TCP over a multi-state Markovian path

Eitan Altman, Konstantin Avrachenkov*, Chadi Barakat**, and Parijat Dube***

INRIA Sophia Antipolis,
2004, route des Lucioles, B.P.93,
06902, Sophia Antipolis Cedex, France.
Email:{altman,k.avrachenkov,cbarakat,pdube}@sophia.inria.fr

Abstract. In this paper we analyze the performance of a TCP-like flow control mechanism. The transmission rate is considered to increase linearly in time until the receipt of a congestion notification (via loss detection in context of TCP) where the transmission rate is multiplicatively decreased. We introduce a general model based on a multi-state Markov chain for the moments at which the congestion is detected. With this model, we are able to account for correlation and burstiness in congestion moments. Furthermore, we specify several simple versions of our general model and then we identify their parameters from real TCP traces.

1 Introduction

We study in this paper the performance of an additive-increase multiplicative-decrease flow control protocol. This is the kind of control used by TCP, the widely-used transport protocol of the Internet [9]. TCP is used as a reference through the present work, however we anticipate that our results will be also applicable for other flow control mechanisms. A fluid approach is used to model the controlled flow. The transmission rate of the source is assumed to grow linearly at a rate α. In the case of TCP where the flow is controlled via a congestion window, the transmission rate at any instant is equal to the window size divided by the Round Trip Time of the connection. The growth of the transmission rate continues until the source receives a notification of congestion from the network. In case of TCP, the congestion is inferred from the loss of packets. It is an implicit notification compared to the explicit notification used by other flow control protocols as the ABR service in ATM or the ECN proposal in the Internet. We call the moment at which the source reduces its transmission rate a loss moment. Upon detection of a loss, the transmission rate is *scaled down* by a (possibly random) factor $a \in [0, 1]$. The scaling factor depends on many factors as the version of TCP, the number of packet losses in the congestion period and the way with which the loss is detected (e.g. duplicate ACK or Timeout [12]). Note that by choosing in some instants $a = 1$ one can introduce potential loss instants.

* The work of this author was financed by a grant of CNET France-Telecom on flow control in High Speed Networks

** The work of this author was financed by an RNRT "Constellations" project on satellite communications

*** The work of this author was partially financed by the French embassy in India

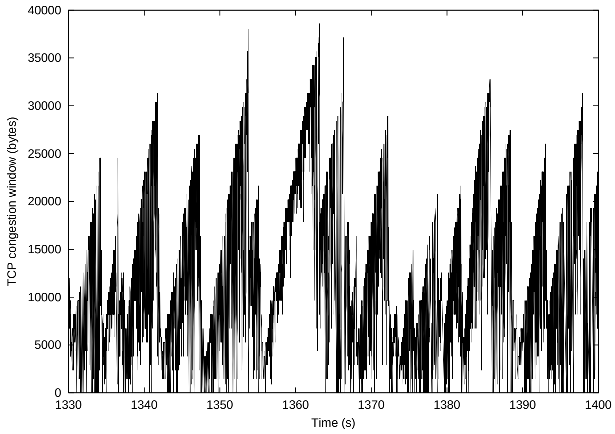

Fig. 1. TCP window evolution

The study of the performance of a flow control mechanism requires a characterization of the moments at which the transmission rate is reduced. These moments can be seen as a point process, where the appearance of a point corresponds to the appearance of a congestion signal or a loss in the context of TCP, causing a reduction in the transmission rate. Simple models as Poisson or iid models may not work in some cases where losses present some kind of burstiness or correlation. For example in Figure 1 one can observe a scenario where the moments of transmission rate reduction are clustered together. This figure corresponds to the window size evolution of a New Reno [5] TCP connection running between two sites at the technology park Sophia Antipolis. Normally in TCP, the window is divided by two upon congestion detection, but we see in this figure a more severe reduction due to multiple consecutive division of the congestion window by two. In a previous paper [1], we present a two-state Markovian model to account for burstiness of losses. In that paper, we considered a lossy path with two states *Good* and *Bad* together with potential loss moments. The transmission rate is reduced upon potential losses. A potential loss can transform into a real loss with probability p_G in the Good state and with probability p_B in the Bad state ($p_G \leq p_B$). The time between potential loss moments is assumed to be independently and identically distributed. Our main contribution in [1] is to show that the throughput of the flow control mechanism increases with the increase in burstiness of losses. However, we validated the model only via simulations, but we did not provide any algorithm for the identification of its parameters from real traces.

The present work is an extension of our previous work [1] to a multi-state Markovian case. Being motivated by some experimentation results (e.g. Figure 1), we allow the path of the connection to be in more than two states. The need for more than two states for describing the channel is also motivated by modelling results from [11,13] on mobile satellite channels, where it was shown that one needs typically at least four states. In [1], the scaling factor a is a random variable equal to either 0.5 (the potential loss becomes a real loss) or 1 (a potential loss is not transformed into a real loss). Here we propose to study the scaling factor with a general distribution that depends on the state of the path. We present then some applications of our general model. These applications can be seen as different ways to infer the parameters of the general model from a real TCP trace. In particular, we provide a method for the parameter identification of our model in [1]. A comparison between the different applications is provided to see which one is the most efficient in predicting TCP performance.

In the following section, we present the general multi-state multi-reduction model for the flow control mechanism. This general model is analyzed in Section 3. In Section 4, we provide several particular cases of the general model as well as their application to TCP modelling. We conclude Section 4 by a comparison between the different particular cases.

2 The model

Let $X(t)$ be the transmission rate at time t. In case of TCP, it is equal to the current window size divided by the Round Trip Time of the connection. Let $K = \{1, 2, \ldots, N\}$ be the set of possible states of the path. We allow losses to occur in any of the N states; the probability of the occurrence of losses in each of these states may be different. To that end, we define a series of potential losses occurring with a certain distribution of times between potential losses. Let T_n denote the time at which the nth potential loss occurs and let X_n denote the transmission rate just prior to T_n. The pair $\{T_n, X_n\}$ can be considered as a marked point process [3]. Let D_n, $n \in \mathbb{Z}$ be a sequence of times between potential losses: $D_n = T_{n+1} - T_n$. D_n are assumed to be i.i.d. with expectation d, second moment $d^{(2)}$ and Laplace Stieltjes Transform $D^*(s) = E[e^{-sD_n}]$. Let Y_n be the state of the channel at the nth potential loss instant. We assume further that the sequences $\{Y_n\}$ and $\{D_n\}$ are independent. We assume that $\{Y_n\}$ is an ergodic Markov chain with the following transition probabilities,

$$p_{ij} = P\{Y_{n+1} = j | Y_n = i\},\ 1 \leq i, j \leq N$$

Let $P = \{p_{ij}\}_{i,j=1}^{N}$ and let π be the stationary distribution of the Markov chain associated to the path. Next we define N random variables (discrete or continuous), $\{A_n^j; 1 \leq j \leq N\}$, which describe the behavior of the transmission rate when a potential loss occurs: is it reduced and if so by how much. These variables $\{A_n^j; 1 \leq j \leq N\}$ correspond to the N possible states of the model for losses. Each random variable $A_n^j, 1 \leq j \leq N$, takes values in the interval $[0, 1]$. It can take rational or real values within this interval. The choice of the interval $[0, 1]$ stems from the fact that we are *scaling down* the transmission rate at the instant of losses. The set includes 1 since it corresponds to the case when a potential loss is not transformed into a real loss and so the transmission rate is unaltered. $A_n^j, 1 \leq j \leq N$ has a distribution function $F^j(a)$ for all $n \in \mathbb{Z}$. That is, we take the distribution of A_n^j to be time homogeneous. Denote

$$a_i := \int_0^1 a dF^i(a), \quad 1 \leq i \leq N.$$

We assume that there is at least one i for which $a_i < 1$. The dynamics of the system can be given by the following stochastic recurrent equation

$$X_{n+1} = \sum_{j=1}^{N} A_n^j X_n 1\{Y_n = j\} + \alpha D_n. \tag{1}$$

3 Performance Analysis

First we observe that equation (1) is a particular case of stochastic linear difference equations of type $X_{n+1} = A_n X_n + B_n$, where $\{A_n, B_n\}$ is a stationary and ergodic processes (one can consider the Markov chain $\{Y_n\}$ in the stationary regime). It follows from [4] and [8] that such equations have a stationary solution X_n^* given by,

$$X_n^* = \sum_{k=0}^{\infty} (\prod_{i=n-k}^{n-1} A_i) B_{n-k-1}.$$

The stationary regime exists under the assumption that there is at least one i for which $a_i < 1$. Moreover, for any arbitrary starting point X_0, the sequence $\{X_n\}$ will converge almost surely to this stationary regime, that is

$$\lim_{n \to \infty} |X_n - X_n^*| = 0, \text{ P-a.s.}$$

Therefore, we can assume without loss of generality that the process $\{X_n\}$ is in the stationary regime in order to compute the limit distribution. Next we compute the moments of X_n in this regime. Let us denote,

$$x_i = E[X_n 1\{Y_n = i\}] \; 1 \le i \le N.$$

Obviously, the expectation of X_n is given by,

$$E[X_n] = \sum_{i=1}^{N} x_i.$$

To compute $x_i, 1 \le i \le N$, we use the Laplace Stieltjes Transform approach. Namely, define the following Laplace Stieltjes Transforms:

$$W(s,i) = E\left[e^{-sX_n} 1\{Y_n = i\}\right], \; 1 \le i \le N,$$

where we assume that X_n is in the stationary regime.

Theorem 1. *The Laplace Stieltjes Transforms $W(s,j), 1 \le j \le N$, are solutions of the following implicit equations,*

$$W(s,j) = D^*(\alpha s) \left[\sum_{i=1}^{N} p_{ij} \int_0^1 W(as,i) dF^i(a)\right] \; 1 \le j \le N \tag{2}$$

Proof: We write for any j, $1 \le j \le N$,

$$
\begin{aligned}
E[e^{-sX_{n+1}}1\{Y_{n+1}=j\}] &= \sum_{i=1}^{N} E[e^{-sX_{n+1}}1\{Y_{n+1}=j\}|Y_n=i\}P(Y_n=i) \\
&= \sum_{i=1}^{N} E[e^{-sX_{n+1}}|Y_n=i]E[1\{Y_{n+1}=j\}|Y_n=i]P(Y_n=i) \\
&= \sum_{i=1}^{N} E[e^{-s(A_n^i X_n+\alpha D_n)}|Y_n=i]p_{ij}P(Y_n=i) \\
&= D^*(\alpha s)\sum_{i=1}^{N}\int_0^1 E[e^{-saX_n}|Y_n=i]dF^i(a)p_{ij}P(Y_n=i) \\
&= D^*(\alpha s)\sum_{i=1}^{N} p_{ij}\int_0^1 E[e^{-saX_n}1\{Y_n=i\}]dF^i(a)
\end{aligned}
$$

This results in the implicit equations (2). $\square$

Although the Laplace Stieltjes Transforms in Theorem 1 are only given as solutions of implicit equations, all moments of $X_n 1\{Y_n = i\}$ for $1 \le i \le N$ (in the stationary regime) can be obtained explicitly. Note that

$$E[X_n^k 1\{Y_n = i\}] = (-1)^k \frac{d^k W(s,i)}{ds^k}\bigg|_{s=0}.$$

We shall now proceed to the calculation of expressions for the first and second moments of $X_n 1\{Y_n = i\}$ for $1 \le i \le N$ from the implicit expressions of the Laplace Stieltjes transforms. Upon differentiating the implicit expressions (2) and using the following relations,

$$W(0,i) = \pi_i, \quad 1 \le i \le N,$$

$$D^*(0) = 1, \quad \frac{dD^*(\alpha s)}{ds}|_{s=0} = -\alpha d,$$

we get N linear equations in N unknowns:

$$x_j = \sum_{i=1}^{N} p_{ij} a_i x_i + \alpha d \pi_j \;\; 1 \le j \le N. \tag{3}$$

We shall now write the above N equations in matrix notation. Let $x = [x_1, x_2, \ldots, x_N]$ and

$$A = \begin{bmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_N \end{bmatrix}.$$

Then the equations (3) take the form

$$x = xAP + \alpha d\pi \tag{4}$$

Recall that $0 \leq a_i \leq 1$ for all i. Furthermore, we assume that there is at least one i for which $a_i < 1$. The latter guarantees that the matrix AP is substochastic (there is an i for which $\sum_{j=1}^{N} p_{ij} a_i < \sum_{j=1} p_{ij} = 1$). Recall that moduli of all eigenvalues of a substochastic matrix are strictly less than one. Therefore, matrix $I - AP$ has no zero eigenvalue, and consequently, equation (4) has a unique solution. Thus we can state the following result:

Theorem 2. *Let X_n be in the stationary regime. Then $E[X_n]$ is given by*

$$E[X_n] = xe = \alpha d\pi (I - AP)^{-1} e$$

where e is a vector of ones.

To compute the second moment of X_n, we first define

$$x_i^{(2)} = E[X_n^2 1\{Y_n = i\}], \quad 1 \leq i \leq N.$$

Clearly,

$$E[X_n^2] = \sum_{i=1}^{N} x_i^{(2)}.$$

Also let $x^{(2)} = [x_1^{(2)}, x_2^{(2)}, \ldots, x_N^{(2)}]$ and

$$A^{(2)} = \begin{bmatrix} a_1^{(2)} & 0 & \ldots & 0 \\ 0 & a_2^{(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_N^{(2)} \end{bmatrix},$$

where

$$a_i^{(2)} = \int_0^1 a^2 dF^i(a), \quad 1 \leq i \leq N.$$

Then in the next Theorem we give an explicit expression for $E[X_n^2]$.

Theorem 3. *Let $\{X_n\}$ be in the stationary regime and there is at least one i for which $a_i < 1$. Then $E[X^2]$ is given by*

$$E[X_n^2] = x^{(2)} e = \left(2\alpha d(xAP) + \alpha^2 d^{(2)} \pi\right) (I - A^{(2)} P)^{-1} e.$$

Proof: Differentiating twice the implicit expressions (2), we obtain

$$\begin{aligned} \frac{d^2 W(s,j)}{ds^2} &= D^*(\alpha s) \left[\sum_{i=1}^{N} p_{ij} \int_0^1 \frac{d^2 W(as,i)}{ds^2} dF^i(a) \right] \\ &+ \frac{d^2 D^*(\alpha s)}{ds^2} \left[\sum_{i=1}^{N} p_{ij} \int_0^1 W(as,i) dF^i(a) \right] \\ &+ 2 \frac{dD^*(\alpha s)}{ds} \left[\sum_{i=1}^{N} p_{ij} \int_0^1 \frac{dW(as,i)}{ds} dF^i(a) \right] \end{aligned}$$

Now evaluating the above derivatives at $s = 0$, we get

$$x_j^{(2)} = \sum_{i=1}^{N} p_{ij} a_i^{(2)} x_i^{(2)} + 2\alpha d \sum_{i=1}^{N} p_{ij} a_i x_i + \alpha^2 d^{(2)} \pi_j .$$

Next we rewrite the equations in matrix notation

$$x^{(2)} = x^{(2)} A^{(2)} P + 2\alpha d (xAP) + \alpha^2 d^{(2)} \pi .$$

Solving for $x^{(2)}$, we get

$$x^{(2)} = \left(2\alpha d (xAP) + \alpha^2 d^{(2)} \pi \right) \left(I - A^{(2)} P \right)^{-1}$$

The existence of $(I - A^{(2)} P)^{-1}$ is guaranteed, because $A^{(2)} P$ is again substochastic as the sum of the elements of the ith row of $A^{(2)} P$ is $\sum_{j=1}^{N} p_{ij} a_j^{(2)} < \sum_{j=1}^{N} p_{ij} = 1$. □

Observe that we computed the expectation of the transmission rate with respect to loss instants. This expectation is also referred to as Palm expectation in the context of marked point processes [3]. Of course, the most interesting is the calculation of the expectation of the transmission rate at an arbitrary time moment. For ergodic processes the latter expectation coincides with the following time average P-a.s.,

$$\bar{x} = \lim_{T \to \infty} \frac{1}{T} \int_0^T X(t) dt$$

This is no other than the throughput of the transfer. It is the total volume of transmitted data over the transfer time. We proceed to evaluate this throughput by employing the concept of Palm probability.

Theorem 4. *The throughput, or the time-average transmission rate, is given by*

$$\bar{x} = E[X(t)] = \sum_{i=1}^{N} a_i x_i + \frac{1}{2} \alpha \frac{d^{(2)}}{d} = \bar{a} x^T + \frac{1}{2} \alpha \frac{d^{(2)}}{d}, \tag{5}$$

where $a = [a_1, a_2, \ldots, a_N]$ and x is given in Theorem 2.

Proof: To compute $E[X(t)]$ one can use the following inversion formula (see e.g., [3] Ch.1 Sec.4)

$$E[X(t)] = \frac{1}{d} E^0 [\int_0^{T_1} X(t) dt] \tag{6}$$

where $E^0[.]$ is an expectation associated with Palm distribution. Thus we can write,

$$E[X(t)] = \frac{1}{d} E^0 [\int_0^{T_1} (\sum_{i=1}^{N} A_0^i X_0 1\{Y_0 = i\} + \alpha t) dt]$$

Because of the independence of X_n and $\{D_k, k \geq n\}$ and also because of the independence of $\{D_n\}$ and $\{Y_n\}$ we can write,

$$
\begin{aligned}
E[X(t)] &= \frac{1}{d}[\sum_{i=1}^{N} \left(E^0[A_0^i] E^0[X_0 1\{Y_0 = i\}] \right) E^0[D_0]] + \frac{\alpha}{2d} E^0[D_0^2] \\
&= \sum_{i=1}^{N} a_i x_i + \frac{1}{2}\alpha \frac{d^{(2)}}{d} = a x^T + \frac{1}{2}\alpha \frac{d^{(2)}}{d}
\end{aligned}
$$

□

In the next theorem we evaluate the second moment of the transmission rate at an arbitrary time instant.

Theorem 5. *Let $d^{(3)}$ be the third moment of the time between potential losses. The second moment of the input rate over a long time interval is equal to:*

$$
\begin{aligned}
\bar{x}^{(2)} &= \lim_{t \to \infty} \frac{1}{t} \int_0^t X^2(t) dt \\
&= \frac{1}{3}\alpha^2 \frac{d^{(3)}}{d} + \frac{1}{d}\alpha d^{(2)} a x^T + a^{(2)} x^{(2)T}
\end{aligned}
$$

where $a^{(2)} = [a_1^{(2)}, a_2^{(2)}, \ldots, a_N^{(2)}]$ and $x^{(2)}$ is given in Theorem 3.

Proof: Again by the inversion formula from Palm probability,

$$
\begin{aligned}
E[X^2(t)] &= \frac{1}{d} E^0[\int_0^{T_1} X^2(t) dt] \\
&= \frac{1}{d} E^0 \left[\int_0^{T_1} \left(\sum_{i=1}^{N} A_0^i X_0 1\{Y_0 = i\} + \alpha t \right)^2 dt \right] \\
&= \frac{1}{d} E^0 \left[\frac{\alpha^2 D_0^3}{3} + \alpha D_0^2 \sum_{i=1}^{N} A_0^i X_0 1\{Y_0 = i\} + \sum_{i=1}^{N} \left(A_0^i\right)^2 X_0^2 1\{Y_0 = i\} D_0 \right] \\
&= \frac{1}{3}\alpha^2 \frac{d^{(3)}}{d} + \frac{1}{d}\alpha d^{(2)} \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N} a_i^{(2)} x_i^2 \\
&= \frac{1}{3}\alpha^2 \frac{d^{(3)}}{d} + \frac{1}{d}\alpha d^{(2)} a x^T + a^{(2)} x^{(2)T}
\end{aligned}
$$

□

Having obtained the expressions for the general case of N states we shall now focus on some particular cases in the following sections. We show how the parameters of our model can be inferred from a real trace of a TCP connection. Different possible applications of the model to the same trace are presented and the results are then compared to show which method is the most efficient. We will see in the sequel how much the model is general and how multiple sub-models can be derived from it by setting differently the parameters.

4 Specifications of the general model

In this section, we present different ways for the application of our general model to predict the performance of a TCP-like flow control mechanism. We chose to work with real loss processes. From the trace of a TCP connection, we determine the moments of window reduction. We reconstruct then the evolution of TCP congestion window over time under the assumption that the window increases linearly between two consecutive losses. We call this reconstructed window evolution the Exact Fluid Model and we use it below as a reference. We try then to derive simple closed form expressions for the throughput of the exact fluid model, and therefore for the throughput of TCP, using simple versions of our general model.

Our experimentation consists of a long-life New-Reno TCP connection running between `clope.inria.fr` at INRIA and `nessie.essi.fr` at ESSI, both located in the technology park Sophia Antipolis in France. The two machines are connected to the same metropolitan network. The TCP connection is run eleven times for approximately 20 minutes each at the most busy periods (between 10 am and 2 pm). The trace of the connection is captured at the source using the `tcpdump` tool and a program is developed to analyze the traces in order to find the moments at which the congestion window is divided by two. We noticed that most of the time, the loss of packets is detected with the Fast Retransmit algorithm (3 Duplicate ACKs) [12]. We noticed also that the maximum window advertised by the receiver is rarely reached due to working at busy periods. Thus, we can expect that our fluid model approximates correctly the behavior of the congestion window.

4.1 The basic model

We consider here the very simple case where the path has a single state and where the transmission rate is divided once by two at every potential loss moment. We assume that the times between losses are iid. This gives the following expression for the throughput,

$$E[X(t)] = \alpha d + \frac{1}{2}\alpha \frac{d^{(2)}}{d}. \tag{7}$$

Obviously, if times between losses are really iid, this model must give very close result to the throughput of the exact fluid model. And indeed, in our experiments we did not find a significant correlation between inter-loss times. Figure 2 confirms this conclusion. The throughput given by formula (7) follows closely the one given by the exact fluid model. However, to use formula (7) for the throughput calculation, one must know the second moment of inter-loss times. Usually, this quantity is difficult to find since it requires the knowledge of all inter-loss times for the modelled connection. Note that, by contrast, d can be easily calculated by dividing the total time of the connection by the number of losses. The number of losses in turn can be calculated using the packet loss probability. One way to eliminate $d^{(2)}$ is to express it as a function of d. For example, one can assume that inter-loss times form a Poisson process and hence take $d^{(2)} = 2d^2$. The problem with this solution is that it hides the impact of burstiness and expresses the throughput only as a function of the average loss rate. Indeed in Figure 2, the throughput calculated according to the Poisson assumption does not match well the throughput of the exact fluid model. The reason for this mismatch is clearly explained by Figure 3 where we plot the histogram of inter-loss times. This figure shows the deviation of the inter-loss time distribution from the exponential shape. This deviation is caused by the appearance of bursts of losses which causes the pulse of probability around the origin. Indeed, we noticed from the real traces of a TCP connection that the congestion window is divided multiple times by two when a congestion occurs and this due to the loss of packets in multiple consecutive Round Trips (see also Figure 1). However, the important notice we made from Figure 3 is that the time between bursts can still be well approximated by the exponential distribution. Figure 4 shows the distribution of times between losses after the elimination of the pulse around the origin. In the next two sections, we will present two methods to account for this bursty behavior of losses.

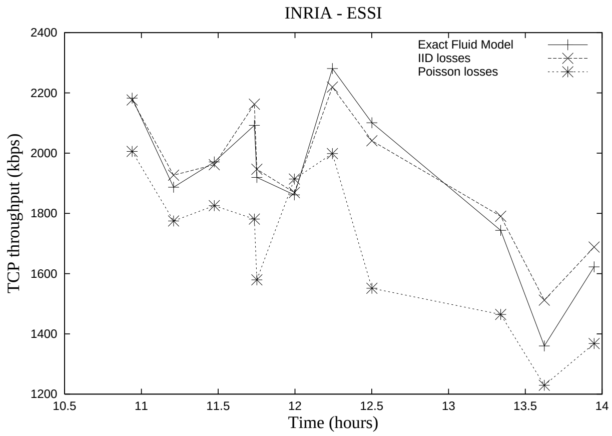

Fig. 2. Comparison of Poisson, iid and exact fluid models

4.2 The aggregate loss method

As was noticed in Figure 3, the inter-loss time distribution is a mixture of two distributions, one around the origin represents the time between losses within bursts and another away from the origin represents the time between bursts. This prompts us to aggregate the losses inside a burst into a single loss and to divide the transmission rate upon an aggregate loss occurrence (or a burst occurrence) by two power the number of aggregated losses inside the burst. The aggregate loss process can be considered now as a Poisson process. Upon the arrival of an aggregate loss, the transmission rate is divided by a random factor that can be greater than two. The question that one may ask here is how to characterize a burst, in other words how

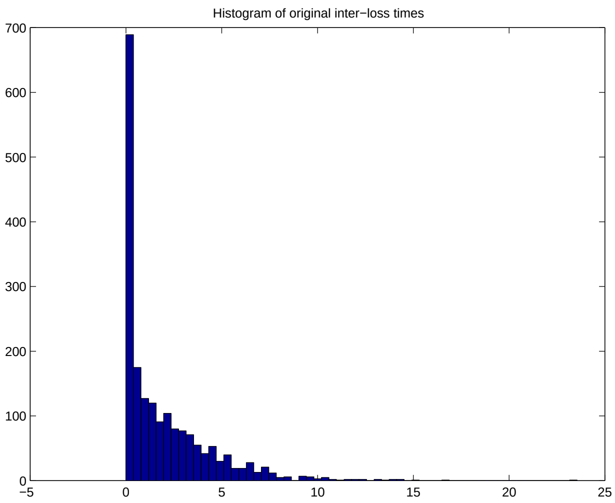

Fig. 3. Histogram of inter-loss times

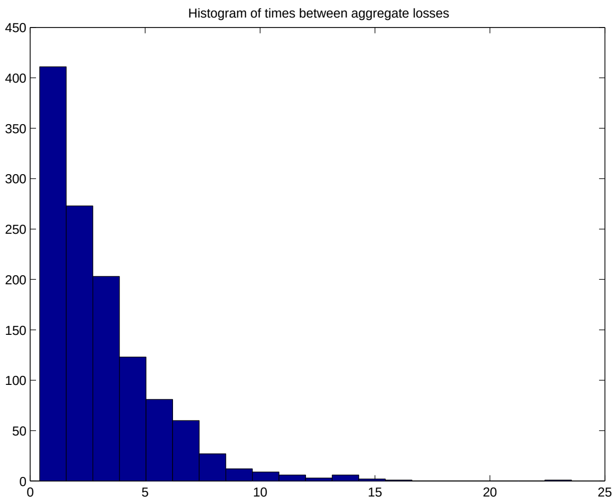

Fig. 4. Histogram of times between bursts

to decide that two consecutive losses are within the same burst or within two different bursts. In this section we use the following empirical method: we look at the distribution of inter-loss times and to try to find a point which clearly separates the two distributions. We zoom in Figure 5, the distribution of inter-loss times (Figure 3) around the origin. It is clear that two bursts are separated by approximately δ =0.4s. We use this δ for the identification of bursts. In the following, we present two different ways to describe the behavior of the random reduction factor. The first way is to assume that it is iid. The second way is to model it with a Markov chain.

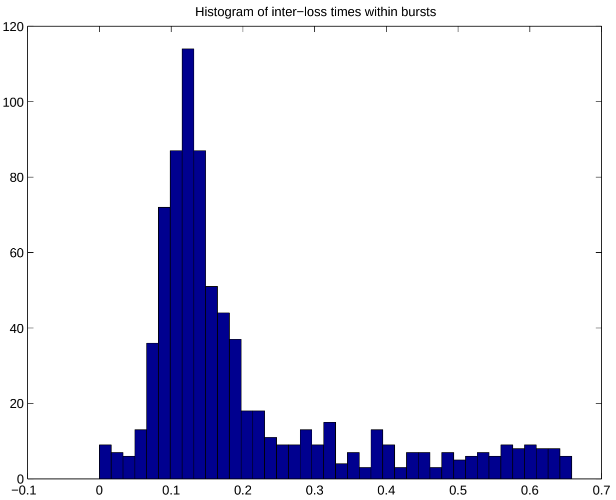

Fig. 5. Histogram of inter-loss times within bursts

First, let us consider the case of iid reduction factor. The evolution of the transmission rate in this case is given by

$$X_{n+1} = A_n X_n + \alpha D_n,$$

where the reduction factor A_n has a distribution function $F(a)$. D_n is the time between bursts which can be approximated by a Poisson process. Of course, this can be viewed as a particular case of our general model where the path of the connection has only one state. The general results of Section 2 can be specified for the present case as follows,

$$E[X_n] = \frac{\alpha d}{1-\bar{a}},$$

$$\bar{x} = E[X(t)] = \frac{\alpha d \bar{a}}{1-\bar{a}} + \frac{1}{2}\alpha \frac{d^{(2)}}{d}, \tag{8}$$

where $\bar{a} = \int_0^1 adF(a)$. Here, the reduction factor A_n is a discrete random variable which takes the values multiple of $1/2$. Thus, we calculated $\bar{a}$ as

$$\bar{a} = \sum_{i=1}^{m} \frac{1}{2^i} p_i,$$

where the probabilities p_i are estimated from the TCP connection trace. Let n be the total number of aggregate losses in the trace. We can write

$$p_i = \sum_{k=1}^{n} 1\{a_k = 1/2^i\}/n$$

Note here that the main gain from aggregation, is that the second moment of D_n can now be taken as $2d^2$. Furthermore, from Figure 4, one can see that the distribution of D_n is a shifted exponential distribution given that the time between two aggregate losses is always larger than δ. Thus, a more correct estimation for the second moment is given by

$$d^{(2)} = \delta^2 - 2\delta d + 2d^2.$$

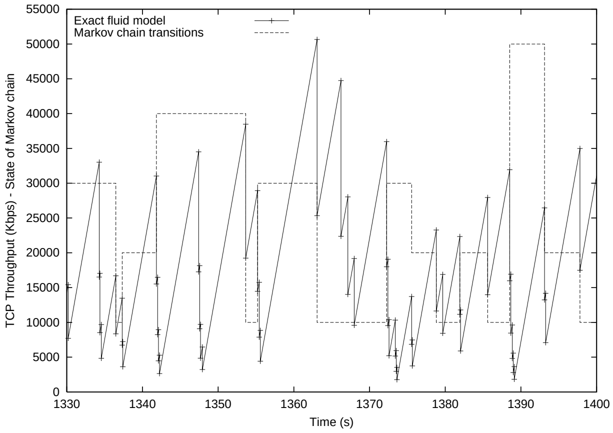

Fig. 6. Transitions of the multi-state Markov chain

Next we consider the case where the reduction factor is modelled using a Markov chain. We associate a multi-state Markov chain to the path. The transitions of the chain occur upon aggregate loss arrival. The state of the chain when an aggregate loss arrives is equal to the number of losses within the burst. The Markov chain determines then how many times the transmission rate is divided by two. Figure 6 explains how the transmission rate and the Markov chain change together. A interval of 0.4s is used to identify the losses belonging to the same burst. The evolution of the transmission rate in this case can be described as follows,

$$X_{n+1} = \sum_{j=1}^{N} a_j 1\{Y_n = j\} X_n + \alpha D_n, \tag{9}$$

where a_j is constant equal to $1/2^j$ and where Y_n is the state of the Markov chain. D_n again represents the time between bursts which can be approximated by a Poisson process. As a corollary of Theorem 3, the throughput can be written as

$$\bar{x} = E[X(t)] = \sum_{j=1}^{N} a_j x_j + \frac{\alpha d^{(2)}}{2d}. \tag{10}$$

The estimations of transition probabilities $\hat{p}_{ij}, i, j = 1, ..., N$, of the Markov chain $\{Y_k\}$ are identified from the trace of the TCP connection as follows,

$$\hat{p}_{ij} = \sum_{k=1}^{n-1} 1\{Y_{k+1} = j | Y_k = i\} / \sum_{k=1}^{n-1} 1\{Y_k = 1\}$$

where the Markov chain state Y_k corresponds to the number of transmission rate reduction at the event of the kth aggregate loss and n is the total number of aggregate loss events. If the number of rate reductions at the aggregate loss moment is greater than N, we assume that the Markov chain is in the state N. Since N is chosen so that it is unlikely to have the rate reduced more than N times during a burst, this assumption should not cause any problem. In the following we take $N = 4$.

Using the maximum distance of 0.4s between losses within a burst (Figure 5), we aggregate in bursts the moments at which the transmission is divided by two. As before, we assume that the resulting aggregate loss process is Poisson. We approximate the throughput of the exact fluid model using equations (8) and (10). Figure 7 shows the results. The iid batch model denotes the first case where the number of losses in a burst is described by an iid random variable. The Markovian batch model denotes the second case where this number is described by a Markov chain. We notice that the two methods give approximately the same result which means that the number of losses within a burst is really iid distributed. The result is closer to that of the exact fluid model than the throughput calculated for the Poisson model. However, it is not as good as we expected. The main reason is that we are ignoring the length of a burst which is here comparable to the time between bursts. Possibly, for other connections where losses are more clustered together, this batch method will have a better performance. One may expect that the Markov version of the batch model will perform better than the iid version on connections where strong correlation exists between burst sizes. In the next subsection, we will present a model that accounts for the time the connection spends during a burst.

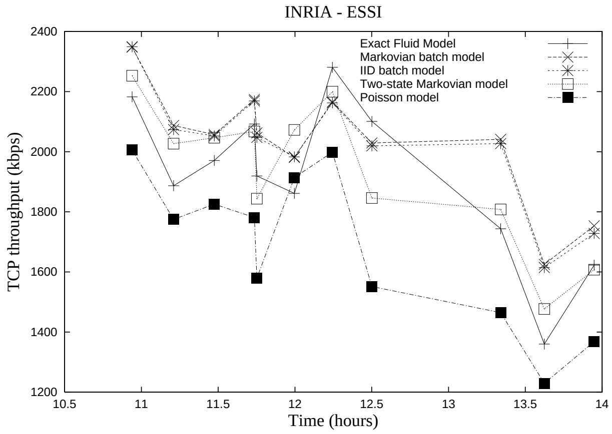

Fig. 7. Comparison between the different methods

4.3 The two-state model

Consider a particular case of our general model where the path switches between two different states. Namely, let $N = 2$ and let the state 1 corresponds to the *Good* state of the path and the state 2 to the *Bad* state. We also denote the transition probabilities of the Markov chain as follows: $p_{11} = g$, $p_{12} = \bar{g} = 1-g$, $p_{21} = \bar{b} = 1-b$ and $p_{22} = b$. The stationary distribution of this chain are equal to,

$$\pi_1 = \frac{\bar{b}}{\bar{b}+\bar{g}}, \qquad \pi_2 = \frac{\bar{g}}{\bar{b}+\bar{g}}$$

The following results can be easily obtained as straightforward corollaries of the theorems for the general N state model.

Corollary 1 *The Laplace Stieltjes Transforms $W(s,i), i = 1,2$, are the solutions of the following implicit equations,*

$$\begin{aligned} W(s,1) &= D^*(\alpha s)[g \int_0^1 W(as,1)dF^1(a)] + D^*(\alpha s)[\bar{b} \int_0^1 W(as,2)dF^2(a)], \\ W(s,2) &= D^*(\alpha s)[\bar{g} \int_0^1 W(as,1)dF^1(a)] + D^*(\alpha s)[b \int_0^1 W(as,2)dF^2(a)]. \end{aligned}$$

We shall now proceed to obtain explicit expressions for the first and second moments of the transmission rate at potential loss instants.

Corollary 2 *The first moment of the transmission rate at a potential loss moment is given by*

$$E[X_n] = x_1 + x_2,$$

where

$$x_1 = \alpha d \frac{a_2(\pi_2 - b) + \pi_1}{1 - a_2 b - a_1 g + a_1 a_2 (g + b - 1)} \tag{11}$$

$$x_2 = \alpha d \frac{a_1(\pi_1 - g) + \pi_2}{1 - a_2 b - a_1 g + a_1 a_2 (g + b - 1)} \tag{12}$$

Corollary 3 *The second moment of the transmission rate at a potential loss moment is given by*

$$E[X_n] = x_1^{(2)} + x_1^{(2)},$$

where

$$x_1^{(2)} = \frac{2\alpha d a_1 a_2^{(2)} x_1 (1 - g - b) + 2\alpha d (a_2 x_2 \bar{g} + a_1 x_1 g) + \alpha^2 d^{(2)} (a_2^{(2)} \pi_2 + \pi_1 - b a_2^{(2)})}{\left(1 - g a_1^{(2)} - b a_2^{(2)} - a_1^{(2)} a_2^{(2)} (1 - g - b)\right)} \tag{13}$$

$$x_2^{(2)} = \frac{2\alpha d a_1^{(2)} a_2 x_2 (1 - g - b) + 2\alpha d (a_1 x_1 \bar{g} + a_2 x_2 b) + \alpha^2 d^{(2)} (a_1^{(2)} \pi_1 + \pi_2 - g a_1^{(2)})}{\left(1 - g a_1^{(2)} - b a_2^{(2)} - a_1^{(2)} a_2^{(2)} (1 - g - b)\right)} \tag{14}$$

Corollary 4 *The throughput, or the time-average of the transmission rate, is given by*

$$E[X(t)] = a_1 x_1 + a_2 x_2 + \frac{1}{2}\alpha \frac{d^{(2)}}{d},$$

where x_1 and x_2 are given in Equations (11) and (12).

Corollary 5 *The second moment of the transmission rate at an arbitrary time instant is given by*

$$E[X^2(t)] = a_1^{(2)} x_1^{(2)} + a_2^{(2)} x_2^{(2)} + \frac{\alpha d^{(2)} (a_1 x_1 + a_2 x_2)}{d} + \frac{1}{3}\alpha^2 \frac{d^{(3)}}{d},$$

where x_1 and x_2 are given in Equations (11) and (12) and $x_1^{(2)}$ and $x_2^{(2)}$ in Equations (13) and (14) respectively.

Next we specialize the model further by taking A_n^j, for $j \in \{1, 2\}$ and $\forall n \geq 0$, to be discrete random variables with values in $\{0.5, 1\}$. Note that $A_n^j = 0.5$ represents the case when a potential loss is transformed into a real loss, namely when it causes a reduction in the transmission rate, whereas $A_n^j = 1$ represents the case when the transmission rate is not reduced at the potential loss moment. We get here the same model as that described in [1]. Note that in [1] we validate via simulation a particular case of this two-state model that corresponds to $p_G = 0$, $p_B = 1$. In the present work, we show how to set the different parameters of the two-state model in its general case. $\{D_n\}$ is the sequence of the times between potential losses. We also denote $p_G := P\{A_n^1 = 0.5\} = 1 - P\{A_n^1 = 1\}$, as the probability of the event when a potential loss is transformed into a real loss in the Good state. Analogously, we define the probability of a potential loss becoming a real loss in the Bad state as $p_B := P\{A_n^2 = 0.5\} = 1 - P\{A_n^2 = 1\}$. We assume that $p_G \le p_B$. Clearly,

$$a_1 = 1 - \frac{1}{2}p_G \quad \text{and} \quad a_2 = 1 - \frac{1}{2}p_B.$$

Next we demonstrate how the introduced above parameters as well as d and the transition matrix P can be determined from the data in real TCP traces. First, we obtain an estimation of the transition matrix for the Markov chain $\{Y_n\}$. Recall that this is the Markov chain obtained when looking at the state of the channel at potential loss moments. Let $\{S_n\}$ be a sequence of inter-loss times measured from a TCP trace. We need to determine when the path is in the "Good" state and when it is in the "Bad" state. We use the following simple method. Choose some time interval τ. We will explain later how to make this choice. If the inter-loss time S_n is less than τ then the path is in the Bad state, otherwise the path is considered to be in the Good state. If two or more inter-loss times correspond to the same state, we will merge these intervals together and call the new interval L_k^G or L_k^B depending on the state. Note that these new intervals represent the time during which the path of the connection is either in the Good or in the Bad state. Denote n_G (resp. n_B) the number of the time intervals S_k^G (resp. S_k^B) during the time interval that we use for measurement. Then, the evolution of the path of the TCP connection can be described by a two-state continuous time Markov process with the following infinitesimal generator matrix,

$$Q = \begin{bmatrix} -\sigma_G & \sigma_G \\ \sigma_B & -\sigma_B \end{bmatrix} \tag{15}$$

where the rates σ_G and σ_B are calculated as follows:

$$\sigma_G = \frac{1}{E[S_k^G]} \simeq \frac{n_G}{\sum_{k=1}^{n_G} S_k^G}, \quad \sigma_B = \frac{1}{E[S_k^B]} \simeq \frac{n_B}{\sum_{k=1}^{n_B} S_k^B}.$$

Note that on some paths, say a wireless link, this Markov chain is a priori known and can be directly used without the need to look at the trace of the TCP connection. In case it is not known, we need to define it using the parameter τ as described above. We present now two approaches for the determination of τ. The first one is more empirical. We look at the histogram of the inter-loss times (Figure 3) and we choose τ as the time separating the two distributions it encloses (0.4s in the figure). The second method is less empirical and was used in the context of Markov-modulated Poisson processes [10]. In this second approach we define parameter τ as the expectation of the inter-loss times, that is

$$\tau = E[S_k] \simeq \frac{1}{n}\sum_{k=1}^{n} S_k,$$

where n is the total number of inter-loss intervals we get from the trace. Given the continuous time Markov chain associated to the channel, we can now extract the parameters of the discrete time Markov chain embedded at the potential loss moments. We use for this purpose the uniformization technique [14]. Let us choose the potential loss process $\{D_n\}$ as a Poisson process with intensity $1/d$ higher than both σ_G and σ_B. For example, a reasonable choice of d is the estimation of the average Round Trip Time of the connection. According to the uniformization technique [14], the state of the path described by the Markov process (15) and sampled at the moments of potential losses can be equivalently given by a discrete time Markov chain with the following transition matrix,

$$P = \begin{bmatrix} 1 - d\sigma_G & d\sigma_G \\ d\sigma_B & 1 - d\sigma_B \end{bmatrix}.$$

Having chosen d and calculated σ_G and σ_B from the trace, we can easily deduce the parameters b and g of the loss model. Namely, $\bar{g} = d\sigma_G$ and $\bar{b} = d\sigma_B$. Now we determine p_G and p_B. Let ω_k^G (ω_k^B) be the number of real losses in the time interval S_k^G (resp. in S_k^B). Then the probabilities p_G and p_B are given by

$$p_G = \frac{\sum_{k=1}^{n_G} \omega_k^G}{\sum_{k=1}^{n_G} S_k^G / d} = \frac{d \sum_{k=1}^{n_G} \omega_k^G}{\sum_{k=1}^{n_G} S_k^G} = d\lambda_G, \quad p_B = \frac{\sum_{k=1}^{n_B} \omega_k^B}{\sum_{k=1}^{n_B} S_k^B / d} = \frac{d \sum_{k=1}^{n_B} \omega_k^B}{\sum_{k=1}^{n_B} S_k^B} = d\lambda_B.$$

$1/\lambda_G$ and $1/\lambda_B$ represent the average time between window reductions in the Good and in the Bad state respectively. For the same eleven traces obtained in our experiments, we calculated the parameters of the model. We use $\tau = \delta =$0.4s to separate the Bad state from the Good state. In Figure 7, we compare the result with that of the exact fluid model. A close match is noticed. In addition to the good results and the closed form expression it provides, this model has the advantage of having simple parameters. All what we need to approximate the throughput is the parameters of the two-state Markov chain associated to the path and the intensity of losses in both states. Concerning the parameter d, it is enough to choose in a way that the intensity of potential losses $1/d$ is higher than the intensity of losses in the Bad state λ_B.

5 Concluding Remarks

We considered in this paper a multi-state Markov path for describing the loss process experienced by a connection that has a linear window increase between losses, and multiplicative decrease upon a loss event. The modelling of some channels using a Markov chain with more than two states have long been advocated, see e.g. [11,13].

Using an approach based on the Laplace Stieltjes Transform, we derived explicit expressions for the two first moments of the transmission rate of the connection just prior to losses, as well as the two first moments of the steady state throughput. We note that the expression for the second moment of the throughput could be useful in designing TCP friendly protocols for real time applications [6] in which other parameters of the linear increase and multiplicative decrease are chosen so as to maintain the same expected throughput (as a function of the loss process and of the round-trip time) as the original TCP protocol. (The latter requirement on the expected throughput stmes from fairness arguments.) Such applications (e.g. interactive voice or video connections) typically require a smaller variance of the throughput than the one of the original TCP in order to ensure a reasonable quality of service.

We have recently succeeded also in analysing non Markovian channels [2], and obtain similar performance measures using a completely different approach (that relies on some covariance functions of the interloss times). The approach obtained here, in contrast, leads to formulae that involve only a finite and small number of parameters that can be easily computed. In addition, we proposed here methods for the identification of such parameters.

References

1. E. Altman, K.E. Avrachenkov, and C. Barakat, "TCP in presence of bursty losses", to appear in *Performance Evaluations.* A shorter version appeared in the *Proceedings of ACM SIGMETRICS*, Santa Clara, California, Jun 2000.
2. E. Altman, K.E. Avrachenkov, and C. Barakat, "A stochastic model of TCP-IP with stationary ergodic random losses", ACM SIGCOMM, Aug. 28 - Sept. 1, Stockholm, Sweden, 2000.
3. F. Baccelli and P. Bremaud, "Elements of queueing theory: Palm-Martingale calculus and stochastic recurrences", *Springer-Verlag*, 1994.
4. A. Brandt, "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients", *Adv. Appl. Prob.*, Vol. 18, pp. 211-220, 1986.
5. K. Fall and S. Floyd, "Simulation-based Comparisons of Tahoe, Reno, and SACK TCP", *ACM Computer Communication Review*, Jul 1996.
6. S. Floyd, M. Handley and J. Padhye, "Equation-based congestion control for unicast applications: the extended version", *ACM Sigcomm*, Aug. 28 - Sept. 1, Stockholm, Sweden, 2000.
7. E.N.Gilbert, "Capacity of a burst-noise channel", Bell Systems Technical Journal, Vol. 39, pp. 1253-1265, Sep 1960.
8. P. Glasserman and D.D. Yao, "Stochastic vector difference equations with stationary coefficients", *J. Appl. Prob.*, Vol. 32, pp. 851-866, 1995.
9. V. Jacobson, "Congestion avoidance and control", *ACM SIGCOMM*, Aug 1988.
10. K.S. Meier-Hellstern, "A fitting algorithm for Markov-modulated Poisson processes having two arrival rates", *Euro. J. Oper. Res.*, Vol. 29, pp. 370-377, 1987.
11. M. Rahman, M. Bulmer and M. Wilkinson, "Error models for land mobile satellite channels", *Australian Telecommunication Research*, Vol. 25 No 2, pp. 61-68, 1991.
12. W. Stevens, "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", *RFC 2001*, Jan 1997.
13. B. Vucetic and J. Du, "Channel modeling and simulation in satellite mobile communication systems", *IEEE J. on Selected Areas in Communnications*, Vol. 10, No. 8, pp. 1209-1218, 1992.
14. J. Walrand, "An introduction to queueing networks", *Prentice Hall*, 1988.