Analysis of AIMD protocols over paths with variable delay

Eitan Altman, Chadi Barakat, Víctor Ramos

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Analysis of AIMD protocols over paths with variable delay

Eitan Altman — Chadi Barakat — Víctor Ramos

N° 5232

June 2004

Thème COM

ISSN 0249-6399 ISRN INRIA/RR--5232--FR+ENG

Analysis of AIMD protocols over paths with variable delay

Eitan Altman , Chadi Barakat , Víctor Ramos*†

Thème COM —Systèmes communicants
Projets Mistral et Planète.

Rapport de recherche n° 5232 —June 2004 —23 pages

Abstract: The throughput of AIMD protocols in general and of TCP in particular, has been computed in many existing works by modeling the round-trip time as a constant and thus replacing it by its expectation. There are however many scenarios in which the delays of packets vary, causing a variation of the round-trip time. Many typical scenarios occur in wireless and mobile networks. We propose in this paper an analytical model that accounts for the variability of delay, while computing the throughput of an AIMD protocol. We derive a closed-form expression for the throughput, that illustrates the impact of delay variability. We show by analysis and simulation, that an increase in the variability of delay improves the performance of an AIMD protocol. Thus, an analytical model that only considers the average delay could underestimate the performance of an AIMD protocol in scenarios where delay is variable.

Key-words: TCP, delay variability, stochastic recursive equations.

This is an extension of a paper that appeared at INFOCOM'04 which was restricted to a Poisson loss process with fixed parameter.

* University of Nice–Sophia Antipolis.
† Also with Universidad Autónoma Metropolitana, Mexico.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Analyse de protocoles AIMD sur des trajectoires à délai variable

Résumé : Les modèles traditionnels des protocoles AIMD, et en particulier ceux du protocole TCP, calculent le débit en modélisant le délai d'aller-retour comme étant une constante, ainsi celui-ci est remplacé par son espérance. Pourtant, il existe plusieurs scénarios pour lesquels le délai des paquets varie, ce qui cause une variation du temps d'aller-retour. Les réseaux mobiles et les réseaux sans fil sont les meilleurs exemples de ce type de scénario. Nous proposons dans ce papier, un modèle analytique prenant en compte la variabilité du délai pour calculer le débit d'un protocole AIMD. Nous obtenons une expression close pour le débit, et montrons par analyse et simulation qu'une augmentation de la variabilité du délai améliore les performances d'un protocole AIMD. En conséquence, un modèle analytique qui ne considère que le délai moyen pourrait sous-estimer les performances de ce type de protocole dans des scénarios où le délai est variable.

Mots-clés : TCP, variabilité du délai, équations récursives stochastiques.

1 Introduction

In computing the throughput of long-lived AIMD connections (and of TCP connections in particular), existing analytical models do not take into account moments of the delay other than the first one. In a recent paper [7], it was observed however that the variability of this quantity impacts the throughput performance. In [7], a model is proposed which is partly analytical and partly empirical: it uses as parameters the probabilities of having a single, a double and a triple loss event, which should be inferred from the trace; it is through these parameters that the variability of delay is accounted for. Then in the rest of the derivation there, delays are replaced by their expectations.

Paths with high variability of delay are common in communication networks, and are typical to wireless networks [7]. Here are some examples of scenarios where the delays of packets can vary. First consider the High Data Rate (HDR) systems. The distribution of delays for these systems is described in [13]. HDR is a Qualcomm proposed CDMA air interface standard (3G1x-EVDO) for supporting high speed asymmetrical data. In HDR, the reason for variability of delay is the fact that the (link layer) packet to be transmitted is chosen dynamically among various connections, according to the channel that has the best state. Another source of delay variation on wireless links is ARQ (at the link level); ARQ can add considerable delay during retransmission, especially on geostationary satellite links where the propagation delay is large. The mobility of users in a mobile network is also an important source for delay variability. A situation that adds to variability in the delays is when a TCP connection has lower priority with respect to other connections which have highly variable transmission rate. Such a situation occurs in UMTS where data packets are most frequently transmitted over shared channels (the FASH and RASH channels) in which higher priority is given to control packets. The delays of packets are also variable because of queuing time in routers. Generally speaking, the variability of delay is a common phenomenon in communication networks, and its consideration in the analytical models for AIMD protocols is important.

We propose in this paper a model for the performance of a window-based AIMD mechanism in presence of variable delay. The model is based on stochastic difference equations. We provide closed-form expressions for the moments of the window size in steady state, as well as for the throughput of the mechanism. The model is validated with simulations using the TCP protocol, which has features of an AIMD policy in its steady state [10].

One of the key results of our paper is that the performance we obtain when considering the variability of delay is better than the one we obtain when we assume the delay to be constant and thus replacing the delay by its average. This result is very important since it means that actual models for TCP, which only consider the mean delay, underestimate the performance of the protocol in environments where the packet delay is variable. Underesti-

mating the performance of AIMD protocols may have a serious impact on the dimensioning of networks, and on the development of TCP-friendly multimedia applications.

In the next section we present our model, then we derive the expectation of its stationary window size in Section 3. The analysis of the throughput and of its dependence on delay variability follow in Sections 4 and 5. Section 6 validates our analytical results using ns-2 [11] simulations, and Section 7 ends the paper with some concluding remarks.

2 The model

We consider systems for which the loss model at the packet level can be described by a Poisson process independent of the window size with a stochastic intensity. The average rate of loss events at the nth round-trip time is λ_n. A loss event in our case corresponds to the loss of a packet.

Our model studies a general window-based fluid AIMD mechanism. It applies to the TCP protocol when the window size is large enough so that the packet nature of TCP is diluted. The specification of our model to TCP will be described throughout the text. The TCP protocol will be used at the end of the paper for the validation of our results. Recall that a TCP connection in its steady state, and in the absence of timeouts and limitation on the throughput caused by the receiver window, can be seen as an AIMD protocol, where the congestion window increases by a constant factor every round-trip time, and where it is divided multiplicatively in presence of loss events [1, 10]. In particular, a TCP connection where the receiver acknowledges every packet, increases its congestion window by one packet every round-trip time, and divides its window by two when a packet loss is detected. The Reno version of TCP divides its window by two for every packet loss [8]. The Newreno and SACK versions divide their windows at most once by two in a round-trip time, regardless of the number of packet losses during the round-trip time [8]. We model both types of TCP behavior in this paper, while giving a particular attention to new versions as Newreno and SACK.

2.1 Stochastic recursive equations

We model the variable delay/rate path as follows. We consider some sequence of instants T_n and define the nth interval as $[T_n, T_{n+1})$. Let $R_n = T_{n+1} - T_n$ be its duration. R_n models the sequence of round-trip times seen by the AIMD control mechanism. We consider the window size W_n at the "end" of the nth interval. W_n is the window size just before instant T_{n+1}. (R_n, λ_n) is some stationary ergodic sequence of random variables. The window is a real number and is measured without loss of generality in terms of packets.

We consider some additive constant $\beta > 0$, and we assume that in the absence of loss, it takes R_n time to the AIMD protocol to increase the window size by β, i.e.,

$$W_{n+1} = W_n + \beta.$$

For example, on a long-lived TCP connection operating in congestion avoidance without the delayed ACK feature, we could take $\beta = 1$, in which case R_n would correspond to a round-trip time (as the window size of TCP increases by roughly one packet every round-trip time). Even though we frequently consider R_n as being the round-trip time, our model

does not require that, and other definitions of R_n are possible. Another definition of R_n might be useful for congestion control mechanisms other than TCP.

We study now the dynamics of the window size when losses occur. Define $Z_n = 1$ if at least one loss occurred during $[T_n, T_{n+1})$ and $Z_n = 0$ otherwise, and set $\overline{Z}_n = 1 - Z_n$. Note that

$$P(Z_n = 0) = \mathbb{E}[\exp(-\lambda_n R_n)] := \mathcal{R}^*(\lambda),$$

λ is some parametric vector describing the intensity of the loss process and to be introduced later.

Consider now that at least one loss occurs during $[T_n, T_{n+1})$. Then the window size at the end of the $(n+1)$th interval is

$$W_{n+1} = \gamma_n W_n + \beta,$$

where γ_n is a random multiplicative factor that allows to account for cases where the multiplicative decrease of the AIMD mechanism is a function of the number of loss events that occur during the nth interval. Later, we will detail on this issue, and provide the expression for γ_n in the case of TCP.

Our dynamics can be interpreted as a simplified model for AIMD in which the multiplicative decrease occurs at the end of the round-trip time. Time interval R_n ends with a window size equal to W_n, and time interval R_{n+1} starts by a window size equal to $\gamma_n W_n$, to end with a window size equal to W_{n+1}. By dividing the window at the end of the round-trip time, we better model the fact that losses are not detected instantaneously. As for the window growth by β, we suppose that is done progressively during the round-trip time

Using the above two expressions for the case of no loss and for the case of at least one loss, we obtain the following dynamics:

$$W_{n+1} = A_n W_n + \beta, \tag{1}$$

where

$$\begin{aligned} A_n &= \overline{Z}_n + \gamma_n Z_n = 1 + Z_n(\gamma_n - 1) \\ &= \gamma_n + \overline{Z}_n(1 - \gamma_n). \end{aligned} \tag{2}$$

(1) is known as a stochastic recursive equation, see [1, 3, 5, 6, 9]. The loss process is assumed to be Poisson with an intensity that depends on the current round-trip time, and the process (R_n, λ_n) is assumed to be stationary ergodic. Moreover, for any integer n, the loss process after time T_n does not depend on $W_j, j \leq n$. Then, A_n is stationary ergodic.

Similar recursive equations have been used in the past to analyze the throughput of TCP and of TCP-friendly applications, see e.g. [2, 14]. The special feature of our present model is that A_n is random and depends on R_n. Another feature is that the increase in the window size between instants $\{T_n\}$ is constant, and is independent of the duration of round-trip times.

Our model can be easily specified to TCP. The process R_n can be seen as the round-trip time of the TCP connection. A loss event corresponds to the loss of a TCP packet. The additive increase constant β is roughly equal to one packet. The multiplicative decrease constant γ_n depends on the TCP version. Some versions as Reno divide their windows by two for every packet loss; in this case $\gamma_n = (1/2)^{N_n}$, where N_n is the number of loss events in the nth round-trip time. Other versions of TCP as Newreno and SACK divide their windows by two whether there is one or more loss events in a round-trip time. For these later versions, γ_n is constant equal to one half.

Our model does not account for some TCP mechanisms as the slow start phase, the timeouts, the receiver window, and so on. Our objective is not to provide an accurate model for TCP, but rather to illustrate the impact of delay variability on protocols implementing the AIMD mechanism. TCP is a typical example of such protocols. For accurate models of TCP that consider some of the non AIMD features (but that do not consider the variability of delay), we refer to [2, 12].

3 Stationary window size analysis

We begin by computing the stationary distribution and moments of the window size at times T_n. We also compute time average quantities. The throughput of the AIMD mechanism is then given in the next section for the case when the sequence $\{R_n\}$ models the round-trip times.

Concerning the process R_n, we consider two cases: $\{R_n\}$ are i.i.d. (independent and identically distributed), and $\{R_n\}$ are Markov correlated. In the i.i.d. case, we obtain closed-form expressions for the throughput and for the moments of the window size. In the Markov correlated case, we obtain linear equations that can be solved for the moments of the window size and for the throughput.

3.1 Window size at times T_n

Applying Theorem 2.A of [9] for which the conditions are easily checked, we get:

Theorem 1. *There exists a unique stationary ergodic process W_n^* that satisfies the same recursion (1) as W_n, and that is defined on the same probability space as (W_n, A_n). For any initial value W_0, $\lim_{n\to\infty} |W_n - W_n^*| \to 0$ and W_n converges to W_n^* in distribution. W_n^* has the explicit form:*

$$W_n^* = \sum_{j=0}^{\infty} \left(\prod_{l=n-j}^{n-1} A_l \right) \beta.$$

To simplify the exposition of the analysis, we consider hereafter that the system is in its stationary regime at time $t = 0$. First, we present the results for the case when the random variables $\{R_n\}$ are i.i.d. and $\lambda_n = \lambda$ are constant (do not depend on n). Then, we explain how to compute the moments of W_0^* in the case the process $\{R_n, \lambda_n\}$ is Markov correlated. Through the analysis, we consider two values of γ_n:

A1.i: The AIMD protocol decreases the window size by a constant factor γ, independently of the number of losses during the round-trip time (provided there is at least one). This models the new versions of TCP as Newreno and SACK [8].

A1.ii: The AIMD protocol decreases its window by a constant factor γ for every loss, which gives $\gamma_n = \gamma^{N_n}$. N_n denotes the number of packets lost in round-trip time R_n. This can be assumed to model old versions of TCP as Reno [8]. In contrast to new versions, the Reno version of TCP can divide its window by more than two in a round-trip time depending on the number of packets lost and the location of these losses in the congestion window.

3.1.1 The i.i.d. case

Taking expectation in (1), we get in the stationary regime:

$$\mathbb{E}[W_0^*] = \frac{\beta}{1 - \mathbb{E}[A_0]} = \frac{\beta}{1 - \mathbb{E}\left[\gamma_0 + \overline{Z}_0(1-\gamma_0)\right]}. \tag{3}$$

This is the average window size sampled just before time T_1 (end of time interval R_0). We give in the following the expression of $\mathbb{E}[W_0^*]$ for the two particular values of γ_0 cited in A1.i and A1.ii. Under A1.i we have

$$\mathbb{E}[A_0] = \gamma + \mathcal{R}^*(\lambda)(1-\gamma).$$

Hence,

$$\mathbb{E}[W_0^*] = \frac{\beta}{1 - \mathbb{E}[A_0]} = \frac{\beta}{1-\gamma} \times \frac{1}{1 - \mathcal{R}^*(\lambda)}, \tag{4}$$

whereas under A1.ii we have, for $k = 0, 1, 2, ...$,

$$P(A_0 = \gamma^k) = P(N_0 = k) = E\left[\frac{(\lambda R_0)^k}{k!}\exp(-\lambda R_0)\right],$$

so that

$$\mathbb{E}[A_0] = E\left[\sum_{k=0}^{\infty} \frac{(\gamma\lambda R_0)^k \exp(-\lambda R_0)}{k!}\right] = \mathcal{R}^*(\lambda(1-\gamma)).$$

Thus

$$\mathbb{E}[W_0^*] = \frac{\beta}{1 - \mathcal{R}^*(\lambda(1-\gamma))}.$$

3.1.2 The correlated case

We model here the correlation that may exist among $\{(R_n, \lambda_n)\}$. We explain how to compute the moments of the window size in presence of such correlation.

Consider an N-state ergodic Markov chain $\zeta(n)$ embedded at times T_n, with transition probabilities P_{ij} and with steady state probabilities π_j, $j = 1, ..., N$. Assume that the distribution of the couple (R_n, λ_n) is only a function of the state of the Markov chain at time T_n and not of the previous history. Hence, given the state of the Markov chain at times T_n and T_m, $m > n$, the coefficients A_n and A_m are independent. We denote by $(R(i), \lambda(i))$ the values of this couple when the Markov chain is at state i.

Our goal is to compute $\mathbb{E}[W_0^*]$. Later, this will serve to compute the throughput of the AIMD mechanism. Define $w_j = \mathbb{E}[W_0^* 1\{\zeta(0) = j\}]$, and define $a_j = \mathbb{E}[A_0|\zeta(0) = j]$. By using the recurrence (1), we have

$$\begin{aligned} w_j &= \mathbb{E}[W_1^* 1\{\zeta(1) = j\}] \\ &= \sum_{i=1}^{N} \mathbb{E}[A_0 W_0^* 1\{\zeta(0) = i\} 1\{\zeta(1) = j\}] + \pi_j \beta \\ &= \sum_{i=1}^{N} a_i w_i P_{ij} + \pi_j \beta. \end{aligned}$$

We obtain a system of linear equations,

$$(I - \mathcal{A})\underline{w} = \mathcal{B} \tag{5}$$

where $\mathcal{A}_{ij} = a_i P_{ij}$, $\mathcal{B}_i = \pi_i \beta$ and $\underline{w} = \{w_i\}$. We may obviously assume that at least for some i, $a_i < 1$ (for either A1.i or A1.ii). Hence $\mathcal{A}$ is a strictly sub-stochastic matrix and its largest eigenvalue is strictly smaller than one. Hence, Equation (5) has a unique solution $\underline{w}$, and finally we obtain $\mathbb{E}[W_0^*] = \sum_{i=1}^{N} w_i$.

Denote by $\mathcal{R}_i^*(s)$ the LST of the round-trip time R_n given that the Markov chain is in state i, i.e., $\mathcal{R}_i^*(s) := \mathbb{E}[\exp(-sR_0)|\zeta(0) = i]$. Then under A1.i we have,

$$a_i = (\gamma + \mathcal{R}_i^*(\lambda(i))(1 - \gamma)). \tag{6}$$

Under A1.ii we have,

$$a_i = \mathcal{R}_i^*(\lambda(i)(1 - \gamma)). \tag{7}$$

3.2 Window size at random time

For the completeness of the study, we compute in this section the expectation in the stationary regime of the process $W(t)$ of the window size. The time average window size is equivalent to the average number of packets in the network at random time. Using Palm calculus, this is given by

$$\mathbb{E}[W(t)] = \frac{\mathbb{E}[S_1]}{\mathbb{E}[R_1]}, \text{ where } S_1 = \int_{T_1}^{T_2} W(s)ds.$$

To sum the window size between T_1 and T_2 (beginning and end of round-trip time R_1), we make the following assumption:

A2: The window size grows linearly during the interval $[T_n, T_{n+1})$ with rate β/R_n, and only at the end of the interval, the window size will decrease if there has been a loss during the interval.

Under A2 we have,

$$S_1 = R_1 \left(A_0 W_0 + \frac{\beta}{2} \right).$$

3.2.1 The $i.i.d.$ case

We consider the case where $\{R_n\}$ are i.i.d. and λ is constant. Although there is a dependence between W_n and R_{n-1}, R_{n-2}, ..., there is no dependence between W_n and R_n, R_{n+1}, ... Thus,

$$\mathbb{E}[W(t)] = \frac{\beta}{2} + \mathbb{E}[A_0]\,\mathbb{E}[W_0^*],$$

where $\mathbb{E}[W_0^*]$ is given by (4), and

$$\mathbb{E}[A_0] = \begin{cases} \gamma + \mathcal{R}^*(\lambda)(1-\gamma), & \text{under A1.i.} \\ \mathcal{R}^*(\lambda(1-\gamma)), & \text{under A1.ii.} \end{cases}$$

3.2.2 The correlated case

We consider the same model for R_n as that in Section 3.1.2. Our problem is to compute $\mathbb{E}[R_1 A_0 W_0]$. We condition on the state of the Markov chain at T_0. This gives,

$$\mathbb{E}[R_1 A_0 W_0] = \sum_{i=1}^{N} w_i a_i \sum_{j=1}^{N} P_{ij} \mathbb{E}[R_1 | \zeta(1) = j].$$

The w_i can be obtained by solving the system of N linear equations in Section 3.1.2. The a_i are given in (6) and (7) for cases A1.i and A1.ii. We put everything together, which gives

$$\mathbb{E}[W(t)] = \frac{\beta}{2} + \frac{1}{\mathbb{E}[R_1]} \sum_{i=1}^{N} w_i a_i \sum_{j} P_{ij} \mathbb{E}[R_1 | \zeta(1) = j].$$

4 Throughput and square-root formula

We compute in this section the throughput of the AIMD mechanism. Consider in what follows the case of i.i.d. round-trip times and λ_n constant equal to λ. This will allow a nice closed-form relating the throughput of an AIMD mechanism like TCP to the packet loss ratio p, and that accounts for the variability of the delay, not simply its average value as in previous models. In the case of correlated round-trip times, we compute the throughput numerically.

A window-based flow control mechanism transmits a window size of packets in every round-trip time. If we look at our model, W_n packets are transmitted in the interval $[T_n, T_{n+1})$. The connection's throughput (in fact it is the sending rate) is simply equal to the average window size $\mathbb{E}[W_0^*]$ computed in Section 3.1 divided by the average round-trip time $\mathbb{E}[R_0]$. Denote by X the throughput of the connection (in packets per second). Therefore, $X = \mathbb{E}[W_0^*]/\mathbb{E}[R_0]$.

Consider in what follows the i.i.d. case, which will allow a nice closed-form expression of the throughput. The Markov correlated case is more complex and requires numerical analysis. We have in the i.i.d. case:

$$X = \frac{\mathbb{E}[W_0^*]}{\mathbb{E}[R_0]} = \frac{\beta}{\mathbb{E}[R_0](1 - \mathbb{E}[A_0])}. \tag{8}$$

It is clear from (8) that the throughput of an AIMD mechanism changes with the variability of the round-trip time. This change is caused by the term $\mathbb{E}[A_0]$ in the denominator of X. In the next section, we will study the relation between this term and the variability of the delay, and in consequence the relation between the variability of the delay and the throughput.

The following analysis holds for A1.i, which is the case when the AIMD mechanism divides its window independently of the number of packets lost in the round-trip time. Under A1.i, we have

$$X = \frac{\beta}{1 - \gamma} \times \frac{1}{\mathbb{E}[R_0](1 - \mathcal{R}^*(\lambda))}. \tag{9}$$

Let us establish the relation between the throughput and the packet loss ratio. Such relation is usually used while modeling the TCP protocol. It is well known in the networking community that the throughput of a long-lived TCP connection (at least in the steady phase where there are no timeouts, no slow-start phases, and no limitation caused by the receiver window) is inversely proportional to the square-root of the packet loss ratio, and to the average round-trip time [1, 12]. We show here what this relation becomes when delay is variable.

Let p denote the packet loss ratio. As before, λ denotes the average rate of loss events. The expressions of TCP throughput in the literature are derived in the case when a packet loss results in a division of the window size. p is not the packet loss probability, but rather the probability that a packet loss results in a division of the window size. Recall that we are working under A1.i. This requires to compute the average rate of loss events that result in a division of the window size. Denote this average rate by $\lambda' \leq \lambda$. This average rate is equal to,

$$\lambda' = \frac{P(Z_0 = 1)}{\mathbb{E}[R_0]} = \frac{1 - \mathcal{R}^*(\lambda)}{\mathbb{E}[R_0]}.$$

Another expression of λ' is $\lambda' = pX$. By equating these two expressions, we obtain

$$1 - \mathcal{R}^*(\lambda) = \mathbb{E}[R_0]\, pX.$$

We substitute this expression in (9), which yields

$$X = \frac{1}{\mathbb{E}[R_0]} \sqrt{\frac{\beta}{(1-\gamma)p}}.$$

This relation is very interesting since it tells us that in an environment where the delay is variable, the throughput is always inversely proportional to the average round-trip time and to the square root of p. The impact of delay variability on the throughput figures in p. This probability represents how many times the window of the AIMD mechanism is divided. The average loss rate λ represents how many packets are lost. The mapping between packet losses and window divisions is dependent, not only on the average round-trip time, but also on its variability. This issue has been addressed in [7], and the probability p has been computed empirically. An interesting result of our model is that it provides a closed-form expression for computing p, without doing measurements. Indeed, if we know the rate of loss events λ and the distribution of round-trip times, and if round-trip times are i.i.d., we can compute p as

$$p = \frac{\lambda'}{X} = \frac{1-\gamma}{\beta}(1 - \mathcal{R}^*(\lambda))^2.$$

This expression of p only holds for A1.i, which models new versions of TCP (Newreno, SACK) that do not divide their windows more than once per round-trip time, regardless of the number of packets lost. It does not hold for A1.ii. Indeed, under A1.ii, every packet lost results in a division of the congestion window, and so $p = \lambda/X$. By substituting in (8), we get

$$X = \frac{\beta}{\mathbb{E}[R_0]\,(1 - \mathcal{R}^*(pX(1-\gamma)))}.$$

This is an implicit equation in X. The square root relation between X and p does not hold in this context.

The square root formula for TCP throughput we present in this section suggests one more thing. If the probability p with which a TCP packet causes a division of the congestion window is constant independent of the round-trip time, the throughput of new TCP versions as Newreno and SACK will depend on the round-trip time only through its average, and so the existing models in the literature hold in this case. The existing models do not hold in the other cases since the distribution of round-trip times is to be considered.

5 Dependence of throughput on delay variability

The analysis we present in this section is done under A1.i for i.i.d. round-trip times and $\lambda_n = \lambda$. It also holds under A1.ii for i.i.d. round-trip times. At the end of the section, we comment on the validity of the result for any stationary ergodic process of round-trip times.

We study here the impact of variability of delay on the expected window size and on the throughput. We consider the expectation of the window size just before times T_n. As we saw above, the throughput of the AIMD mechanism, X, is proportional to the expected window size, so studying one of the two quantities is equivalent to studying the other.

Using our above results, we conclude the following:

Theorem 2. *Consider two AIMD systems having the same loss process and the same average delay. Both systems are identical. Let R_n and $\overline{R}_n$ be their round-trip times and suppose them to be i.i.d. Denote $\mathcal{R}^*(\lambda) = \mathbb{E}[\exp(-\lambda R_n)]$ and $\overline{\mathcal{R}}^*(\lambda) = \mathbb{E}\left[\exp(-\lambda \overline{R}_n)\right]$. Assume that*

$$\mathcal{R}^*(\lambda) \geq \overline{\mathcal{R}}^*(\lambda). \tag{10}$$

Let W_n (resp. $\overline{W}_n$) be the window process corresponding to R_n (resp. $\overline{R}_n$). The throughputs of the two systems are X and $\overline{X}$. We have the following,

$$\mathbb{E}[W_0^*] \geq \mathbb{E}\left[\overline{W}_0^*\right], \qquad X \geq \overline{X}.$$

Proof: The proof easily follows from (4) and (9). ■

We explain now the relation between (10) and the variability of round-trip times. A popular measure of variability of random variables is the convex increasing stochastic order. We say that the variable R_n is larger than the variable $\overline{R}_n$ in the convex increasing order (or more variable) if for any convex increasing function h, we have $\mathbb{E}[h(R_n)] \geq \mathbb{E}\left[h(\overline{R}_n)\right]$. We denote this by $R_n \geq_{conv} \overline{R}_n$. R_n is greater than $\overline{R}_n$ in the increasing convex order if and only if there exists a joint probability space such that $\overline{R}_n \leq \mathbb{E}\left[R_n | \overline{R}_n\right]$. For more details see [4, Chp.4 (2.3.2)].

Since the function $g(R) := \exp(-\lambda R)$ is convex in R, we then obviously have

Remark 1. Either one of the following is a sufficient condition for (10):
(i) $R_n \geq_{conv} \overline{R}_n$, or
(ii) Let $\overline{R}$ be a constant and let $\overline{R} = \overline{R}_n \leq \mathbb{E}[R_n]$.

The first property ensures that the variability of R_n is larger than that of $\overline{R}_n$. If it is the case, the condition (10) is satisfied and the throughput of the AIMD mechanism in presence of R_n is larger than the one in presence of $\overline{R}_n$. The second property, combined with Theorem 2, implies that if we replace delays by their expectations, the throughput of an AIMD mechanism decreases. Our main result is then: *the larger the variability of the delay, the better the throughput of AIMD mechanisms in general, and of TCP in particular.* A related result has been found in [2], but in another context. [2] shows that the larger the variability of times between loss events, the better is the throughput of TCP.

Consider yet the case of i.i.d. round-trip times. When the average time between packet losses is large compared to the average round-trip time, the i.i.d. case converges to the constant round-trip time case. Indeed, in (9), the term $\mathcal{R}^*(\lambda)$ can then be approximated by the first two terms of the Taylor expansion, i.e. by $1 - \lambda\mathbb{E}[R_0]$. This leads to a throughput $X = \beta/((1-\gamma)\lambda\mathbb{E}^2[R_0])$, which equals what we obtain in the constant round-trip time case. The variability of the round-trip time in this case has almost no impact on the throughput, and the round-trip time can be safely substituted by its average. Thus, models assuming constant round-trip times hold in the case of i.i.d. round-trip times that are on average small compared to times between loss events. In all other cases where the delay varies on time scales of the order of the average time between packet losses, a model as the one we propose in this paper is required.

6 Numerical results

6.1 Simulated scenario

Instead of creating an artificial model for a variable delay, we preferred to simulate a realistic scenario that induces a large variability in the delay. We use the ns-2 simulator [11] to study the scenario depicted in Figure 1. Each simulation is run for 2000 simulation seconds. The TCP source starts at time $t = 0$ and the ON-OFF source starts at time $t = 3$ secs. A Poisson ON-OFF source is attached to node n_1, and a Newreno TCP source is attached to node n_2. The lengths of ON and OFF periods are exponentially distributed and are set to the same average value. During ON periods, packets are sent according to a Poisson process.

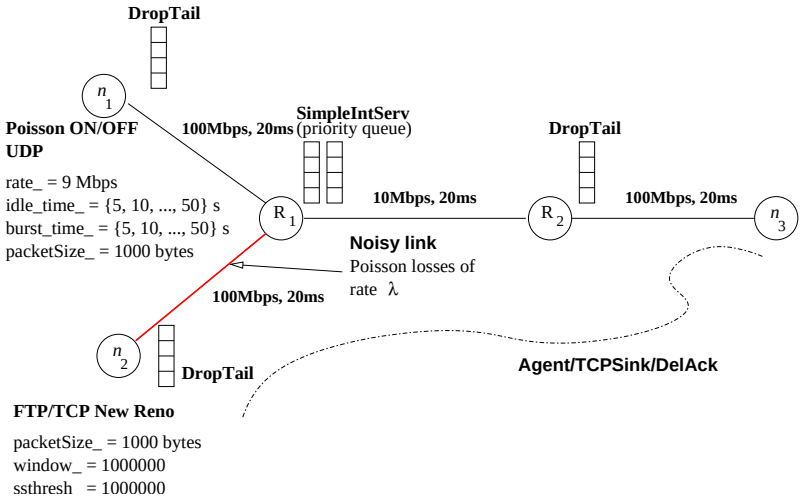

Figure 1: The simulated topology.

The link between nodes n_2 and R_1, which is called *lossy link*, drops packets according to a Poisson process. The packet loss rate in the lossy link is assumed to be fixed and equal to λ packet-losses/second. The SimpleIntServ queue at the input of node R_1 is a 2-level priority queue, where packets from the ON-OFF source get higher priority. The scenario is configured in a way such that there are no congestion losses in the network (this is done by setting a large buffer at the input to the bottleneck link). Our objective is to validate the model in presence of a loss process independent of the window size.

6.2 Results

The results we present in what follows consider only the case where $\lambda_n = \lambda$. Figure 3 compares the performance of the simulations for three analytic models that correspond to three different assumptions on the round-trip time (RTT) process:

- Only the **mean RTT** value is considered when computing TCP throughput. This means in particular that we ignore the variability and correlation of round-trip time samples. Since our simulations are run with the Newreno version of TCP, this corresponds to case $A1-A2.i$. So, throughput is computed using the following expression:

$$X = \frac{\beta}{1-\gamma} \times \frac{1}{\mathbb{E}[R_0]\,(1 - e^{-\lambda \mathbb{E}[R_0]})},$$

  with $\beta = 0.5$ (since the delayed ACK feature is used) and $\gamma = \frac{1}{2}$.

- The **whole RTT distribution** is considered when computing TCP's throughput, but the correlation is ignored. RTTs are thus considered to be i.i.d. Equation (9) is used for computing the throughput and the Laplace-Stieltjes transform in this equation is evaluated using all RTT measurements instead of just using the mean RTT.

- **The correlated case** or Markov case. The throughput is computed in the following way.

  – We first use a simple empirical approach to model RTT as being modulated by a two state (ON-OFF) Markov chain. This is done by associating large RTT values to an ON state and small values to an OFF state, and then computing the empirical transition rates between the states of the modulating chain. More precisely, we sort in ascending order the RTT values, then choose a sample subset to be the OFF state (the lowest RTT values) and the rest of the samples are considered the ON state. The ON state is chosen from the point where the ordered RTT process increases very fast (generally, the ON state is taken above the first 85% samples of the ordered RTT process). Figure 2 shows a typical RTT trace and the corresponding ordered RTT process (right). So, in this case, the RTT process is considered to be in the ON state starting from about the sample number 9600 on the ordered RTT process.

  – By analyzing the original RTT process we obtain the vector R_i, where $i \in \{\text{on, off}\}$ represents the state of the Markov chain underlying the RTT process.

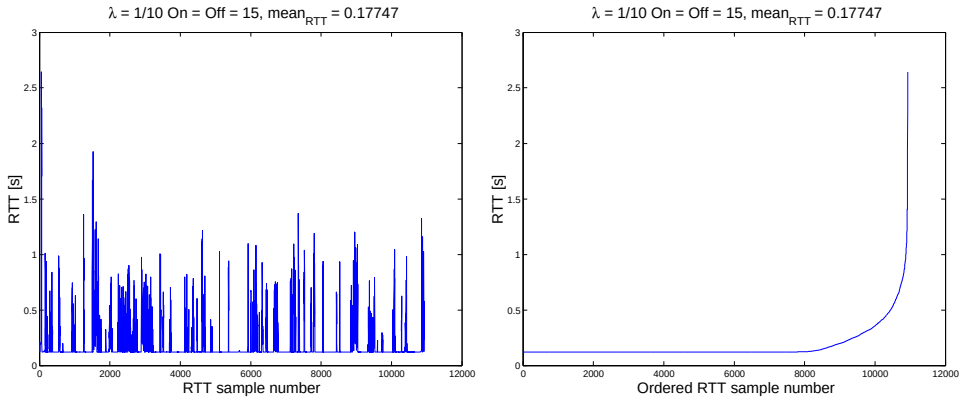

Figure 2: A typical RTT process and the ordered RTT process for deciding ON and OFF states.

– We get the transition probabilities of the Markov chain modulating the RTT process as follows:

$$p_{ij} = \frac{N_{ij}}{Ni} \quad \text{and} \quad p_{ii} = 1 - p_{ij}\,,$$

where $\{(i,j), i \neq j\} \in \{\text{on, off}\}$, N_{ij} is the number of transitions between on-off (resp. off-on) states, and N_i is the total number of RTT samples occurred during state i.

– The steady-state probabilities are computed as:

$$\pi_i = \frac{N_i}{N_{\text{RTT}}}\,,$$

where $i \in \{\text{on, off}\}$, N_i is the number of samples of the RTT process during state i, and N_{RTT} is the number of samples of the RTT process.

– Finally, we compute $\underline{w}$ as in (5), then we calculate the throughput in [bps] as:

$$X = \frac{8B\sum_i w_i}{E[R_0]}\,,$$

where $B = 1000$ bytes is the size of TCP packets and, as above, $i \in \{\text{on, off}\}$.

To compute each point in the following figures, the same simulation is run M times with different seeds, the throughput is computed for each run, and it is finally averaged over M. We consider a value of $M = 50$ in our simulations. As an example, we show in Table 1 the 95% confidence intervals for the Markov approach for $M = 50$. Since confidence intervals are small enough, $M = 50$ is well justified.

Table 1: Confidence intervals for the Markov approach when $\lambda = \frac{1}{10}$.

ON-OFF [s]	$\mathbb{E}[X_{\mathrm{ON-OFF}}]$	Confidence interval
5	2.3542×10^6	$\pm 0.0237 \times 10^6$
10	2.4471×10^6	$\pm 0.0201 \times 10^6$
15	2.4850×10^6	$\pm 0.0254 \times 10^6$
20	2.4318×10^6	$\pm 0.0257 \times 10^6$
25	2.6546×10^6	$\pm 0.0177 \times 10^6$
30	2.6344×10^6	$\pm 0.0185 \times 10^6$
35	2.3678×10^6	$\pm 0.0319 \times 10^6$
40	2.4649×10^6	$\pm 0.0270 \times 10^6$
45	2.4993×10^6	$\pm 0.0266 \times 10^6$
50	2.6316×10^6	$\pm 0.0125 \times 10^6$

Figure 3 shows the throughput results for $\lambda = \frac{1}{15}$ losses/s and $\lambda = \frac{1}{20}$ losses/s. Each point in the plot corresponds to the average throughput computed for the corresponding ON-OFF period lengths. The x-axis represents the average duration of ON and OFF periods. For all cases, the ON and OFF periods are set to the same average durations. The plot labelled as "Fixed" represents the case when the mean RTT value is only considered for throughput computation. The plot labelled as "Variable" represents the case when the whole RTT distribution is considered but not the correlation, and the plot labelled as "Markov" graphs the correlated case. For all cases, the Newreno version of TCP was considered.

The first thing we conclude from our simulation results is that considering delay variability (Variable and Markov cases) leads to a higher throughout than in the case when only the average delay is considered (the Fixed case). The second remark is that the real TCP throughput computed from measurements is closer to the Variable and Markov cases than the Fixed case. This indicates that our model is able to provide a better approximation of TCP throughput than existing models which only consider the mean delay. Note how, starting from ON-OFF period lengths of 20 seconds, the Markov approach is the nearest to the real measured throughput.

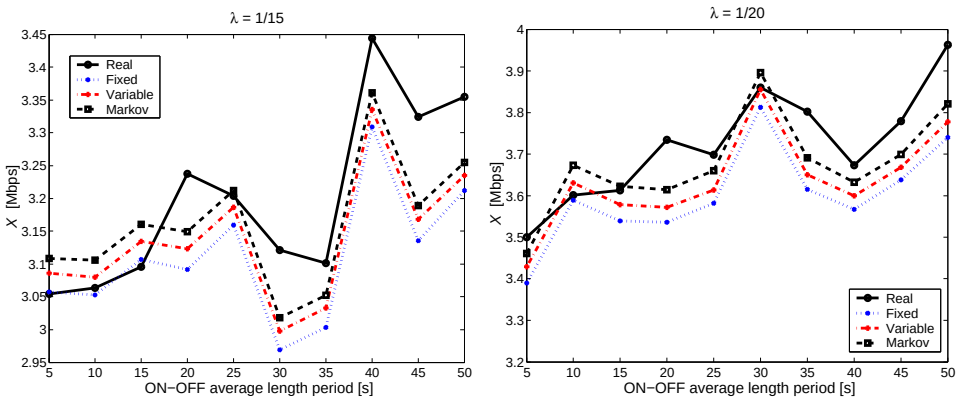

Figure 3: Throughput computed when considering delay variability.

7 Conclusions

We presented in this paper a model for an AIMD mechanism that accounts for delay variability. The model is based on stochastic difference equations. We solved this model for the moments of the window size in the stationary regime as well as for the throughput for the case of i.i.d. and Markov correlated round-trip times. We then studied the dependence of the throughput on delay variability. Our main analytical result was that the throughput of an AIMD mechanism increases when the delay becomes more variable.

In our analysis, we covered two AIMD versions. The first one divides its window in a lossy round-trip time by a constant which is independent of the number of losses. This mechanism models the new versions of TCP as Newreno and SACK. We also considered another AIMD mechanism that divides its window by a constant for each loss event. This latter mechanism models the Reno version of TCP.

We validated our analytical results with ns-2 simulations. We can summarize our conclusions from the numerical validation as follows:

- A model that only considers the mean RTT underestimates the total throughput of TCP.
- A model that accounts for the distribution of RTT but not the correlation (i.e. Equation (9)) is more accurate than the model that replaces RTT by its mean.
- The Markov model that takes into account also the correlation is the most accurate among the class of models we considered. We may expect further accuracy to be obtained by using a higher order Markov chain (with more than two states).

References

[1] Eitan Altman. Stochastic recursive equations with applications to queues with dependent vacations. *Annals of Operations Research*, 112(1):43–61, April 2002.

[2] Eitan Altman, Konstantin Avratchenkov, and Chadi Barakat. A stochastic model of TCP/IP with stationary random losses. In *Proceedings of the ACM Sigcomm*, Stockholm, Sweden, August 2000.

[3] V. Anantharam and T. Konstantopoulos. Stationary solutions of stochastic recursions describing discrete event systems. *Stochastic Processes and Their Applications*, 68:181–194, August 1997. Correction published in vol. 80, no. 2, pp. 271–278, April 1999.

[4] François Baccelli and P. Brémaud. *Elements of queueing theory: Palm-Martingale calculus and stochastic recurrences*. Springer-Verlag, 1994.

[5] A. A. Borovkov and S. G. Foss. Stochastically recursive sequences and their generalizations. *Siberian Advances in Mathematics*, 2(1):16–81, 1992.

[6] A. Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability*, 18, 1986.

[7] M. Choon Chan and R. Ramjee. TCP/IP performance over 3G wireless links with rate and delay variation. In *Proceedings of Mobicom*, pages 71–82, Atlanta, Georgia, USA, September 2002.

[8] K. Fall and Sally Floyd. Simulation-based comparisons of Tahoe, Reno, and SACK TCP. *ACM Computer Communication Review*, 26(3):5–21, July 1996.

[9] P. Glasserman and D. D. Yao. Stochastic vector difference equations with stationary coefficients. *Journal of Applied Probability*, 32:851–866, 1995.

[10] Van Jacobson and Michael J. Karels. Congestion avoidance and control. In *Proceedings of the ACM Sigcomm*, pages 314–329, 1988.

[11] University of California at Berkeley. Network simulator v.2. Available via http://www-nrg.ee.lbl.gov/ns-2.

[12] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: a simple model and its empirical validation. In *Proceedings of the ACM Sigcomm*, 1998.

[13] Qualcomm. Delays in the HDR system, June 2000.

[14] Milan Vojnovic and Jean-Yves Le Boudec. On the long-run behavior of equation-based rate control. In *Proceedings of the ACM Sigcomm*, August 2002.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

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