely the inverse of the asymptotic throughput according relation (25). If we take $f(x) = \exp(-\lambda x)$ for x in identity (28) and $\lambda \geq 0$, the Laplace transform of $\overline{W}_\infty$ is given by

$$\begin{aligned}
\mathbb{E}\left(\exp\left(-\lambda \overline{W}_\infty\right)\right) &= \overline{\rho}(\delta)\, \mathbb{E}\left(\int_0^{\overline{G}_{\overline{V}_\infty}} \exp\left(-\lambda\left(\overline{V}_\infty + s\right)\right)\, ds\right) \\
&= \frac{\overline{\rho}(\delta)}{\lambda}\left(\mathbb{E}\left(\exp(-\lambda \overline{V}_\infty) - \exp\left(-\lambda\left(\overline{V}_\infty + \overline{G}_{\overline{V}_\infty}\right)\right)\right)\right).
\end{aligned}$$

Relation (16) shows that this last expression is

$$\begin{aligned}
\mathbb{E}\left(\exp\left(-\lambda \overline{W}_\infty\right)\right) &= \frac{\overline{\rho}(\delta)}{\lambda}\left(\mathbb{E}\left(\exp(-\lambda \overline{V}_\infty)\right) - \mathbb{E}\left(\exp\left(-\lambda \overline{V}_\infty/\delta\right)\right)\right) \\
&= \overline{\rho}(\delta)\, \mathbb{E}\left(\int_{\overline{V}_\infty}^{\overline{V}_\infty/\delta} e^{-\lambda s}\, ds\right) \\
&= \int_0^{+\infty} e^{-\lambda s} \mathbb{P}\left(\overline{V}_\infty \leq s \leq \overline{V}_\infty/\delta\right)\, ds.
\end{aligned}$$

From Relation (28) we get that the density of $\overline{W}_\infty$ is given by

$$H_\delta(x) = \overline{\rho}(\delta)\, \mathbb{P}\left(\overline{V}_\infty \leq x \leq \overline{V}_\infty/\delta\right) = \overline{\rho}(\delta)\, \mathbb{P}\left(\delta x \leq \overline{V}_\infty \leq x\right),$$

for $x \geq 0$. The expression (23) of the density of $\overline{V}_\infty$ is used to conclude the proof. □

4.2. **Finite maximum window size.**

The embedded Markov chain. In this section, we assume that the rescaled maximum window size $\overline{w}_{\max}$ is finite. As we shall see, by using equation (16), it is also possible to explicitly compute the distribution of the limiting random variable $\overline{V}_\infty$. The following proposition is the analogue of Proposition 12, it gives an explicit probabilistic representation of $\overline{V}_\infty$.

Proposition 17. *If (E_i) is an i.i.d. sequence of exponential random variables with parameter 1, the invariant probability measure of the limiting Markov chain can be represented as follows,*

$$\overline{V}_\infty \stackrel{dist.}{=} \sqrt{\inf_{n\geq 0}\left(\delta^{2n}\overline{w}_{max} + 2\sum_{i=1}^{n}\delta^{2i}E_i\right)}, \tag{29}$$

where (E_n) is an i.i.d. sequence of exponentially distributed random variables with parameter 1.

Proof. From the identity (21) and Equation (16), we get that

$$\overline{V}_\infty^2 \stackrel{\text{dist.}}{=} \delta^2 \min\left(\overline{w}_{\max}^2, 2E_1 + \overline{V}_\infty^2\right), \tag{30}$$

where E_1 is an exponential random variable with unit mean, independent of $\overline{V}_\infty$. If we iterate this equation, by induction we obtain that for $N \geq 1$,

$$\overline{V}_\infty^2 \stackrel{\text{dist.}}{=} \min_{0\leq n\leq N-1}\left(\delta^{2n}\overline{w}_{\max}^2 + 2\sum_{i=1}^{n}\delta^{2i}E_i\right) \wedge \left(\delta^{2N}\overline{V}_\infty^2 + 2\sum_{i=1}^{N}\delta^{2i}E_i\right).$$

The proposition follows by letting N go to infinity. □

Notice that when $\overline{w}_{\max}$ goes to infinity, the representation (29) converges to the expression (22) obtained for the infinite window size. We now give the explicit representation of the distribution of $\overline{V}_\infty$. The equation (30) shows that this variable has a mass at $\delta\overline{w}_{\max}$. As we shall see in the next proposition, the distribution of $\overline{V}_\infty$ is a convex combination of a Dirac mass at $\delta\overline{w}_{\max}$ and a density function on $[0, \delta\overline{w}_{\max}]$.

Proposition 18. *The distribution of $\overline{V}_\infty$ has a mass η at $\delta\overline{w}_{max}$, with*

$$1/\eta = 1 + \prod_{n=1}^{+\infty}(1-\delta^{2n})\left(\sum_{n=0}^{+\infty}\frac{1}{\prod_{k=1}^{n}(1-\delta^{2k})}\left(e^{\delta^{2n}\overline{w}_{max}^2/2} - e^{\delta^{2(n+1)}\overline{w}_{max}^2/2}\right)\right) \tag{31}$$

and a density function $\tilde{h}_\delta$ on $[0, \delta\overline{w}_{max}]$ given by

$$\tilde{h}_\delta(x) = h_\delta(x) + x\eta\sum_{n=1}^{+\infty}\left(k_n\left(x^2 - \delta^{2(n+1)}\overline{w}_{max}^2\right) - k_n\left(x^2 - \delta^{2n}\overline{w}_{max}^2\right)\right), \tag{32}$$

where h_δ is the density function given by (23) and for, $n \geq 1$, $k_n(2x)$ is the density of $\delta^2 E_1 + \cdots + \delta^{2n}E_n$ when the random variables (E_i) are i.i.d. exponentially distributed with parameter 1.

The function k_n can be expressed explicitly as a linear combination of the functions $\exp(-\delta^{-2k}x)$, $x \geq 0$, $k = 1, \ldots, n$. (Notice that k_n vanishes on $\mathbb{R}_-$).

With the expressions (31) and (32), it is easily seen that the distribution of $\overline{V}_\infty$ converges to the distribution with density h_δ when $\overline{w}_{\max}$ tends to infinity.

Proof. If we set $Z = \overline{V}_\infty^2/2$, $\beta = \delta^2$ and $w = \beta\,\overline{w}_{\max}^2/2$, the relation (30) shows that the variable Z satisfies the following relation

$$Z \stackrel{\text{dist.}}{=} \min(\beta(Z+E), w),$$

where E is an exponential variable with parameter 1 independent of Z. If ϕ denotes the Laplace transform of Z, the above equation can be written as, for $\xi \geq 0$,

$$\phi(\xi) = \mathbb{E}\left(e^{-\xi Z}\right) = \mathbb{E}\left(e^{-\xi\beta(Z+E)} 1_{\{\beta(Z+E)<w\}}\right) + e^{-\xi w} P(\beta(Z+E) \geq w).$$

By using the fact that E and Z are independent, we get

$$\phi(\xi) = \mathbb{E}\left(e^{-\xi\beta Z}\frac{\left(1-e^{-(\xi\beta+1)(w/\beta - Z)}\right)}{1+\xi\beta}\right) + e^{-\xi w}\mathbb{E}\left(e^{-(w/\beta - Z)}\right), \tag{33}$$

therefore,

$$\phi(\xi) = \frac{1}{1+\beta\xi}\phi(\xi\beta) + \frac{\xi\beta}{1+\xi\beta}e^{-(\xi+1/\beta)w}\phi(-1), \tag{34}$$

if the above relation is used recursively, we obtain for $n \geq 1$,

$$\phi(\xi) = \frac{1}{\prod_{k=1}^{n}(1+\xi\beta^k)}\phi(\xi\beta^n) + \phi(-1)e^{-w/\beta}\sum_{k=0}^{n-1}\frac{\xi\beta^{k+1}}{\prod_{i=1}^{k+1}(1+\xi\beta^i)}e^{-\xi\beta^k w},$$

since $\phi(\xi\beta^n) \to 1$ as n goes to infinity (recall that $\beta < 1$), the Laplace transform ϕ can be expressed as

$$\phi(\xi) = \frac{1}{\prod_{k=1}^{+\infty}(1+\xi\beta^k)} + \phi(-1)e^{-w/\beta}\sum_{k=0}^{+\infty}\frac{\xi\beta^{k+1}}{\prod_{i=1}^{k+1}(1+\xi\beta^i)}e^{-\xi\beta^k w}.$$

If we take $\xi = -1$ in this identity we get that

$$1/\phi(-1) = \left(1 + e^{-w/\beta}\sum_{n\geq 0}\frac{\beta^{n+1}}{\prod_{k=1}^{n+1}(1-\beta^k)}e^{\beta^n w}\right)\prod_{n=1}^{+\infty}(1-\beta^n).$$

If we set $\eta = \phi(-1)e^{-w/\beta}$, then clearly

$$\eta = \mathbb{P}(\beta(Z+E) > w) = \mathbb{P}(\overline{V}_\infty + \overline{G}_{\overline{V}_\infty} > \overline{w}_{\max}) = \mathbb{P}(\overline{V}_\infty = \delta\overline{w}_{\max}),$$

according to relation (30). It is then easily seen that ϕ can be written as

$$\phi(\xi) = \frac{1}{\prod_{k=1}^{+\infty}(1+\xi\beta^k)} + \eta\left(e^{-\xi w} + \sum_{n=1}^{+\infty}\frac{1}{\prod_{i=1}^{n}(1+\xi\beta^i)}\left(e^{-\xi\beta^n w} - e^{-\xi\beta^{n-1}w}\right)\right).$$

The above Laplace transform is then easy to invert and yields the identity (32). $\square$

The throughput. With the same argument leading to the relation (24) for the infinite maximum window size, the throughput $\rho^\alpha(\delta)$ (Definition 14) can be expressed as

$$\rho^\alpha(\delta)\mathbb{E}(\tau^\alpha) = \mathbb{E}\left(\sum_{k=0}^{\tau^\alpha - 1} W_k^\alpha\right),$$

where τ^α is the first time there is a loss when the initial window size is V_∞^α. Notice that τ^α is $G_{V_\infty^\alpha}^\alpha$ only if a loss occurs before (W_n^α) hits the level $w_{\max}^\alpha$. When the maximum window size is reached, the window size remains constant for a geometrically distributed period H^α with parameter $\exp(-\alpha w^\alpha_{\max})$. This gives the following identities

$$\tau^\alpha = G^\alpha_{V^\alpha_\infty} 1_{\left\{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} < w^\alpha_{\max}\right\}} + (w^\alpha_{\max} - V^\alpha_\infty + H^\alpha)\, 1_{\left\{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} \geq w^\alpha_{\max}\right\}}, \tag{35}$$

$$\sum_{k=0}^{\tau^\alpha - 1} W^\alpha_k = \sum_{k=V^\alpha_\infty}^{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty}} k + \left(w^\alpha_{\max} H^\alpha - \sum_{k=w^\alpha_{\max}}^{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty}} k \right) 1_{\left\{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} \geq w^\alpha_{\max}\right\}}. \tag{36}$$

If η^α is the probability of this event (by invariance $\eta^\alpha = \mathbb{P}(V^\alpha_\infty = \delta w^\alpha_{\max})$), we get

$$\begin{aligned} \alpha \mathbb{E}\left(\sum_{k=0}^{\tau^\alpha - 1} W^\alpha_k \right) &= \frac{1}{2} \mathbb{E}\left(\left(\sqrt{\alpha} V^\alpha_\infty + \sqrt{\alpha} G^\alpha_{V^\alpha_\infty} \right)^2 \right) - \frac{1}{2} \mathbb{E}\left((\sqrt{\alpha} V^\alpha_\infty)^2 \right) + o(\sqrt{\alpha}) \\ &+ \eta^\alpha \sqrt{\alpha} w^\alpha_{\max} \mathbb{E}\left(\sqrt{\alpha} H^\alpha \right) - \frac{\eta^\alpha}{2} \left(\mathbb{E}\left(\sqrt{\alpha} w^\alpha_{\max} + \sqrt{\alpha} G^\alpha_{w^\alpha_{\max}} \right)^2 - \left(\sqrt{\alpha} w^\alpha_{\max} \right)^2 \right), \end{aligned}$$

this identity is a consequence of the Markov property of (W^α_n). The variable $V^\alpha_\infty + G^\alpha_{V^\alpha_\infty}$ conditionally on the event

$$\left\{ V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} \geq w^\alpha_{\max} \right\}$$

has the same distribution as $w^\alpha_{\max} + G^\alpha_{w^\alpha_{\max}}$. The scaling relation (4) for $w^\alpha_{\max}$ implies that $\sqrt{\alpha} H^\alpha$ converges in distribution to an exponentially distributed random variable with mean $1/\overline{w}_{\max}$, using Theorem 9, we obtain

$$\begin{aligned} \lim_{\alpha \to 0} \alpha \mathbb{E}\left(\sum_{k=0}^{\tau^\alpha - 1} W^\alpha_k \right) &= \frac{1}{2} \left(\mathbb{E}\left(\overline{V}_\infty + \overline{G}_{\overline{V}_\infty} \right)^2 - \mathbb{E}\left(\overline{V}_\infty \right)^2 \right) \\ &+ \eta - \frac{\eta}{2} \left(\mathbb{E}\left(\overline{w}_{\max} + \overline{G}_{\overline{w}_{\max}} \right)^2 - \overline{w}^2_{\max} \right) = 1, \end{aligned} \tag{37}$$

by Proposition 11. Similarly for τ^α,

$$\tau^\alpha = (w^\alpha_{\max} - V^\alpha_\infty) \wedge G^\alpha_{V^\alpha_\infty} + H^\alpha 1_{\left\{V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} \geq w^\alpha_{\max}\right\}}$$

therefore

$$\begin{aligned} \mathbb{E}(\tau^\alpha) &= \mathbb{E}\left(w^\alpha_{\max} \wedge \left(V^\alpha_\infty + G^\alpha_{V^\alpha_\infty} \right) - V^\alpha_\infty \right) + \eta^\alpha \mathbb{E}(H^\alpha) \\ &= \frac{1-\delta}{\delta} \mathbb{E}(V^\alpha_\infty) + \eta^\alpha \mathbb{E}(H^\alpha) \end{aligned}$$

by the invariance relation (16) for V^α_∞.

The following proposition is therefore a consequence of the last identity and relation (37).

Proposition 19. *If $\rho^\alpha(\delta)$ is the throughput of an AIMD algorithm with multiplicative decrease factor δ and maximum window size w^α_{max}, then*

$$\lim_{\alpha \to 0} \sqrt{\alpha} \rho^\alpha(\delta) = \frac{\delta}{(1-\delta)\mathbb{E}(\overline{V}_\infty) + \delta \eta / \overline{w}_{max}}, \tag{38}$$

the constant η and the distribution of $\overline{V}_\infty$ are given by Proposition 18.

The above formulas have been used in Figure 3 and 4 to represent the dependence of the throughput and η with respect to the maximum congestion window size.

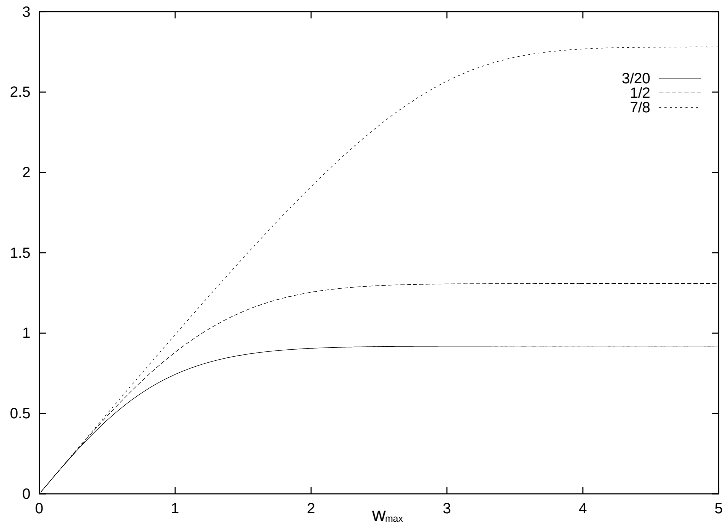

FIGURE 3. The throughput for $\delta = 3/20$, $1/2$ and $7/8$

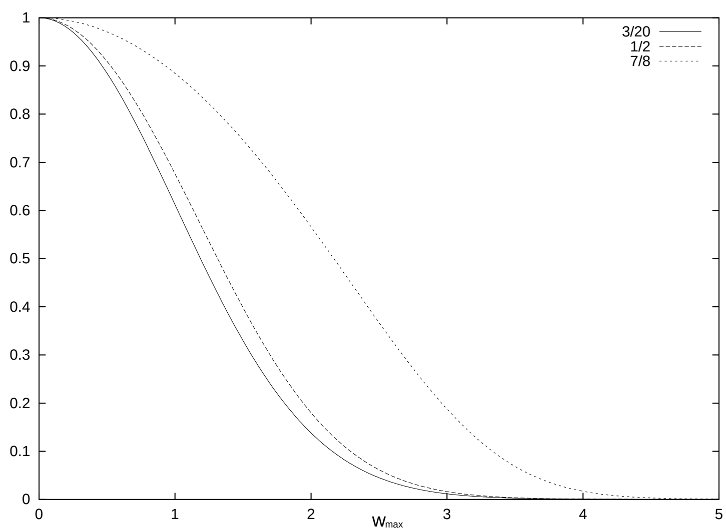

FIGURE 4. The stationary probability of hitting the maximum window size before a loss

5. THE DISTRIBUTION OF THE HITTING TIMES

In this section we study the hitting time of some level by the size of the congestion window. Its practical importance is fairly clear since the performance of the transmission is optimal when the maximal congestion window is reached.

Definition 20. *If $x>0$, S_x^α is the first time (W_n^α) reaches the level $x/\sqrt{\alpha}$,*

$$S_x^\alpha = \inf\left\{n\geq 1 : W_n^\alpha \geq x/\sqrt{\alpha}\right\} \text{ and } \overline{S}_x = \inf\left\{t>0 : \overline{W}(t)\geq x\right\}$$

The convergence result of Proposition 5 gives the key to the limiting behavior of S_x^α, it suggests that S_x^α is of order $1/\sqrt{\alpha}$. The next result show that this is indeed the case. Moreover, the limiting distribution is expressed with the help of an eigenvector of the infinitesimal generator Ω defined by (12). These results are an illustration of the interest of the functional limit theorems proved in Section 2.

Theorem 21. *If $\lim_{\alpha\to 0}\sqrt{\alpha}W_0^\alpha = x_0 < \overline{W}(0)=x$, the variable $\sqrt{\alpha}S_x^\alpha$ converges in distribution to $\overline{S}_x$ as α tends to 0. Its Laplace transform given by, for $\xi\geq 0$,*

$$\mathbb{E}\left(e^{-\xi\overline{S}_x}\right) = \frac{f_\xi(x_0)}{f_\xi(x)},$$

where f_ξ is the unique solution of the equation

$$f'(y) + yf(\delta y) = (\xi + y)f(y), \tag{39}$$

with $f(0)=1$.

Proof. For $a\geq 0$, by definition

$$\left\{\sqrt{\alpha}S_x^\alpha > a\right\} = \left\{\sup_{0\leq t\leq a}\sqrt{\alpha}W^\alpha_{\lfloor t/\sqrt{\alpha}\rfloor} < x\right\},$$

since the function $g\to\sup\{g(u);\,0\leq u\leq a\}$ is continuous on the Skorohod space of functions on $\mathbb{R}_+$ (see Ethier and Kurtz [12]), Proposition 5 shows that

$$\lim_{\alpha\to 0}\mathbb{P}\left(\sqrt{\alpha}S_x^\alpha > a\right) = \mathbb{P}\left(\sup_{0\leq t\leq a}\overline{W}(t) < x\right) = \mathbb{P}\left(\overline{S}_x > a\right)$$

the variable $\sqrt{\alpha}S_x^\alpha$ converges in distribution to $\overline{S}_x$ as α tends to 0.

We now prove that equation (39) has a unique solution. If f is such a solution, taking $g(y)=\exp(-(\xi+y)^2/2)f(y)$, we get the differential equation,

$$g'(y) = -ye^{-(1-\delta)y(\xi+(1+\delta)y)/2}g(\delta y),$$

hence

$$g(y) = e^{-\xi^2/2} - \int_0^y ue^{-(1-\delta)u(\xi+(1+\delta)u)/2}g(\delta u)\,du. \tag{40}$$

Since $\delta<1$, the above equation can be seen as a fixed point equation on the space $C([0,1/2])$ of the continuous function on $[0,1/2]$. If $\psi(g)$ denotes the right hand side of (40), it is clear that ψ is a contracting functional on $C([0,1/2])$ endowed with the uniform norm. The operator ψ has therefore a unique fixed point on $C([0,1/2])$. If g is this fixed point, then g can be continued on the real line. For $y\in\mathbb{R}_+$, according to (40), the value of $g(y)$ is expressed with the values of g on the interval $[0,\delta y]$. The existence and uniqueness of equation (39) are proved.

Using expression (20) of the infinitesimal generator Ω, equation (39) can also be written as $\Omega(f)(x)=\xi f(x)$ for $x\geq 0$. (Here $\overline{w}_{\max}$ does not play a role, it is assumed to be infinite). Using a classical result on the martingales of Markov processes (see Rogers and Williams [23] for example), we get that

$$\left(e^{-\xi\, t\wedge\overline{S}_x}f\left(\overline{W}(t\wedge\overline{S}_x)\right)\right)$$

is a local martingale. Since f is continuous on $[0,x]$, this local martingale is bounded hence a regular martingale, therefore

$$f(x_0) = \mathbb{E}\left(e^{-\xi\,\overline{S}_x} f\left(\overline{W}(\overline{S}_x)\right)\right) = f(x)\mathbb{E}\left(e^{-\xi\,\overline{S}_x}\right).$$

The theorem is proved. □

We have not been able to find a closed form expression for the solution of equation (39). Nevertheless it is possible to get some explicit results on the distribution of the hitting times $(\overline{S}_x)$. If f is the solution of the following equation, for $x>0$,

$$\Omega(f)(x) = 1,$$

with $f(0)=0$, when $\overline{W}(0)=x_0<x$ the same arguments as in the proof of the previous theorem show that $\mathbb{E}(\overline{S}_x) = f(x)-f(x_0)$. The functional equation to solve is

$$f'(x) + x(f(\delta x) - f(x)) = 1,$$

for $x\geq 0$, with $f(0)=0$. If $g(x)=\exp(-x^2/2)f(x)$, then this equation becomes

$$g(x) - \int_0^x e^{-u^2/2}\,du = \psi(g)(x) \stackrel{\text{def}}{=} -\int_0^x ue^{-(1-\delta^2)u^2/2} g(\delta u)\,du, \tag{41}$$

for $x\geq 0$. As before this fixed point equation has a unique solution which can be obtained by iteration. If this equation has some similarity with identity (40), its solution can be represented explicitly with the following trick. For a, $b\in\mathbb{R}_+$, we denote by $H[a,b]$ the function

$$H[a,b](x) = e^{-ax^2/2}\int_0^x e^{-bu^2/2}\,du,$$

for $x\geq 0$; the operator ψ applied to $H[a,b]$ gives the relation

$$\psi\left(H[a,b]\right) = \frac{\delta}{a+1-\delta^2}\Big(H\left[a+\delta^2 b+1-\delta^2, \delta^2 b\right] - H\left[0, a+\delta^2 b+1-\delta^2\right]\Big). \tag{42}$$

Definition 22. *The countable subset $\mathcal{T}$ of $\mathbb{R}_+^2$ and the function $L:\mathcal{T}\to\mathbb{R}_+$ are defined as follows:*

- *$(0,1)\in\mathcal{T}$ and $L((0,1)) = (1-\delta^2)/(1+\delta-\delta^2)$;*
- *if $z=(a,b)\in\mathcal{T}$ then the elements $e_0(z) = (a+\delta^2 b+1-\delta^2, \delta^2 b)$ and $e_1(z) = (0, a+\delta^2 b+1-\delta^2)$ are also in $\mathcal{T}$ with*

$$L(e_i(z)) = \frac{(-1)^i\delta}{a+1-\delta^2}L(z),$$

for $i=0$ and for $i=1$ if $z\neq(0,1)$.

In this manner $\mathcal{T}$ has a binary natural tree structure with $(0,1)$ as the ancestor, the children of $z\in\mathcal{T}$ are $e_0(z)$ and $e_1(z)$. Notice that the function $z\to e_0(z)$ has no fixed point and $(0,1)$ is the only one for $z\to e_1(z)$ If we combine the representation of the solution of the equation (41) by iteration together with the identity (42), we obtain the following proposition.

Proposition 23. *If* $\overline{W}(0) = x_0 < x$ *then* $\mathbb{E}(\overline{S}_x) = f(x) - f(x_0)$ *with*

$$f(x) = e^{x^2/2} \sum_{z \in \mathcal{T}} L(z) h[z](x), \tag{43}$$

where

$$h[z](x) = e^{-ax^2} \int_0^x e^{-bu^2}\, du,$$

if $z = (a, b)$ *and the set* $\mathcal{T}$ *and the function* $L(\cdot)$ *are given by definition 22.*

We finish by an estimation of the mean value of $\overline{S}_x$, it can be refined with an arbitrary precision by using the above proposition.

Corollary 24. *With the notation of the above proposition, if* $\overline{W}(0) = 0$*, the inequality* $m(x) \le \mathbb{E}\left(\overline{S}_x\right) \le M(x)$ *holds, with*

$$m(x) = \frac{e^{x^2/2}}{1-\delta^2}\left((1-\delta-\delta^2)\int_0^x e^{-u^2/2}\, du + \delta \int_0^x e^{-\delta^2 u^2/2}\, du\right),$$

$$M(x) = e^{x^2/2} \int_0^x e^{-u^2/2}\, du.$$

Proof. Since the solution of the fixed point equation (41) is clearly non negative, the relation

$$g(x) \le \int_0^x e^{-u^2/2}\, du,$$

holds, therefore the upper bound is true. This inequality applied in the right hand side of (41) gives the lower bound. ☐

The set $\mathcal{T}$ is apparently not easy to describe explicitly (if we forget the tree structure).

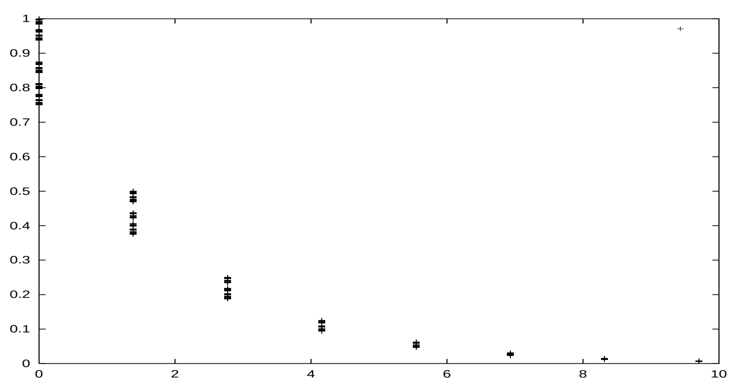

FIGURE 5. The set $\mathcal{T}$ for $\delta = 1/2$

We conclude with a convergence result on the hitting times of the embedded Markov chain $(\overline{V}_n)$,

$$\tau_x^\alpha = \inf\left\{n \ge 1 : V_n^\alpha \ge x/\sqrt{\alpha}\right\} \text{ and } \overline{\tau}_x = \inf\left\{n \ge 1 : \overline{V}_n \ge x\right\}.$$

Theorem 25. *If* $\lim_{\alpha\to 0}\sqrt{\alpha}V_0^\alpha = x_0 < x$*, the variables* τ_x^α*,* $\alpha > 0$ *converges in distribution to* $\overline{\tau}_x$ *when* $\alpha \to 0$*. The generating function of* $\overline{\tau}_x$ *is given by, for* $0 < u < 1$*,*

$$\mathbb{E}\left(u^{\overline{\tau}_x}\right) = \frac{f_u(x_0)}{f_u(x)}, \text{ with } f_u(x) = \sum_{n\geq 0} \frac{u^{n+1}}{\prod_{k=1}^{n}(1-\delta^{2k})} e^{x^2\delta^{2n}/2}, \tag{44}$$

Proof. The convergence is fairly clear from Corollary 4. The equation

$$u\,\mathbb{E}\left(f\left(\overline{V}_1\right) \middle| \overline{V}_0 = x\right) = f(x) \tag{45}$$

is equivalent to

$$f'(x) = xf(x) - uxf(\delta x).$$

The function defined by (44) clearly satisfies such an equation, therefore the identity (45) implies that the sequence

$$\left(u^{n\wedge\overline{\tau}_x} f_u\left(\overline{V}_{n\wedge\overline{\tau}_x}\right)\right)$$

is a bounded martingale, we conclude by taking its expected value at infinity. □

References

1. Cédric Adjih, Philippe Jacquet, and Nikita Vvedenskaya, *Performance evaluation of a single queue under multi-user TCP/IP connections*, Tech. Report RR-4141, INRIA, March 2001.
2. M. Allman, V. Paxson, and W. Stevens, *TCP congestion control*, RFC 2581, April 1999, URL `http://www.rfc-editor.org/rfc/rfc2581.txt`.
3. Eitan Altman, Konstantin Avrachenko, and Chadi Barakat, *A stochastic model of TCP/IP with stationary random losses*, ACM-SIGCOMM'00 (Stockholm), no. 4, 2000, pp. 231–242.
4. Eitan Altman, Kostya Avrachenkov, Chadi Barakat, and Rudesindo Nunez Queija, *State-dependent M/G/1 type queueing analysis for congestion control in data networks*, Proceedings of IEEE INFOCOM (Anchorage, Alaska), April 2001 2001.
5. Søren Asmussen, *Applied probability and queues*, John Wiley & Sons Ltd., Chichester, 1987.
6. François Baccelli and Dohy Hong, *TCP is max-plus linear*, ACM-SIGCOMM'00 (Stockholm), no. 4, 2000, pp. 219–230.
7. Patrick Billingsley, *Convergence of probability measures*, Wiley series in probability and mathematical statistics, John Wiley & Sons Ltd, New York, 1968.
8. Patrick Brown, *Resource sharing of tcp connections with different round trip times*, Proc. IEEE Infocom (Tel-Aviv, Israel), March 2000.
9. Vincent Dumas, Fabrice Guillemin, and Philippe Robert, *Limit results for Markovian models of TCP*, March 2001, URL `http://algo.inria.fr/robert/src/papers/tcp/globe.ps.gz`.
10. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996.
11. Arthur Erdélyi, Wilhelm Magnus, Fritz Oberhettinger, and Francesco G. Tricomi, *Higher transcendental functions. Vol. III*, McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955, Based, in part, on notes left by Harry Bateman.
12. Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons Inc., New York, 1986.
13. Sally Floyd, *Connections with multiple congested gateways in packet-switsched networks part 1: One way traffic*, Computer Communication Review **21** (1991), no. 5, 30–47.
14. Sally Floyd, Mark Handley, and Jitendra Padhye, *A comparison of equation-based AIMD congestion control*, Preprint, May 2000.
15. V. Jacobson, *Congestion avoidance and control*, SIGCOMM '88 Symposium: Communications Architectures and Protocols, ACM, August 1988.
16. Franck P. Kelly, *Mathematical modeling of the internet*, Mathematics Unlimited - 2001 and Beyond (B. Engquist and W. Schmid, eds.), Springer Verlag, Berlin, 2001, pp. 685–702.
17. J. Madhavi and S. Floyd, *TCP-friendly unicast rate-based flow control*, End2end-interest mailing list, January 1997.
18. S. Meyn and R. Tweedie, *Markov chains and stochastic stability*, Communications and control engineering series, Springer, 1993.

19. Esa Nummelin, *General irreducible Markov chains and nonnegative operators*, Cambridge University Press, Cambridge, 1984.
20. Teunis J. Ott, J.H.B. Kemperman, and Matt Mathis, *The stationary behavior of ideal TCP congestion avoidance*, Unpublished manuscript, August 1996.
21. J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, *Modeling TCP throughput: A simple model and its empirical validation*, IEEE/ACM Transactions on Networking **8** (2000), 133–145.
22. James Roberts and Laurent Massoulié, *Bandwidth sharing and admission control for elastic traffic*, Proc. Infocom'99, 1998.
23. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 2: Itô calculus*, John Wiley & Sons Inc., New York, 1987.
24. W. Richard Stevens, *TCP-IP illustrated vol. 1 : the protocols*, Professional Computing Series, Addison-Wesley, 1994.
25. M. Vojnović and J.-Y. Le Boudec, *Some observations on equation-based rate control*, Proc. ITC'17 (Salvador de Bahia, Brazil), 2001.
26. M. Vojnović, J.-Y. Le Boudec, and C. Boutremans, *Global fairness of additive-increase and mutliplicative-decrese with heteregeneous round trip times*, IEEE Infocom (Tel Aviv, Israel), March 2000.

(V. Dumas) MAB, Université Bordeaux I, 351, cours de la Libération, 33405 Talence Cedex, France

E-mail address: dumas@math.u-bordeaux.fr

Current address, V. Dumas: INRIA, domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France

(F. Guillemin) France Telecom R&D, DAC/CPN, 2, Avenue Pierre Marzin, 22 300 Lannion, France

E-mail address: Fabrice.Guillemin@rd.francetelecom.fr

(Ph. Robert) INRIA, domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France

E-mail address: Philippe.Robert@inria.fr

URL: http://algo.inria.fr/robert/