A model for TCP congestion control capturing the correlations in times between the congestion events

Esa Hyytiä and Peder J. Emstad
Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence,†
Norwegian University of Science and Technology, Trondheim, Norway

***Abstract*— We consider a simplified model for the rate control of TCP sources. In particular, we assume idealized negative feedbacks upon reaching a certain total sending rate, i.e., at the moment when the total sending rate attains a given capacity limit c one of the TCP sources is given a negative feedback and the source reduces its sending rate in a multiplicative manner. Thus, the model takes into account the interactions between different flows appropriately at the microscopic level instead of assuming independence. For this model we are able to derive steady state equations and solve them. Furthermore, we are able to compute several important performance measures such as the mean and the variance of the total sending rate. Moreover, we are able to characterize the packet loss process at the bottleneck link and, in particular, the correlations therein.**

I. Introduction

Transmission control protocol (TCP) is without doubt the most popular transmission rate control mechanism due to its success in the Internet. The rate control mechanism in TCP belongs to the family of the so-called additive increase and multiplicative decrease (AIMD) schemes, which have been studied extensively in the literature, see, e.g., [1] and [2]. An introduction to TCP rate control mechanism can also be found from several text books, see, e.g., [3]. One important aspect of bandwidth allocation protocols is the fairness, i.e., how the available bandwidth is allocated between the traffic flows sharing the same resources. For example, it is well-known that in TCP/IP networks the flows with higher round trip times (RTT) tend to get smaller shares. Perhaps the most famous results on TCP are the results on the throughput analysis, e.g., the famous square root formula for the TCP throughput by Floyd and Fall [4], and the more accurate expression by Padhye et al. [5]. The above results provide the average throughput per flow as a function of the packet loss probability.

The common assumption in TCP throughput analysis has been the independence between congestion epochs, i.e., it is assumed that the packet losses occur in fixed time intervals ([4], [5], [6], [7]), or originate from a (non-uniform) Poisson process [8]. A more general approach can be found from [9], where the losses are generated by an arbitrary exogenous random process allowing one to model correlations between inter-loss times.

In contrast to above work, in this paper we consider an elementary model for the TCP rate control mechanism, where the loss process gets explicitly defined by the sending rates of all TCP sources, as the case also is in practice. The aim is to characterize the behavior of the concurrent TCP flows at the microscopic level, e.g., in order to study the influence of the TCP traffic to other traffic flows such as real time voice or video streams. To this end, we study a single bottleneck link and make several simplifying assumptions about the rate control mechanism. Firstly, we consider a continuous model, i.e., a fluid flow model, where the sending rates are some non-negative real numbers. Secondly, the decision to send a negative feedback is based on the current arrival rate of the packets into the bottleneck link, not directly on the occupancy level of the buffer (unfinished work). Thirdly, we assume a constant delay before sources react on the negative feedback signals (i.e., a constant RTT).

The model is analyzed in the standard framework of Markov processes. First we present the steady state equations for the underlying embedded Markov chain with continuous state space. We are able to solve the special case of two TCP flows exactly, while for more than two flows the number of possible transitions increases considerably and we resort to a flow aggregation approach. In particular, we choose one flow as a targeted flow and assume that the rest of the flows share the remaining bandwidth equally. For this model we are able to compute several important performance measures such as the mean and the variance of the total sending rate for different number of flows, the distribution for the window size upon a negative feedback and full

†“Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence” is appointed by The Research Council of Norway, and funded by the Research Council, NTNU and UNINETT. (http://www.ntnu.no/Q2S/)

characterization of times between congestion events. The knowledge of the packet loss process at the bottleneck link, and the correlations therein, allows us to further present an elementary model for the loss process for the UDP flow(s) sharing the same link.

The rest of the paper is organized as follows. First, in Section II, we present our model in detail and comment about the assumptions. Then, in Section III, we analyze the model and derive expressions for several important performance measures and characterize the loss process caused by the TCP mechanism. In Section IV, we study the correlations between the consecutive packet losses, i.e., how often the buffer occupancy achieves the given level and how the time intervals between such events are correlated. Section V contains some numerical examples, and finally, in Section VI, we present our conclusions and comment on the future work.

II. TCP MODEL DESCRIPTION

We consider n TCP flows in an idealized system where the sending rate of one source is always reduced in multiplicative manner when the total rate would exceed the capacity of the (single) bottleneck link. In particular, let $r_i(t)$ denote the sending rate of source i at time t, so that the total sending rate is simply

$$R(t) = \sum_i r_i(t).$$

We can assume such a scaling that the total capacity of the bottleneck link is one, $c = 1$, so that we have

$$R(t) = \sum_i r_i(t) \leq 1, \qquad \forall \;\; t.$$

Furthermore, we assume a proportional marking, i.e., upon reaching the capacity limit c the flow to be downsized is chosen randomly with the probabilities proportional to the sending rates r_i. With $c = 1$ the probability of choosing flow i is simply $r_i(t)$. Let ν be the multiplicative factor, i.e., upon a negative feedback flow i reduces its rate according to

$$r_i(t + dt) = \nu \cdot r_i(t).$$

Thus, the drop in total sending rate is $(1-\nu) \cdot r_i(t)$ with probability of r_i and, consequently, the mean drop in total rate (conditional to the current state $\mathbf{r}$) is given by

$$\mathrm{E}\left[\Delta_R \,|\, \mathbf{r}\right] = (1-\nu) \cdot \sum_i r_i^2.$$

Moreover, the mean time to achieve the same total rate after a negative feedback is given by

$$\mathrm{E}\left[\Delta_T \,|\, \mathbf{r}\right] = (1-\nu) \cdot \sum_i r_i^2,$$

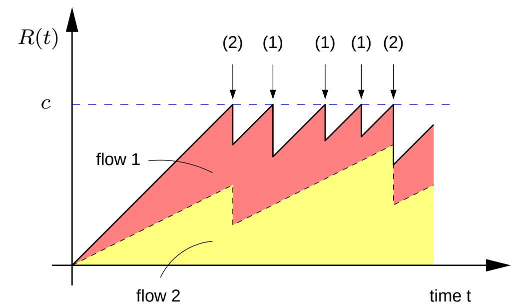


Fig. 1. Sample realization of two competing flows.

with an appropriate scale of time, i.e., when the linear increase factor is equal to 1, $d/dt\,R(t) = 1$. With this choice of the time scale we can write

$$\Delta_R = \Delta_T = \Delta.$$

A. Relationship to real systems

As mentioned already, this is a rather idealized model for TCP congestion control. Firstly, we consider a continuous model where sending rate of each source is some positive real number. This is rather standard assumption in the analysis of the TCP. Secondly, the decision to send a negative feedback signal (either by explicit congestion notification (ECN) [10] or by dropping a packet) is not based on the (averaged) occupancy level of the buffer, but the total sending rate of the sources sharing the same bottleneck link. Thirdly, we neglect the delays in the feedback loop, or rather assume that the delay is some constant for all sources, which yields the same additive increase rate for all sources.

Next we comment a bit on the second assumption as it is unique to this model. In practice the IP routers base their decisions on the (averaged) occupancy level of the buffer (see, e.g., random early detection in [11]). However, if the arrival rate is substantially lower than the service rate c, no queue accumulates in the router. Thus, if during each epoch the drop in the total sending rate is large enough for the router to empty the buffer, the assumption of the negative feedback criteria based on the current arriving rate is not really limiting. This should be the case with a small number of TCP flows (i.e., access network).

III. ANALYSIS

A. Embedded Markov chain

The process described in the previous section clearly constitutes a Markov process. In particular, we can associate an embedded Markov chain to this process

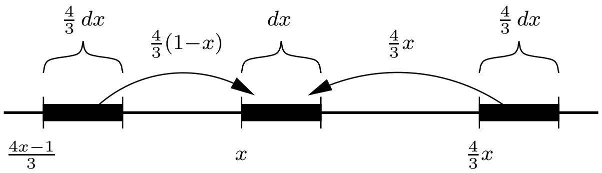


Fig. 2. Transitions to the interval $(x, x + dx)$.

by considering, e.g., the time instances when the total rate attains the capacity of the bottleneck link. Let $\mathbf{X}^{(k)}$ denote the sending rates of the TCP sources at the k^{th} point, $k = 1, 2, \ldots$. Then, with the probability of $X_i^{(k)}$ the next state $\mathbf{X}^{(k+1)}$ of the embedded Markov chain is

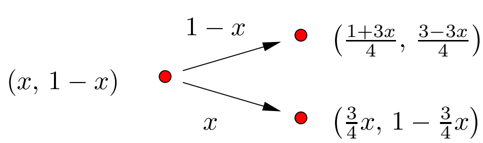
$$(\,X_1^{(k)} + \Delta^*,\, \ldots,\, \nu X_i^{(k)} + \Delta^*,\, \ldots,\, X_n^{(k)} + \Delta^*),$$

where $\Delta^* = (1 - \nu)X_i^{(k)}/n$. We note that the above Markov chain has a state space in $\mathbb{R}^n$ dimensional hyperplane, $\sum_{i=1}^n X_i = 1$ with $X_i \in (0, 1)\ \forall\, i$.

Note that, if the multiplicative factor ν and the initial state $\mathbf{X}^{(0)}$ of the system are fractional numbers, $\nu \in \mathbb{Q}$ and $\mathbf{X}^{(0)} \in \mathbb{Q}^n$, then the system remains in $\mathbb{Q}^n$, i.e., $\mathbf{X}^{(k)} \in \mathbb{Q}^n\ \forall\ k = 0, 1, \ldots$. However, in this paper we consider only "smooth solutions", which are relevant from the practical point of view.

B. Solution for 2 flows

With two flows and $X_1 = x$ the state of the embedded Markov chain is $(x,\, 1 - x)$. The transitions to the next embedded point (with $\nu = 1/2$) are then as follows.



where, e.g., at state $(x, 1 - x)$ the reduction occurs for flow 1 with probability of x and the embedded Markov chain moves to state $(3x/4, 1 - 3x/4)$. Let $f(x)$ denote the pdf for the state of the flow 1 at the embedded points, $\mathrm{P}\{x < X_1 \le x + dx\} = f(x)\,dx$. We define explicitly,

$$f(x) = 0 \quad \forall\ x < 0 \text{ and } x > 1.$$

Due to the symmetry we also have

$$f(x) = f(1 - x). \tag{1}$$

In order to determine the exact form of $f(x)$ we write the global balance equations. In particular, we can consider a small interval $(x,\, x + dx)$ and the probability flows to and from it. For a small interval dx there are no self-transitions in this system and the flow out is simply

$$\text{flow out} = f(x) \cdot dx.$$

For the flow in we need to consider transitions from two intervals corresponding to transitions up and down as illustrated in Fig. 2. Hence, the probability flow into small interval $(x, x + dx)$ is

$$f\left(\frac{4x-1}{3}\right) \cdot \frac{4}{3}\,dx \cdot \frac{4}{3}\,(1-x) + f\left(\frac{4x}{3}\right) \cdot \frac{4}{3}\,dx \cdot \frac{4}{3}\,x.$$

These probability flows are equal in the steady state, which yields

$$f(x) = \left(\frac{4}{3}\right)^2 \left[(1-x) f\left(\frac{4x-1}{3}\right) + x f\left(\frac{4x}{3}\right)\right]. \tag{2}$$

Letting $f(1/2) = a$ we obtain

$$\begin{aligned} f(1/2) = a &= \left(\frac{4}{3}\right)^2 \left[(1/2)\, f\left(\frac{1}{3}\right) + (1/2)\, f\left(\frac{2}{3}\right)\right] \\ &= \left(\frac{4}{3}\right)^2 \cdot f\left(\frac{1}{3}\right), \end{aligned}$$

and we get

$$f(1/3) = f(2/3) = \left(\frac{3}{4}\right)^2 \cdot a.$$

Similarly, for each $x = \left(\frac{3}{4}\right)^n \cdot \frac{1}{4}$ we can recursively obtain an accurate value. Thus,

$$f(1/2) = a \qquad f(1/4) = \frac{a}{4} \qquad f(9/64) = \frac{a}{48}$$

$$f(1/3) = \left(\frac{3}{4}\right)^2 \cdot a \qquad f(3/16) = \frac{a}{12}$$

Recursively, one obtains

$$\begin{aligned} f\left(\frac{1}{4}\left(\frac{3}{4}\right)^n\right) &= \frac{a}{4 \cdot 3^n} \cdot \left(\frac{3}{4}\right)^{n(n-1)/2} \\ &= a \cdot \sqrt{\frac{3^{n^2-3n}}{4^{n^2-n+2}}}, \quad \forall\ n = 0, 1, 2, \ldots \end{aligned}$$

The above suggests a solution for interval $x \in (0, 1/4]$,

$$f(x) = \frac{a}{4} \cdot \sqrt{\frac{3^{n(n-3)}}{4^{n(n-1)}}} = \frac{a}{8}\left(\frac{2}{3}\right)^n x^{(n-1)/2} \qquad \text{with } n = \frac{\log 4x}{\log 3/4}. \tag{3}$$

The other values in $(1/4, 1)$ can be computed recursively using the identities (1) and (2), and the constant a follows from the normalization condition. The resulting pdf is illustrated in Fig. 3. The distribution has a mean 0.5 and variance $\sigma^2 \approx 0.0192$. Note that in order to get the time averages for the actual sending rate one must integrate over $f(x)$ with appropriate weights ($f(x)$ is merely the steady state distribution of the embedded Markov chain!).

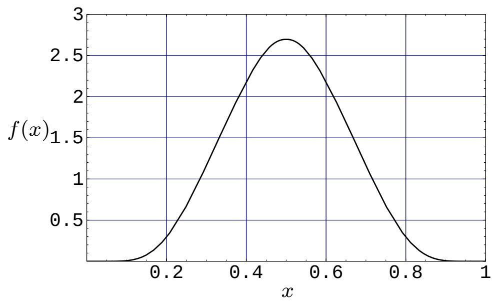
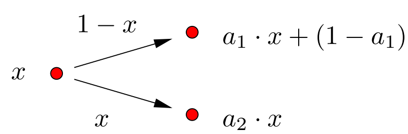


Fig. 3. Steady state distribution $f(x)$ of the embedded chain for two competing flows.

C. General case

The above Markov chain is a special case of a discrete time Markov process with a continuous state space in $(0, 1)$ of $\mathbb{R}$ with the transition probabilities according to



where $0 < a_1 < 1$ and $0 < a_2 < 1$. Let $\alpha_1 = 1/a_1$, and $\alpha_2 = 1/a_2$. Then the *global balance equations* for this system can be written as

$$\begin{aligned} f(x) &= \alpha_1^2 \cdot (1-x) \cdot f(1-\alpha_1+\alpha_1 x) \\ &\quad + \alpha_2^2 \cdot x \cdot f(\alpha_2 x). \end{aligned} \tag{4}$$

Again, we know that

$$f(x) = 0 \quad \forall\, x < 0 \text{ and } x > 1.$$

In particular, for $0 \leq x \leq 1 - a_1$ no transitions are possible from t to x, for $t < x$, and (4) simplifies to

$$f(x) = \alpha_2^2 \cdot x \cdot f(\alpha_2 x). \tag{5}$$

Values for $x > 1 - a_1$ can then be computed recursively using (4) once the solution for (5) is known. In particular, it turns out that for $0 < x \leq 1 - a_1$ and $n = 0, 1, \ldots$ we have

$$f(x) = \alpha_2^{(n^2+3n)/2} \cdot x^n \cdot f(\alpha_2^n x),$$

which can be written as

$$f(a_2^n x) = a_2^{(n^2-3n)/2} \cdot x^n \cdot f(x).$$

Choosing $x = 1 - a_1$ (the upper bound of the interval), we have

$$f(a_2^n(1-a_1)) = a_2^{(n^2-3n)/2} \cdot (1-a_1)^n \cdot f(1-a_1).$$

Letting $r = a_2^n(1 - a_1)$ we finally have a continuous solution for the interval $x \in (0, 1 - a_1)$,

$$f(r) = a_2^{(n^2-3n)/2} \cdot (1-a_1)^n \cdot f(1-a_1), \tag{6}$$

with

$$n = \frac{\log r - \log(1-a_1)}{\log a_2}.$$

The value $f(1 - a_1)$ is obtained by normalization. With $a_1 = a_2 = 3/4$ the above reduces into (3).

D. Flow aggregation approach

Generally there are m TCP flows sharing the same bandwidth. The straightforward analysis, however, leads to complicated equations as the number possible transitions from each state increases linearly as a function of m. In order to reduce the complexity of the model let us next consider an approximation where we have $m - 1$ aggregated flows and a "targeted flow" 1.

Let $X_1 = x$ denote the sending rate of flow 1 upon a negative feedback signal, so that the aggregated flows have a total sending rate of $1 - x$. We assume that the aggregated flows share this bandwidth of $1 - x$ equally at the point of time when the negative feedback signal is generated.

In this case the state transition probabilities are as follows. With probability of $1-x$ at state x (corresponding to the sending rate of flow 1) the negative feedback is sent to one of the aggregated flows corresponding to a decrease in the sending rate equal to

$$\Delta = (1-\nu) \cdot \frac{1-x}{m-1},$$

so that the next state of the embedded chain is

$$\begin{aligned} x + \frac{\Delta}{m} &= \frac{x(m^2-m) + (1-x)(1-\nu)}{m(m-1)} \\ &= \frac{m^2-m+\nu-1}{m(m-1)} \cdot x + \frac{1-\nu}{m(m-1)}. \end{aligned}$$

Similarly, with the probability of x the rate of flow 1 is reduced and the decrease is

$$\Delta = (1-\nu)x,$$

and the next state of the chain is

$$\nu x + \frac{\Delta}{m} = \frac{(m-1)\nu + 1}{m} \cdot x.$$

In particular, this corresponds to the model described at the start of Section III-D with

$$a_1 = \frac{m^2 - m - (1-\nu)}{m^2 - m} \quad \text{and} \quad a_2 = \frac{(m-1)\nu + 1}{m}.$$

With $\nu = 1/2$ the above reduces into

$$a_1 = \frac{2m^2 - 2m - 1}{2m^2 - 2m} \quad \text{and} \quad a_2 = \frac{m+1}{2m}. \tag{7}$$

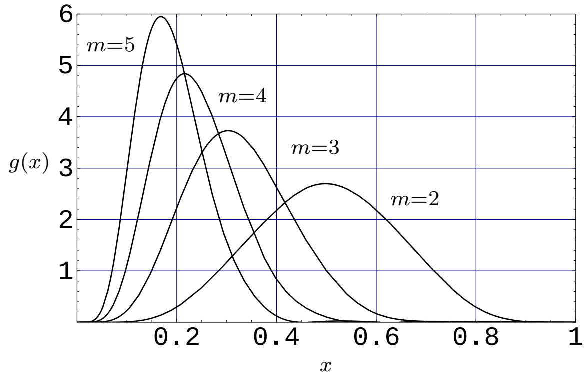


Fig. 4. Steady state distribution $g(x)$ of the embedded Markov chain in aggregated flow approach with $m = 2, 3, 4, 5$ flows.

The steady state distribution $g(x)$ of this system is illustrated in Fig. 4 for $m = 2, 3, 4, 5$ flows. It can be seen that as the number of flows increases the mean and the variance of the distribution decrease, as expected.

In summary, constants a_1 and a_2 depend on the number of flows m and the multiplicative factor ν according to (7), and the possible transitions and total decrements from a given state x are as follows:

trans.	$\mathrm{P}(\cdot)$	new state	decrement Δ
up	$1-x$	$a_1 x + (1-a_1)$	$m((1-a_1) - (1-a_1)x)$
down	x	$a_2 x$	$\dfrac{m}{m-1}(1-a_2)x$

E. Mean period and mean sending rate

Let random variable Δ denote the drop in the total sending rate upon a negative feedback, which, with our choice of time scale, also corresponds to the time between two negative feedbacks (i.e., packet losses). During a long time interval of length T there are on average $M = T/\mathrm{E}[\Delta]$ periods. The total "area" of these triangles is (on average) $M \cdot \frac{1}{2} \cdot \mathrm{E}\left[\Delta^2\right]$. Hence, the mean sending rate is given by

$$\mathrm{E}[R] = 1 - \frac{1}{2} \frac{\mathrm{E}\left[\Delta^2\right]}{\mathrm{E}[\Delta]}. \tag{8}$$

Generally one can consider a single epoch during which the sending rate $R(t)$ is given by $R(t) = t + 1 - \Delta$. Hence, we immediately obtain for the nth moment,

$$\mathrm{E}[R^n] = \frac{\mathrm{E}\left[\int_0^\Delta (t+1-\Delta)^n \, dt\right]}{\mathrm{E}[\Delta]} = \frac{1-\mathrm{E}\left[(1-\Delta)^{n+1}\right]}{(n+1)\cdot \mathrm{E}[\Delta]}.$$

In order to evaluate the mean sending rate, $\mathrm{E}[R]$, we need to know the first two moments of random variable Δ. Generally, we have

$$\mathrm{E}\left[\Delta^k\right] = \int_0^1 x\, f(x) \cdot [(1-\nu)x]^k \, dx + \int_0^1 (1-x)\, f(x) \cdot \left[(1-\nu)\frac{1-x}{m-1}\right]^k dx,$$

which reduces into

$$\mathrm{E}\left[\Delta^k\right] = (1-\nu)^k \int_0^1 f(x) \left[x^{k+1} + \frac{(1-x)^{k+1}}{(m-1)^k}\right] dx. \tag{9}$$

Substituting (9) into (8) gives us then an expression for determining the mean sending rate with m flows,

$$\mathrm{E}[R] = 1 - \frac{1}{2} \cdot \frac{\mathrm{E}\left[\Delta^2\right]}{\mathrm{E}[\Delta]} = 1 - \frac{(1-\nu)}{2} \cdot \frac{\int_0^1 f(x) \left[x^3 + \frac{(1-x)^3}{(m-1)^2}\right] dx}{\int_0^1 f(x) \left[x^2 + \frac{(1-x)^2}{m-1}\right] dx}. \tag{10}$$

Note that even though the constant ν does not explicitly exist in the integrand(s), the pdf $f(x)$ depends on it. Furthermore, due to our idealized assumption of immediate negative feedback signals, the optimal ν is clearly $1-\epsilon$, which yields a constant total sending rate of 1. In practice this is of course not possible due to strictly positive RTT's, but one must choose a smaller ν, e.g., $\nu = 1/2$, in order to ensure stability in the network.

At the limit $m \to \infty$ (10) reduces into

$$\mathrm{E}[R] = 1 - \frac{1-\nu}{2} \cdot \frac{\mathrm{E}\left[X^3\right]}{\mathrm{E}\left[X^2\right]}, \tag{11}$$

where the random variable X denotes the rate of the targeted flow 1 at the moment when the capacity limit is reached and a negative feedback is sent.

1) two flows with $\nu = 1/2$: In this case we can easily write the pdf for the drop in total sending rate by giving each event "sending rate is reduced by Δ" a weight according to the duration of the event, i.e., also Δ in our case. If the capacity limit is reached with sending rates $X_1 = x$ and $X_2 = 1-x$ for flows 1 and 2, respectively, then with probability of x the drop Δ is equal to $x/2$ and with probability of $1-x$ it is equal to $(1-x)/2$. Pdf $f(x)$ corresponds to the steady state distribution of X_1 and we obtain for the pdf $h(x)$ of random variable Δ,

$$h(x) = 8x \cdot f(2x),$$

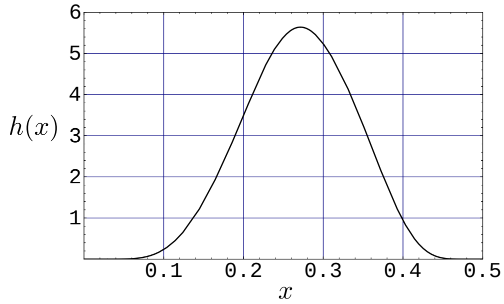


Fig. 5. Steady state distribution $h(x)$ for the drop Δ.

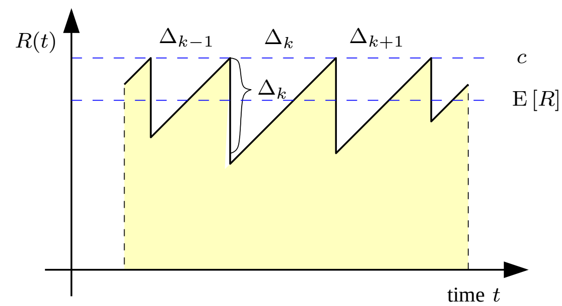


Fig. 6. Random variable Δ_k corresponds to the drop in total sending rate upon a negative feedback, and also, to the time interval between two consecutive negative feedback signals.

which is illustrated in Fig. 5. Consequently, we have

$$\mathrm{E}\left[\Delta^k\right] = 2^{1-k} \cdot \mathrm{E}\left[X^{k+1}\right],$$

and the following numerical values:

$$\begin{aligned} \mathrm{E}[\Delta] &= 0.269, & \sigma_\Delta^2 &= 0.0044, \\ \mathrm{E}\left[\Delta^2\right] &= 0.0769, & \mathrm{E}[R] &= 0.857. \end{aligned}$$

Thus, the rate of the negative feedback signals in this system is about 3.71 of which half belong to each particular flow.

2) $m-1$ aggregated flows: In this case, for $\nu = 1/2$ and $m = 2,3,4,5$, the mean drop $\mathrm{E}[\Delta]$ is

$$\mathrm{E}[\Delta] = \big(0.27, \quad 0.17, \quad 0.13, \quad 0.10\big),$$

and the standard deviation of Δ, similarly,

$$\sigma_\Delta = \big(0.067, \quad 0.036, \quad 0.022, \quad 0.020\big).$$

Consequently, the mean total rate $\mathrm{E}[R]$ for $m = 2,3,4,5$ TCP flows is approximately

$$\mathrm{E}[R] = \big(0.86, \quad 0.91, \quad 0.93, \quad 0.95\big).$$

Using (11) instead of (10) gives

$$\hat{R} = \big(0.93, \quad 0.95, \quad 0.97, \quad 0.97\big).$$

F. Constant bit rate UDP flow

In this model UDP flow(s) simply consume a fraction of the total capacity c. When reaching the capacity limit the negative feedback is sometimes sent (erroneously) to UDP source, which does not adjust its sending rate, and consequently, another negative feedback will be sent (or the packet dropped).

Let r_{udp} denote the rate of the (CBR) UDP traffic and c the total capacity of the link. Then, each time the total rate achieves the link capacity (with Bernoulli trial assumption) on average

$$\frac{r_{\text{udp}}}{c - r_{\text{udp}}},$$

UDP packets are dropped (unnecessarily). Thus, we have an elementary model for the loss process in UDP flows.

IV. Correlations in the Loss Process

Let us again first consider ideal TCP flows sharing a bottleneck link with unit capacity, $c = 1$. In the previous sections we have derived, both accurate and approximative, expressions for the pdf $f(x)$ corresponding to the bandwidth sharing at the time instances when the system reaches the capacity limit $c = 1$ for different number of TCP flows sharing the bottleneck link. At those time instances a negative feedback signal is sent to a randomly chosen source based on their current sending rates causing a reduction in the total sending rate $\Delta = R(t^-) - R(t^+)$ according to the multiplicative decrease scheme. As before, we let Δ_k denote the kth drop in the total sending rate upon a negative feedback, i.e., Δ_k is also the time interval between the kth and the $(k+1)$th packet loss (or more generally, the time interval between two negative feedback signals), as illustrated in Fig. 6. The random variables Δ_k, however, are not independent, i.e., the consecutive time intervals between packet losses have a correlation which we will study in this section.

A. Two TCP flows

Assuming two TCP flows and a multiplicative factor of $1/2$ we have

$$\Delta_{k+1} = \begin{cases} \frac{3}{4}\,\Delta_k, & \text{with probability of } \frac{3}{2}\,\Delta_k, \\ \frac{1}{2} - \frac{3}{4}\,\Delta_k, & \text{with probability of } 1 - \frac{3}{2}\,\Delta_k. \end{cases} \tag{12}$$

This is an interesting expression as it defines a Markov chain between two consecutive time intervals. Moreover, from (12) we obtain the conditional expectation,

$$\begin{aligned} \mathrm{E}[\Delta_{k+1} \,|\, \Delta_k] &= \frac{1}{2}\left(\frac{3}{2}\Delta_k\right)^2 + \frac{1}{2}\left(1 - \frac{3}{2}\Delta_k\right)^2 \\ &= \frac{9}{4}\Delta_k^2 - \frac{3}{2}\Delta_k + \frac{1}{2}. \end{aligned} \tag{13}$$

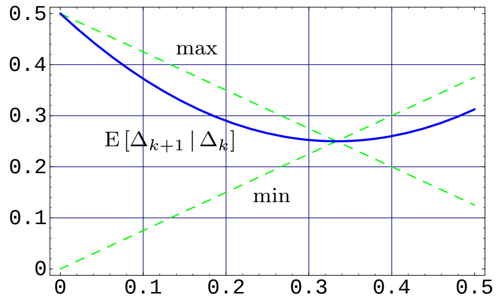


Fig. 7. The minimum, maximum and mean of Δ_{k+1} conditioned on Δ_k (x-axis) in case of two flows according to (12) and (13).

These are illustrated in Fig. 7 where x-axis corresponds to Δ_k and y-axis to Δ_{k+1}. The covariance between two consecutive time intervals between packet losses is

$$\begin{aligned}
&\mathrm{Cov}\left[\Delta_k,\, \Delta_{k+1}\right] = \mathrm{E}\left[\Delta_k \Delta_{k+1}\right] - \mathrm{E}\left[\Delta_k\right] \cdot \mathrm{E}\left[\Delta_{k+1}\right] \\
&\quad = \mathrm{E}\left[\mathrm{E}\left[\Delta_k \Delta_{k+1} \,|\, \Delta_k\right]\right] - \mathrm{E}\left[\Delta\right]^2 \\
&\quad = \mathrm{E}\left[\frac{9}{4}\Delta_k^3 - \frac{3}{2}\Delta_k^2 + \frac{1}{2}\Delta_k\right] - \mathrm{E}\left[\Delta\right]^2 . \\
&\quad = \frac{9}{4}\mathrm{E}\left[\Delta^3\right] - \frac{3}{2}\mathrm{E}\left[\Delta^2\right] + \frac{1}{2}\mathrm{E}\left[\Delta\right] - \mathrm{E}\left[\Delta\right]^2 ,
\end{aligned}$$

where the last two terms can be also written as

$$\mathrm{E}\left[\Delta\right](1/2 - \mathrm{E}\left[\Delta\right]).$$

The different moments of random variable Δ can be computed using (9). Numerically we obtain

$$\mathrm{E}\left[\Delta^3\right] \approx 0.0231,$$

and get,

$$\mathrm{Cov}\left[\Delta_k,\, \Delta_{k+1}\right] \approx -0.00134.$$

Thus, the correlation coefficient is

$$\rho = \frac{\mathrm{Cov}\left[\Delta_k,\, \Delta_{k+1}\right]}{\sqrt{\mathrm{V}\left[\Delta_k\right] \cdot \mathrm{V}\left[\Delta_{k+1}\right]}} \approx -0.303.$$

Note that if time intervals Δ_k and Δ_{k+1} were independent the covariance would be equal to zero, and this clearly is not the case (not even approximately when the number of flows is small).

B. *General case, m flows*

We note that for the general case (i.e., the flow aggregation approach) a similar analysis is straightforward by first conditioning on the state of the system, $X_k = x$, just before the first negative feedback. On this condition there are two possible outcomes for drop Δ_k, and recursively, $2 \cdot 2 = 4$ possible outcomes for the pair (Δ_k, Δ_{k+1}). Combining this with an appropriate pdf $f(x)$ allows one to compute the covariance in the general case. Due to lack of space we omit the details here.

V. Numerical Examples

It should be emphasized again that the purpose of this model is not to model the behavior of the actual TCP variant accurately, but rather to serve as an elementary model for studying the behavior of TCP-like congestion control mechanism at the microscopic level, i.e., at the time scales of the time interval between packet losses.

Next we will, however, discuss how the model parameters can be chosen in order to match a realistic scenario in some degree. The nature of the fluid model implies that the model is only accurate when the packet size (MTU) is small when compared to the product of the link capacity (c) and the round trip time (RTT), i.e.,

$$\mathrm{MTU} \ll c \cdot \mathrm{RTT}. \tag{14}$$

In congestion avoidance phase, a TCP source is supposed to increase its window approximately by one packet per round trip time (RTT), i.e.,

$$\frac{\Delta \mathrm{W}}{\Delta t} = \frac{1}{\mathrm{RTT}}.$$

On the other hand, the sending rate of a TCP source equals the current congestion window, W, times the maximum segment size, MTU, divided by the RTT,

$$r_i = \frac{\mathrm{W} \cdot \mathrm{MTU}}{\mathrm{RTT}},$$

and consequently,

$$\frac{\partial r_i}{\partial t} = \frac{\mathrm{MTU}}{\mathrm{RTT}^2}.$$

In our model we have scaled the maximum capacity to 1 and the time so that the linear increase rate equals to $1/m$, where m is the total number of TCP flows. Let

$$r_i = \alpha \cdot r_i^*, \quad \text{and} \quad t = \beta \cdot t^*,$$

where the starred versions correspond to the rates and the time in our model. With these,

$$\frac{\Delta r_i}{\Delta t} = \frac{\alpha \Delta r_i^*}{\beta \Delta t^*} = \frac{\alpha}{\beta} \cdot \frac{1}{m},$$

and hence

$$\frac{\alpha}{\beta} = \frac{m \cdot \mathrm{MTU}}{\mathrm{RTT}^2}.$$

Next we must fix the level at which a negative feedback is sent (or reaches the source to be exact). To this we can choose any α for which

$$\alpha \cdot \mathrm{E}\left[R\right] < c,$$

where c corresponds to nominal link service rate (stability condition). We note that the interesting area is when

$$c < \alpha < c \,/\, \mathrm{E}\left[R\right],$$

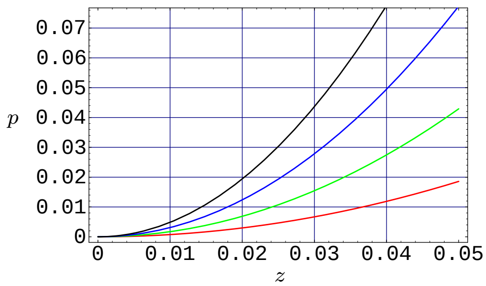


Fig. 8. The packet loss probability p according to (15) for $m = 2, 3, 4, 5$ flows (from lowest to highest) as a function of quantity $z = \mathrm{MTU}/(c \cdot \mathrm{RTT})$.

as then the arrival rate exceeds the service rate momentarily and the transmission queue builds up causing queueing delays within the router.

The *blocking probability* we can estimate as follows. The average total sending rate (in packets/s) is equal to

$$\alpha \cdot \frac{\mathrm{E}[R]}{\mathrm{MTU}}.$$

On the other hand, there are on the average

$$\frac{\Delta T}{\beta \cdot \mathrm{E}[\Delta]}$$

packet losses during a long time interval of ΔT. Thus,

$$p = \frac{\mathrm{MTU}}{\alpha\beta \cdot \mathrm{E}[\Delta] \cdot \mathrm{E}[R]} = \frac{\mathrm{MTU}}{\alpha\beta \cdot (\mathrm{E}[\Delta] - \mathrm{E}[\Delta^2]/2)}$$
$$\approx \left(\frac{\mathrm{MTU}}{c \cdot \mathrm{RTT}}\right)^2 \cdot \frac{m}{\mathrm{E}[\Delta]}. \tag{15}$$

The behaviour of the loss probability is illustrated in Fig. 8 for $m = 2, 3, 4, 5$ flows (using the flow aggregation approximation).

Example: Consider a 10 Mbit/s link with $\mathrm{MTU} = 576$ bytes and $\mathrm{RTT} = 20$ ms, i.e.,

$$\frac{\mathrm{MTU}}{c \cdot \mathrm{RTT}} \approx 0.023.$$

The packet loss probabilities for $m = 2, 3, 4, 5$ flows are

$$\mathbf{p} = \begin{pmatrix} 0.0039, & 0.0091, & 0.016, & 0.026 \end{pmatrix}.$$

VI. Conclusions

We have presented an elementary model for TCP congestion control mechanism where the focus has been on the interactions between the concurrent TCP flows at the microscopic level. In contrast to the previous work on the TCP analysis, this model manages to catch the coupling with different concurrent TCP flows correctly (a lost packet in one flow allows higher window sizes in the other flows). Especially, the correlation between the consecutive time intervals between the packet losses is modelled appropriately instead of assuming independence among them (the usual approach). We were also able to solve the proposed model by analytical means and compute several key performance figures numerically.

Moreover, by using this model it is possible to characterize the loss process of TCP and UDP flows sharing the same bottleneck link. The future work includes incorporating unequal RTT times in the model and evaluating the accuracy of the model against simulation results and actual measurements from the network.

References

[1] Dah-Ming Chiu and Raj Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Computer Networks and ISDN Systems*, vol. 17, no. 1, pp. 1–14, 1989.

[2] Paul Hurley, Jean-Yves Le Boudec, and Patrick Thiran, "A note on the fairness of additive increase and multiplicative decrease," in *Proceedings of ITC-16*, Edinburgh, UK, June 1999.

[3] Jim Kurose and Keith Ross, *Computer Networking: a top-down approach featuring the Internet*, Addison-Wesley, 2001.

[4] Sally Floyd and Kevin Fall, "Promoting the use of end-to-end congestion control in the Internet," *IEEE/ACM Trans. Networking*, vol. 7, no. 4, pp. 458–472, 1999.

[5] Jitendra Padhye, Victor Firoiu, Donald F. Towsley, and James F. Kurose, "Modeling TCP reno performance: a simple model and its empirical validation," *IEEE/ACM Trans. Networking*, vol. 8, no. 2, pp. 133–145, Apr. 2000.

[6] T. V. Lakshman and Upamanyu Madhow, "The performance of TCP/IP for networks with high bandwidth-delay products and random loss," *IEEE/ACM Trans. Networking*, vol. 5, no. 3, pp. 336–350, 1997.

[7] Matthew Mathis, Jeffrey Semke, and Jamshid Mahdavi, "The macroscopic behavior of the TCP congestion avoidance algorithm," *SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 3, pp. 67–82, 1997.

[8] S. Savari and E.Telatar, "The behavior of certain stochastic processes arising in window protocols," in *Global Telecommunications Conference, 1999. (GLOBECOM'99), Volume 1B*, 1999, vol. 1b, pp. 791–795.

[9] Eitan Altman, Konstantin Avrachenkov, and Chadi Barakat, "A stochastic model of TCP/IP with stationary random losses," *IEEE/ACM Trans. Networking*, vol. 13, no. 2, pp. 356–369, 2005.

[10] Sally Floyd, "TCP and explicit congestion notification," *ACM Computer Communication Review*, vol. 24, no. 5, pp. 10–23, Oct. 1993.

[11] Sally Floyd and Van Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Networking*, vol. 1, no. 4, pp. 397–413, 1993.