Analysis of MIMD Congestion Control Algorithm for High Speed Networks [1]

E. Altman, K. Avrachenkov, C. Barakat, A. A. Kherani and B. J. Prabhu

INRIA Sophia-Antipolis, 2004 route des lucioles, 06902 Sophia-Antipolis, France
Email:{altman, k.avrachenkov, cbarakat, alam, bprabhu}@sophia.inria.fr

Abstract

Proposals to improve the performance of TCP in high speed networks have been recently put forward. Examples of such proposals include High-Speed TCP, Scalable TCP, and FAST. In contrast to the additive increase multiplicative decrease algorithm used in the standard TCP, Scalable TCP uses a multiplicative increase multiplicative decrease (MIMD) algorithm for the window size evolution. In this paper, we present a mathematical analysis of the MIMD congestion control algorithm in the presence of random losses. Random losses are typical to wireless networks but can also be used to model losses in wireline networks with a high bandwidth delay product. Our approach is based on showing that the logarithm of the window size evolution has the same behaviour as the workload process in a standard G/G/1 queue. The Laplace-Stieltjes transform of the equivalent queue is then shown to directly provide the throughput of the congestion control algorithm and the higher moments of the window size. Using *ns-2* simulations, we validate our findings using Scalable TCP.

Key words: MIMD congestion control, Scalable TCP, Laplace-Stieltjes transform, discrete time queues, random losses.

1 Introduction

In high speed networks, the congestion avoidance phase of TCP takes a long time to increase the window size and fully utilize the available bandwidth.

[1] This work was partially supported by the EURO NGI network of Excellence, by the INRIA's TCP ARC collaboration project and by the *Indo–French Center for Promotion of Advanced Research (IFCPAR)* under research contract number 2900–IT.

S. Floyd writes in [1]: "for a Standard TCP connection with 1500-byte packets and a 100 ms round-trip time, achieving a steady-state throughput of 10 Gbps would require an average congestion window of 83,333 segments, and a packet drop rate of at most one congestion event every 5,000,000,000 packets (or equivalently, at most one congestion event every $1\frac{2}{3}$ hours).The average packet drop rate of at most 2×10^{-10} needed for full link utilization in this environment corresponds to a bit error rate of at most 2×10^{-14}, and this is an unrealistic requirement for current networks." Thus, in the context of high speed networks, it is essential to study the effect of random packet losses on TCP, since they may limit the TCP throughput more than the congestion losses do and may lead to a poor utilization of the large available capacity. The modeling of random losses is also essential in the study of TCP performance over wireless channels.

In order to improve the utilization of the available capacity in high speed networks, modifications to the standard TCP have been proposed in [1]-[3]. In [4], T. Kelly has proposed a variation of TCP, called Scalable TCP, wherein upon each ACK it receives, the sender increases its congestion window ($cwnd$) by 0.01 packets. When a loss event is detected, the sender decreases $cwnd$ by a factor of 0.125. Hence, if the window size is $W(t)$ at some time t (meaning that there are $W(t)$ unacknowledged packets in the network) then, in the absence of losses, the window size after an RTT (round-trip time), $W(t+RTT)$, would be $1.01 \times W(t)$, whereas if there are losses during $(t, t+RTT)$, $W(t+RTT)$ will be around $0.875 \times W(t)$ (here we assume that, as in New Reno and SACK, the window is reduced only once during a round trip time even if there are several losses). A feature of this algorithm is that, starting from a window size of some fraction of the bandwidth-delay product (BDP), the number of $RTTs$ required to reach BDP is independent of the link speed.

Consider the class of Multiplicative Increase and Multiplicative Decrease (MIMD) congestion control algorithms where each ACK results in a window increment of $\alpha - 1 > 0$ and a loss event is responded with a reduction of window size by a fraction $1 - \beta < 1$. Scalable TCP can then be viewed as an instance from this class with $\alpha = 1.01$ and $\beta = 0.875$. This motivates us to study the window behaviour of MIMD congestion control algorithms for the purpose of studying Scalable TCP. We focus on the analytical performance study of these algorithms, and, hence, of Scalable TCP, in the presence of random as well as congestion losses. In the rest of the paper, Scalable TCP and MIMD algorithm will be used interchangeably.

Related Work: TCP is frequently modeled using window-dependent losses where sessions with large window sizes typically create congestion so that the loss probability increases with window size. Indeed, in the wireline environment, losses are frequently assumed to be caused by buffer overflow, see e.g. [5] that studies Scalable TCP where losses are caused by buffer overflow which occurs when the window is sufficiently large. In contrast, the main emphasis of our study is on the performance of a single MIMD source in the presence of random window-independent losses. The independence has been observed in connections over wide area networks [6], but has also been studied and advocated in other contexts, see e.g. [7], [8]. Another important research direction related to congestion control protocols is fairness. Fairness issues arise when multiple sources share a common link. Chiu and Jain [9] showed that the MIMD algorithm is unfair in the presence of synchronous losses. However, in [10] Altman *et al* show that fairness amongst MIMD sources can be achieved by introducing some asynchronous losses. Furthermore, the authors also study inter-protocol fairness (i.e., fairness among MIMD and AIMD sources) in the presence of synchronous losses.

In Section 2 we present three models based on different assumptions on the window size. Then we present a general analysis of these models. Our approach is based on showing that an invertible transformation applied to the window size process results in a process that has the same evolution as the total workload process in a standard G/G/1 queue. The Laplace-Stieltjes transform of the equivalent queueing process thus obtained provides the throughput of the connection as well as the higher moments of the window size of the given MIMD algorithm (Section 3). We study a model with only random losses (Section 4) as well as a model where, in addition to random losses, there are either congestion losses or the window size is upper bounded (Section 5). We also present an exact analysis of a model with upper and lower bound on the window size (Section 6). In Section 7 we first present a model in which losses in an RTT depend on the window size in that RTT. Then, we propose an approximation to this model, and relate this approximate model with the model in Section 5. We validate our findings using *ns-2* simulations (Section 8) and end with a concluding section.

2 Discrete Time Models

We consider the scenario where a single FTP application transfers data using an MIMD flow control protocol with parameters α and β as mentioned in the Introduction. The transfer is done over a high speed path consisting of one or more high speed links. We assume that the file is sufficiently large to ensure the convergence to a stationary regime. In this section, we introduce different models of MIMD schemes for different network conditions.

Let τ denote the round-trip propagation delay of the high speed path (in literature this is also referred to as the fixed part of the round-trip time). Let c packets per second be the link capacity of the slowest link. Let $\{W(t), t \geq 0\}$ denote the window process evolving over time. Use $W_0 \triangleq W(0)$ and let $\tau_1 \triangleq$

$\tau + \left(\frac{W_0}{c} - \tau\right)^+$ denote the first round-trip time. Let $W_1 \triangleq W(\tau_1)$ and define $\tau_2 = \tau + \left(\frac{W_1}{c} - \tau\right)^+$. Proceeding in this manner, we get a sequence $\{\tau_n, n \geq 1\}$ of round-trip times and a sequence $\{W_n, n \geq 0\}$ of window sizes. Consider a sequence of time instants $\{t_n, n \geq 0\}$ where t_n is the end of n^{th} round-trip time, i.e., $t_0 = 0$, $t_1 = \tau_1$ and $t_n = t_{n-1} + \tau_n$. Under our definition, W_n is the window size at time instant t_n. The window evolution can now be written as

$$W_{n+1} = \begin{cases} \alpha W_n, \text{ if there was no loss in interval } [t_n, t_{n+1}], \\ \beta W_n, \text{ if there were one or more losses in } [t_n, t_{n+1}]. \end{cases}$$

We shall consider the following models for the evolution of $\{W_n\}$ under random losses:

(i) There is no upper bound on the window size.
(ii) There is an upper bound B on the window size which corresponds to an explicit limitation of the window size. When this value is reached then the window stops growing.
(iii) There is an upper bound B on the window size. However, when this value is reached, the connection suffers a congestion loss (this is in addition to the random losses) and the multiplicative decrease of window is invoked.

The first model approximates the scenario where the link BDP is high and there is a significant probability of loss in a round-trip time so that the practical upper bound of BDP is reached with negligible probability. The second model corresponds to the case where the window is bounded by the receiver's advertised window. The last model corresponds to the case where the window reaches the value of round-trip pipe size (BDP+Buffer) and suffers a loss owing to buffer overflow.

2.1 Window Evolution for the Proposed Models

Let $A_n, n \geq 1$, be a random variable such that $A_n = \alpha$ if there was no loss in the interval $[t_n, t_{n+1}]$, and $A_n = \beta$ otherwise. Throughout this section, $\{A_n, n \geq 1\}$ will be assumed to be a general stationary ergodic sequence. Now we describe the evolution of $\{W_n\}$ in terms of $\{A_n\}$ recursively for the models described above.

Model (i): Taking into account the fact that the window size of TCP is bounded below by a value of one packet [2], the window size evolution for this

[2] There is no loss of generality as one can consider any value of the minimal window size and then rescale the model.

model can be written as

$$W_{n+1} = \max(A_n W_n, 1). \tag{1}$$

In practice, the sender's window is either upper bounded by the receiver's advertised window or by the pipe size (BDP+Buffer). However, in large BDP networks, when the losses are sufficiently frequent so that the upper bound of the window, B, is rarely reached, this model can be used to approximate the behaviour of the sender's window.

Model (ii): On the contrary, if the losses are infrequent and B is often reached and is sufficiently large, we can ignore the lower bound on the window size (which would be rarely attained). Thus, the window evolution for this model can be written as

$$W_{n+1} = \min(A_n W_n, B). \tag{2}$$

Here B models the limitation due to the receiver's advertised window.

Model (iii): The window evolution in this model is similar to that of model (ii). However, there is an instantaneous drop in the window upon reaching the upper bound. In this case, B models the pipe size. The receiver's advertised window is assumed not to limit the performance.

In the next section we relate the window process under the different models introduced in this section to the workload evolution in a discrete time G/G/1 queue.

3 Preliminary Analysis

Consider the following discrete time stochastic recursive equation

$$W_{n+1} = \max(A_n W_n, 1). \tag{3}$$

The process, $\{W_n\}$, can be viewed as a sequence of observations of a continuous time process sampled at certain, not necessarily equal, time intervals. The sequence $A_n \in (0, \infty)$ is assumed to be stationary and ergodic.

Taking the logarithm of equation (3), we obtain

$$\log[W_{n+1}] = \max(\log[A_n] + \log[W_n], 0).$$

Using the substitutions $Y_n = \log[W_n]$, and $U_n = \log[A_n]$ in the above equation, we obtain

$$Y_{n+1} = \max(Y_n + U_n, 0). \tag{4}$$

We now make the following observation: The recursive equation (4) has the same form as the equation describing the workload process in a G/G/1 queue observed at, say, just after an arrival (see, for example, [11]). U_n denotes the difference between the service time of the n^{th} customer and the interarrival time between the n^{th} and the $(n+1)^{th}$ customer. Since the introduced transformation, $\log(\cdot)$, is invertible, there is a one to one correspondence between the processes $\{Y_n, n \geq 0\}$ and $\{W_n, n \geq 0\}$. This observation allows us to study the stability of the window process $\{W_n, n \geq 0\}$ via that of $\{Y_n, n \geq 0\}$. Furthermore, the analogy with queueing theory of the process $\{Y_n, n \geq 0\}$ allows us to obtain the steady state moments of W_n.

Theorem 3.1 *Assume that $E[\log A_0] < 0$. Then there exists a unique stationary ergodic process $\{W_n^*\}$, defined on the same probability space as $\{W_n\}$, that satisfies the recursion (3). Moreover, for any initial value $W_0 = w$, there is a random time T_w, which is finite with probability 1, such that $W_n = W_n^*$ for all $n \geq T_w$. If $E[\log A_0] > 0$ then W_n tends to infinity w.p.1 for any initial value $W_0 = w$.*

PROOF. According to Theorem 2A [12], if $E[\log A_0] < 0$ then the stochastic process $\{Y_n\}$ converges to a stationary ergodic process $\{Y_n^*\}$ which is defined on the same probability space as $\{Y_n\}$ and is the unique stationary regime that satisfies (4). This implies the statement for $W_n = \exp(Y_n)$. The last part of theorem similarly follows from [13, p. 36]. □

Remark 1 *Due to Jensen's inequality and the concavity of the logarithmic function, $E[\log A_0] \leq \log E[A_0]$. Hence, $\log E[A_0] < 0$, or equivalently $E[A_0] < 1$, is a sufficient condition for the stability of the window process $\{W_n\}$ (for the existence of a unique stationary ergodic regime and the convergence to this regime). However this condition is in general not a necessary one.*

Remark 2 *We stress the importance of the maximum operator in equation (3). Indeed, if we eliminate it and write simply $W_{n+1} = A_n W_n$ then on taking the log, instead of (4) we get $Y_{n+1} = \log[A_n] + Y_n$. Its solution is*

$$Y_n = Y_0 + \sum_{i=0}^{n-1} \log[A_i].$$

Since $\{A_n\}$ is stationary ergodic, the strong law of large numbers implies that if $E \log[A_i] < 0$ then Y_n converges to $-\infty$, and, therefore, W_n converges to 0 which is clearly a bad estimate for the window size process. (If $E \log[A_i] > 0$ then Y_n and, therefore, W_n converge to ∞ which was also predicted by the model that took the minimum window into account.) Note that in the limiting case of $E[\log A_i] = 0$, if A_i's are independent and identically distributed (i.i.d.) then Y_n is a null recurrent Markov chain and thus unstable.

The log transformation allows us to obtain the moments of W_n in the stationary regime (i.e., moments of W_n^*) from the Laplace-Steiltjes Transform (LST) of Y_n in the stationary regime (i.e., LST of Y_n^*). The LST of Y_n^* is given by

$$G(s) = E[e^{-sY_n^*}],$$

which is defined for $s \in S$, where S is the region of convergence of $G(s)$. For a given integer $k \geq 0$, the k^{th} moment of W_n^* is obtained as follows

$$E[(W_n^*)^k] = E[\exp(kY_n^*)] = G(-k), \tag{5}$$

where $-k$ is assumed to belong to S. If $-k \notin S$ then the corresponding moment is ∞. Thus, all finite moments of W_n^* can be obtained from the LST of Y_n^*.

A similar analysis can be done for the stochastic recursive equation

$$W_{n+1} = \min(A_n W_n, B) \tag{6}$$

by making the transformation $Y_n = \log[B] - \log[W_n]$. The moments of W_n^* can then be obtained from the LST of Y_n^* using the relation

$$E[(W_n^*)^k] = E[B^k \exp(-kY_n^*)] = B^k G(k). \tag{7}$$

All the moments of W_n^* are finite since $G(s)$ is finite for $s \geq 0$.

The recursive equation for model (i), as given by (1), is similar to equation (3). Therefore, the analysis of this model can be done along the lines of the analysis of (3). Similarly, the analysis of models (ii) and (iii) can be done along the lines of the analysis of (6). We note that the analysis of model (iii) is similar to that of model (ii). The equivalent queueing system of model (iii) can be obtained by deleting the idle periods of the equivalent queueing system of model (ii). The throughput of the MIMD algorithm, or the first moment of the window size, under different models, can be obtained from equations (5) and (7).

In the rest of the paper, we make the following assumption.

Assumption 3.1 *: In each RTT, the probability of one or more loss events occurring is p. The loss probability is independent of the window size and independent from one RTT to another.*

We note that the z-transform, which is defined for integer valued random variables, is a discrete analog of the LST. In the following sections, under the above assumption, we derive the LST of the window size, W_n, for the three models.

4 Model (i): Lower Bound on Window and Random Losses

In this section, we analyse model (i) in which the window size is given by

$$W_{n+1} = \max(A_n W_n, B_l),$$

where B_l is a lower bound on the window size, and W_n denotes the window size at the end of the n^{th} RTT. Under the assumption of independent loss probability, p, in each RTT, the sequence A_n is i.i.d with the following distribution

$$A_n = \begin{cases} \alpha & \text{w.p. } 1-p, \\ \beta & \text{w.p. } p. \end{cases}$$

As noted in the previous section, we make the transformation $Y_n = \frac{\log[W_n]-\log[B_l]}{\log[\alpha]}$ (the division by $\log[\alpha]$ is made for convenience). The recursive equation for the process $\{Y_n\}$ is given by (from equation (4))

$$Y_{n+1} = \begin{cases} Y_n + 1 & \text{w.p. } 1-p, \\ (Y_n - k)^+ & \text{w.p. } p. \end{cases}$$

where $k = -\frac{\log[\beta]}{\log[\alpha]}$. The number k is positive, since $\beta < 1$. For this model to be stable, the necessary and sufficient condition is $E\log[A_n] < 0$, or, equivalently,

$$(k+1)p > 1. \tag{8}$$

In the rest of this section, we assume that the above condition is satisfied.

We also make the following two assumptions.

Assumption 4.1 $k = -\frac{\log[\beta]}{\log[\alpha]}$ *is an integer.*

Assumption 4.2 $\frac{\log[B_l]}{\log[\alpha]}$ *is an integer.*

These two assumptions allow us to use a discrete state space, $\mathcal{S} = \{0, 1, 2, ...\}$ for Y_n. Thus, Y_n can be modelled as a discrete state space Markov chain. The state $Y_n = i$ corresponds to $W_n = B_l\alpha^i$. The transition probabilities for this model are shown in Figure 1.

Let $P_n(j), j \in \mathcal{S}$, be the probability of Y_n being in state j at the end of the n^{th} RTT. The probability of being in state j at the end of the $(n+1)^{th}$ RTT is given by

$$\begin{aligned} P_{n+1}(j) &= (1-p)P_n(j-1) + pP_n(j+k), \; j \geq 1 \\ &= p\textstyle\sum_{i=0}^{k} P_n(i), \qquad\qquad\qquad j = 0. \end{aligned} \tag{9}$$

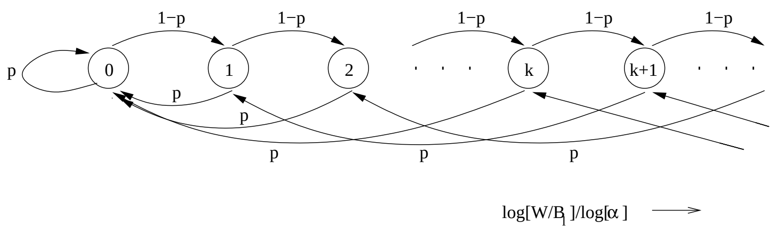


Figure 1. Transition probabilities of Y_n.

Denote the z-transform of Y_n by $\mathbf{Y}_n(z)$. $\mathbf{Y}_n(z)$ is defined as

$$\mathbf{Y}_n(z) = \sum_{j=0}^{\infty} P_n(j) z^j. \tag{10}$$

From (9) and (10) we obtain

$$\begin{aligned} \mathbf{Y}_{n+1}(z) - P_{n+1}(0) &= \sum_{j=1}^{\infty} P_{n+1}(j) z^j \\ &= (1-p) \sum_{j=1}^{\infty} P_n(j-1) z^j + p \sum_{j=1}^{\infty} P_n(j+k) z^j. \end{aligned} \tag{11}$$

Let $\mathbf{Y}(z) = \lim_{n\to\infty} \mathbf{Y}_n(z)$. We note that the steady state window size, W, is given by $W = B_l \alpha^Y$. Therefore, the moments of W can be obtained as follows.

$$E[W^n] = E[(B_l \alpha^Y)^n] = B_l^n \mathbf{Y}(\alpha^n). \tag{12}$$

Theorem 4.1 *(a)* $\mathbf{Y}(z)$ *is given by*

$$\mathbf{Y}(z) = \frac{1 - 1/z_0}{1 - z/z_0}, \tag{13}$$

where z_0 *is the unique root of*

$$\frac{(1-p)}{p} z^{k+1} - \frac{1}{p} z^k + 1 = 0 \tag{14}$$

that lies outside the closed unit disc.
(b)

$$P(W \geq w) = \left(\frac{w}{B_l}\right)^{-\log[z_0]/\log[\alpha]}. \tag{15}$$

PROOF. From Theorem 3.1 we can conclude that the Markov chain Y_n is stationary and ergodic. In particular, this implies that $\mathbf{Y}_n$ converges to $\mathbf{Y}$ and that $P_n(\cdot)$ converges to $P(\cdot)$.

To prove part (a), from (11) we obtain

$$\mathbf{Y}(z) - P(0) = (1-p)z\mathbf{Y}(z) + pz^{-k}\sum_{j=0}^{\infty} P_n(j+1+k)z^{j+1+k}$$

$$\mathbf{Y}(z)(1-(1-p)z) = P(0) + pz^{-k}\sum_{j=0}^{\infty} P_n(j+1+k)z^{j+1+k}$$

$$= P(0) + pz^{-k}(\mathbf{Y}(z) - \sum_{i=0}^{k} P(i)z^i),$$

and hence,

$$\mathbf{Y}(z)((1-(1-p)z)z^k - p) = z^k P(0) - p\sum_{i=0}^{k} P(i)z^i.$$

Since $P(0) = p\sum_{i=0}^{k} P_n(i)$, $\mathbf{Y}(z)$ can be expressed as

$$\mathbf{Y}(z) = \frac{\sum_{i=0}^{k-1} P(i)(z^k - z^i)}{-\frac{(1-p)}{p}z^{k+1} + \frac{1}{p}z^k - 1}. \tag{16}$$

Under the stability condition (8), $\mathbf{Y}(z)$ exists and is analytic in the open disc $\{z : |z| < 1\}$. The numerator of equation (16) has at most $k-1$ zeros inside the unit circle and one zero on the unit circle. Hence, there can be at most $k-1$ zeros of the denominator of equation (16) within the unit circle as any more zeros will make $\mathbf{Y}(z)$ non-analytic. Using Rouche's theorem [14] we can show that there are at least k zeros of the denominator inside and on the unit circle. As $z = 1$ is a zero of the denominator, there are at least $k-1$ zeros inside the unit circle. From the two previous arguments, there are exactly $k-1$ zeros of the denominator within the unit circle and they must be the same as those of the numerator for $\mathbf{Y}(z)$ to be analytic [11]. Hence, $\mathbf{Y}(z)$ reduces to (13).

To prove part (b), we note that the distribution of Y can be obtained by inverting $\mathbf{Y}(z)$, and is given by

$$P(Y = j) = (1 - 1/z_0)(1/z_0)^j, \quad j \geq 0. \tag{17}$$

This, together with the relation $W = B_l \alpha^Y$, gives (15). □

Corollary 3 *Let* $a = \frac{\log[z_0]}{\log[\alpha]}$. *The* n^{th} *moment of* W *is given by*

$$E[W^n] = \begin{cases} B_l \frac{a}{a-n} & n < a \\ \infty & n \geq a \end{cases}. \tag{18}$$

This follows from (12) and (13).

The z-transform $\mathbf{Y}(z)$ is analytic for $z < z_0$. Hence, the n^{th} moment of W is finite if $n < \frac{\log[z_0]}{\log[\alpha]}$. The window size distribution can be seen to become heavy tailed for $a \leq 2$. Thus, for a given loss rate, p, either α or β can be suitably chosen in order to reduce the variance of the window size.

5 Upper Bound on the Window Size or Congestion Losses

In this section, we consider the models where the window at the sender is limited to B_u. This limitation could be due either to the receiver's advertised window (model (ii)) or to the BDP+Buffer limitations (model(iii)). We make the following transformation

$$Y_n = \frac{\log[B_u] - \log[W_n]}{\log[\alpha]}. \tag{19}$$

We make the following assumption along with assumptions (3.1) and (4.2).

Assumption 5.1 $L = \frac{\log[B_u]}{\log[\alpha]}$ *is an integer.*

5.1 Model (ii)

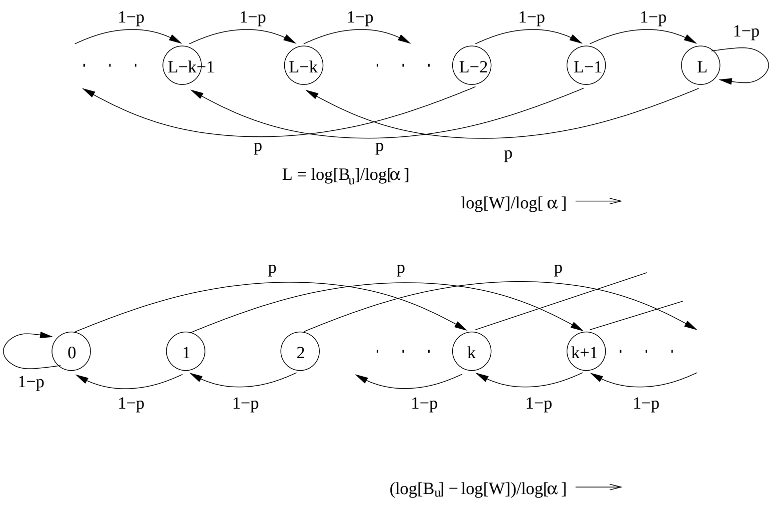


Figure 2. Transition probabilities of $\frac{\log[W_n]}{\log[\alpha]}$ (top figure) and Y_n (bottom figure).

The transition probabilities of $\frac{\log[W_n]}{\log[\alpha]}$ and of Y_n are shown in Figure 2. The state $Y_n = i$ corresponds to the state $W_n = B_u\alpha^{-i}$. The recursive equation for the process $\{Y_n\}$ is given by

$$Y_{n+1} = \begin{cases} Y_n + k & \text{w.p. } p, \\ (Y_n - 1)^+ & \text{w.p. } 1 - p. \end{cases}$$

For this model to be stable, the necessary and sufficient condition is

$$(k+1)p < 1. \tag{20}$$

In the rest of this section, we assume that the above condition is satisfied.

The balance equations for Y_n in steady state can be written as

$$\begin{aligned} P(0) &= \frac{(1-p)}{p} P(1), \\ P(i) &= (1-p)P(i+1), \quad i = 1, ..., k-1. \\ P(i) &= pP(i-k) + (1-p)P(i+1), \quad i \geq k. \end{aligned}$$

These equations are similar to those of a bulk arrival queue.

The z-transform of the steady state probability distribution of Y is defined as

$$\mathbf{Y}(z) = \sum_{j=0}^{\infty} P(j) z^j.$$

Theorem 5.1 $\mathbf{Y}(z)$ *is given by*

$$\mathbf{Y}(z) = (1 - (k+1)p) \frac{1 - z}{pz^{k+1} - z + (1-p)}. \tag{21}$$

PROOF. Following similar arguments from Kleinrock [11], we can write the z-transform as follows.

$$\sum_{i=1}^{\infty} P(i) z^i = \sum_{i=1}^{\infty} pP(i-k) z^i + \sum_{i=1}^{\infty} (1-p) P(i+1) z^i,$$

where $P(i-k) = 0$ for $i < k$. Therefore,

$$\begin{aligned} \mathbf{Y}(z) - P(0) &= pz^k \sum_{i=1}^{\infty} P(i-k) z^{i-k} + z^{-1} \sum_{i=1}^{\infty} (1-p) P(i+1) z^{i+1} \\ &= pz^k \mathbf{Y}(z) + z^{-1}(1-p)(\mathbf{Y}(z) - zP(1) - P(0)), \end{aligned}$$

which gives

$$\mathbf{Y}(z)(z - pz^{k+1} - (1-p)) = P(0)(z - (1-p) - zp)).$$

Using $\mathbf{Y}(1) = 1$ and the L'Hôpital's rule, we get $P(0) = [1 - (k+1)p]/[1-p]$. Hence, we obtain $\mathbf{Y}(z)$ as given by (21). □

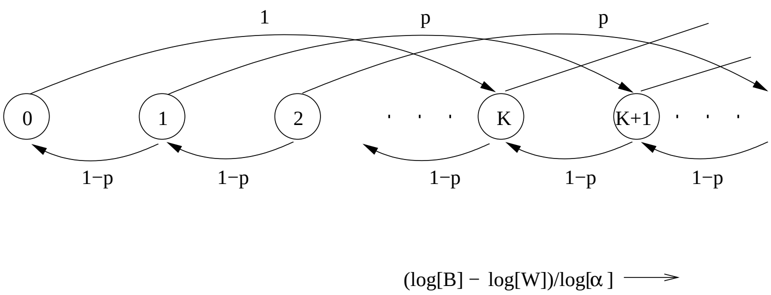
Corollary 4 *The moments of* $W = B_u \alpha^{-Y}$ *are given by*

$$E[W^n] = E[(B_u \alpha^{-Y})^n)] = B_u^n E[\alpha^{-nY}] = B_u^n \mathbf{Y}(\alpha^{-n}). \tag{22}$$

The distribution of Y can be found by inverting $\mathbf{Y}(z)$ using partial fraction expansion. The distribution can be seen to be a weighted sum of geometric distributions.

5.2 Model (iii)

In this model, along with random losses, a loss occurs when the window size reaches B. The transition probabilities of Y_n are shown in Figure 3. As in the previous model, the state $Y_n = i$ corresponds to the state $W_n = B_u \alpha^{-i}$. Note that the transition probability is different from the previous model only



Congestion limited with random losses

Figure 3. Transition probabilities of Y_n.

at $Y = 0$. In this model, there is a jump with probability 1 to state k. The balance equations can be written as,

$$\begin{aligned} P(i) &= (1-p)P(i+1), \quad i = 0, ..., k-1. \\ P(i) &= pP(i-k) + (1-p)P(i+1) + (1-p)P(0)\delta_{i-k}, \quad i \geq k. \end{aligned}$$

Let $\mathbf{Y}(z)$ be the z-transform of the steady state probability distribution of Y.

Theorem 5.2 $\mathbf{Y}(z)$ *is given by*

$$\mathbf{Y}(z) = \left(\frac{1-(k+1)p}{(k+1)}\right) \frac{1 - z^{k+1}}{pz^{k+1} - z + (1-p)}. \tag{23}$$

PROOF. We can find $\mathbf{Y}(z)$ as follows.

$$\sum_{i=1}^{\infty} P(i)z^i = pz^k \sum_{i=1}^{\infty} P(i-k)z^{i-k} + \frac{1}{z}\sum_{i=1}^{\infty}(1-p)P(i+1)z^{i+1} + z^k(1-p)P(0),$$

which implies

$$\mathbf{Y}(z) - P(0) = pz^k\mathbf{Y}(z) + z^{-1}(1-p)(\mathbf{Y}(z) - zP(1) - P(0)) + z^k(1-p)P(0),$$

and gives the relation

$$\mathbf{Y}(z)(z - pz^{k+1} - (1-p)) = P(0)(1-p)(z^{k+1} - 1).$$

Using $\mathbf{Y}(1) = 1$ and L'Hôpital's rule we get $P(0) = [1-(k+1)p]/[(k+1)(1-p)]$, and hence, we obtain $\mathbf{Y}(z)$ as given by (23). □

As before, we can obtain the distribution of Y, and hence that of W, by inverting the z-transform. We can also obtain the moments of W directly from $\mathbf{Y}(z)$.

6 Model (iv): Upper and Lower Bounds on the Window Size

The models analysed in the previous two sections assumed either a lower or an upper bound on the congestion window. Although approximate, models $(i)-(iii)$ are easy to evaluate and provide simple expressions to obtain the throughput of the connection. In this section, we present an analysis when the congestion window is bounded from above and from below. The recursive equation for this model is given by

$$W_{n+1} = \max(\min(A_n W_n, B_u), B_l), \tag{24}$$

where B_l and B_u are the lower and upper bounds, respectively. We make the assumptions (3.1), (4.1), (4.2), and (5.1).

Let $Y_n = \frac{\log[B_u]-\log[W_n]}{\log[\alpha]}$ be the transformation of W_n. Let $L = \frac{\log[B_u]-\log[B_l]}{\log[\alpha]}$ be the number of states of Y_n. The transition probabilities of Y_n are shown in Figure 4.

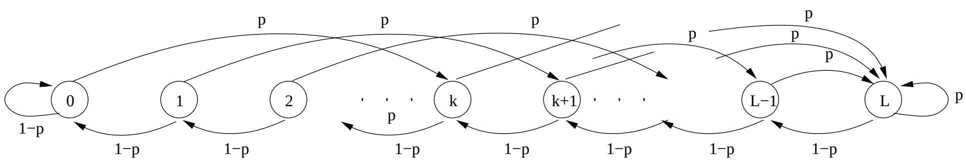


Figure 4. Transition probabilities of Y_n.

The recursive equation for the process $\{Y_n\}$ is given by

$$Y_{n+1} = \begin{cases} \min(Y_n + k, L) & \text{w.p. } p, \\ \max(Y_n - 1, 0) & \text{w.p. } 1-p, \end{cases}$$

Y_n is a finite state space communicating Markov chain with L states. Therefore, its steady state probabilities exist. The steady state probability, P_m, of being in state m satisfies the balance equation

$$\begin{aligned} P_m &= (1-p)P_{m+1} + pP_{m-k}, \quad m = 1, 2, ..., L-1, \\ P_0 &= \frac{1-p}{p} P_1. \end{aligned} \tag{25}$$

where we define for convenience $P_m = 0$ for $m < 0$.

Proposition 5 *Let $m = n(k+1) + j$, where $j = m \mod (k+1)$. The steady state probability, P_m, is given by*

$$P_m = P_0 \frac{p}{(1-p)^m} \cdot \sum_{i=0}^{n} (-1)^i \Big(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1} \Big) (1-p)^{ik}. \tag{26}$$

P_0 can be found using the equation $\sum_{m=0}^{L} P_m = 1$. The coefficients, $S_{i,j}$, are given by

$$S_{i,j} = S_{j,i} = \binom{i+j}{i} = \binom{i+j}{j}.$$

$S_{i,j}$ can also be calculated from the recursion

$$S_{i,j} = S_{i,j-1} + S_{i-1,j}.$$

The first few values of $S_{i,j}$ are given in the array below

$$\begin{array}{c|cccccc} (i,j) & -1 & 0 & 1 & 2 & 3 & ... \\ \hline -1 & 0 & 1 & 0 & 0 & 0 & ... \\ 0 & 0 & 1 & 1 & 1 & 1 & ... \\ 1 & 0 & 1 & 2 & 3 & 4 & ... \\ 2 & 0 & 1 & 3 & 6 & 10 & ... \\ 3 & 0 & 1 & 4 & 10 & 20 & ... \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \tag{27}$$

PROOF. It can be shown using induction that the steady state probabilities given by (26) satisfy the steady state equations in (25). The complete proof is given in Appendix A. □

The steady state probabilities of the window size, $P(W = j)$, can be obtained from the relation

$$P(W = j) = P\left(Y = \frac{\log[B_u] - \log[j]}{\log[\alpha]}\right).$$

We note that $\frac{\log[B_u]-\log[j]}{\log[\alpha]}$ is an integer since the state space of W, $\mathcal{W}$, is of the form $\mathcal{W} = \{B_l \cdot \alpha^i, i = 0, 1, ..., L\}$. The n^{th} moment of W can be obtained from

$$E[W^n] = \sum_{j \in \mathcal{W}} j^n P(W = j). \tag{28}$$

The computational cost for obtaining the moments using (28) is much higher than if we use either (18) or (22).

A similar expression can be derived when there is a congestion loss at B_u (i.e., model (iii) with lower bound on the window size). In this case, the recursion of (25) is modified as

$$\begin{aligned} P_m &= (1-p)P_{m+1} + pP_{m-k}, \quad m = k+1, k+2, ..., L-1, \\ P_i &= (1-p)^{k-i} P_k, \quad i = 0, 1, ..., k. \end{aligned} \tag{29}$$

It can be seen that equation (29) for $m \geq k + 1$ is same as equation (25) for $m \geq 1$. Therefore, we can use proposition 5 to obtain the probabilities $P_j, j \geq k+1$ as a function of P_k. Since the probabilities are obtained as a function of P_k, we first substitute $P_0 = P_k \frac{(1-p)}{p}$ in (26), and then we obtain

Proposition 6 *The steady state probabilities for the transition probabilities as given by equation (29) are given by*

$$
\begin{aligned}
P_m &= P_k \frac{1}{(1-p)^{m-k}} \cdot \sum_{i=0}^{n} (-1)^i \Big(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1} \Big) (1-p)^{ik}, \\
&\quad m = k+1, k+2, ..., L \\
P_m &= P_k (1-p)^{k-i}, \quad m = 0, 1, ..., k.
\end{aligned}
$$

P_k can be obtained from the normalizing condition $\sum_{i=0}^{L} P_i = 1$.

We note that in the above equations we compute the probabilities with respect to P_k unlike in (26) where the probabilities are computed relative to P_0.

7 Upper Bound on Window Size and Window Dependent Random Losses : An Approximation

In the models considered in the previous sections, the probability of a loss in an RTT was independent of the window size in that RTT. In this section, we consider a model in which the losses in an RTT depend on the window size in that RTT. Specifically, we assume that each packet is dropped (or, equivalently, is in error) with a constant probability q. As a consequence of this assumption, the probability of packet drops in an RTT is no longer independent of the window size in that RTT. First, we present the model with window dependent losses. Then we propose an approximation to this model which will enable us to compute the throughput in the window dependent model using the expression for throughput in the window independent model (model (ii)).

We make the following assumption.

Assumption 7.1 *In each RTT, the window is reduced only once even in the presence of multiple packet drops.*

We note that this assumption is consistent with loss recovery mechanisms of the recent TCP flavours such as New Reno and SACK.

Let W_n be the window size in the n^{th} RTT. Let p_n be the probability that the window is reduced in the n^{th} RTT. Then, p_n is given by

$$p_n = 1 - (1-q)^{W_n}. \tag{30}$$

The window size evolution for this model can be written as

$$W_{n+1} = \min(A_n W_n, B_u),$$

where B_u is the upper bound on the window size, and A_n is now given by

$$A_n = \begin{cases} \alpha & \text{w.p. } 1 - p_n, \\ \beta & \text{w.p. } p_n. \end{cases}$$

Next, we propose an approximation to the above model, and relate this approximation with model (*ii*).

For $q \cdot B_u << 1$, we can approximate p_n as

$$p_n \approx qW_n. \tag{31}$$

Using the approximation (31), the average window drop probability, $E[p]$, in an RTT is given by

$$E[p] = qE[W]. \tag{32}$$

We now substitute $E[p]$ as the probability of random loss in the expression for computing the average window size in the model (*ii*). From equation (22) with $n = 1$, we have

$$E[W] = B_u(1 - (k+1)p)\frac{1 - \alpha^{-1}}{p\alpha^{-(k+1)} - \alpha^{-1} + (1-p)},$$

and this together with (32) gives

$$E[W] = B_u(1 - (k+1)qE[W])\frac{1 - \alpha^{-1}}{qE[W]\alpha^{-(k+1)} - \alpha^{-1} + (1 - qE[W])}. \tag{33}$$

The above equation is a quadratic equation in $E[W]$, namely

$$c_2 E[W]^2 + c_1 E[W] + c_0 = 0, \tag{34}$$

where $c_2 = q\frac{1-\alpha^{-(k+1)}}{1-\alpha^{-1}}$, $c_1 = -(1 + (k+1)qB_u)$, and $c_0 = B_u$. Therefore, its roots can be explicitly written as

$$E[W]_{1,2} = \frac{-c_1 \pm \sqrt{c_1^2 - 4c_2c_0}}{2c_2}. \tag{35}$$

Proposition 7 *The solution of equation (34) which satisfies the inequality $E[W] \leq B_u$ is*

$$E[W] = \frac{-c_1 - \sqrt{c_1^2 - 4c_2c_0}}{2c_2}. \tag{36}$$

PROOF. Please see Appendix B. □

We can now obtain an approximate throughput of the session in the window dependent loss model by using (36). We note that this expression is an approximation and we shall compare this approximation with actual simulation results at the end of Section 8.

8 Simulation Results

Scalable TCP was proposed as a modification to the existing standard TCP for high speed networks. In the congestion avoidance phase, Scalable TCP uses the following algorithm to update the sender's window at the end of every RTT:

$$W_{n+1} = \begin{cases} 1.01 \times W_n & \text{if no losses are detected during the } n^{th} \text{ RTT,} \\ 0.875 \times W_n & \text{if one or more losses are detected during the } n^{th} \text{ RTT.} \end{cases}$$

As mentioned in the Introduction, Scalable TCP is an instance of MIMD protocols, and therefore, we validate our models by performing simulations with Scalable TCP. The simulation are performed using *ns-2* [15]. The simulation setup has a source and a destination node. The source node has infinite amount of data to send and uses Scalable TCP with New Reno flavor. The link bandwidth is 150 Mbps and the two way propagation delay is 120 ms. The window at the source is limited to 500 packets to emulate the receiver advertised window. The BDP for this system is approximately 2250 packets (packet size is 1040 bytes). In the Scalable TCP we have implemented in *ns-2*, the following assumptions are made:

- The minimum window size, B_l, is 8. The growth rate of Scalable TCP is very small for small window sizes. It has been recommended in [4] to use the Scalable algorithm after a certain threshold.
- There is no separate slow start phase since slow start can be viewed as a multiplicative increase algorithm with $\alpha = 2$.
- For each positive ACK received, the window is increased by $\alpha - 1$ packets. When a loss is detected, the window is reduced by a factor of β. α is taken as 1.01 and β is taken as 0.86. This value of β gives $k = -\frac{\log[\beta]]}{\log[\alpha]} \approx 15$. We set α and β in this way so as to be close to the values recommended in [4] ($\alpha = 1.01$, $\beta = 0.875$).

The expression for the density function of W, $f(w)$, and the moments of W modified for the minimum window at 8 is given by

$$f(w) = \frac{a}{8}\left(\frac{w}{8}\right)^{-(a+1)}, \tag{37}$$

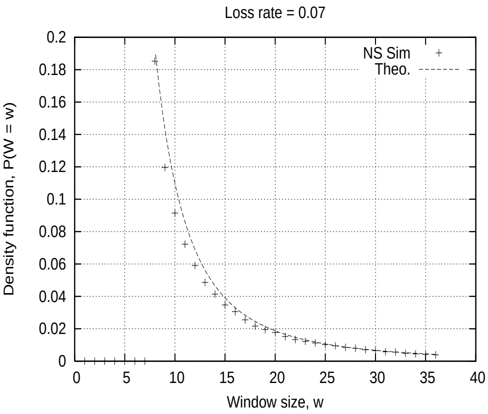


Figure 5. Density Function of the window size, W. $a = 1.55$

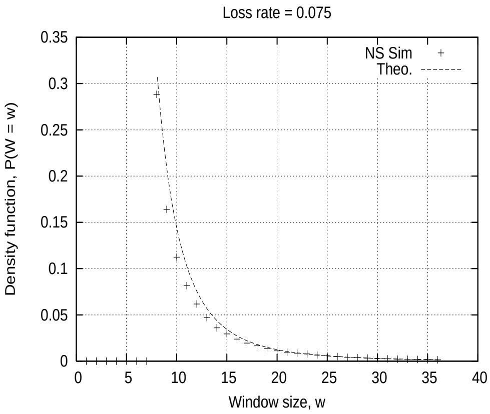


Figure 6. Density Function of the window size, W. $a = 2.53$

and

$$E[W^n] = 8^n \frac{a}{a-n},$$

respectively. In the simulations, the density function of W is obtained by sampling the window at an interval of $RTT = 0.12s$. We would like to note that the RTT is very close to the propagation delay in the present setting, and does not vary much.

Figures 5 and 6 show the PDF of W for two different values of loss rate, p. Simulation results are observed to match well with the analysis (equation (37)). Depending on the value of the root, z_0, of equation (14), the distribution can be seen to become heavy tailed. For example, for $p = 0.07$, the tail decreases at rate 1.55 indicating the heavy tailed nature of the window size. In the models which we considered, the window size was assumed to take rational values. In practice, however, the window size (or, strictly speaking, the number of packets in the network) takes only integer values. For example, when the window size is 8.5, the sender sends 8 packets. The density for the window size through simulations is, therefore, defined only at integer values whereas the theoretical plot is shown for real values. This results in a small discrepancy between the simulations and the theoretical function. Figure 7 shows the throughput in (TCP packets)/RTT as a function of the loss rate, p. The error bars are the 99% confidence intervals. Figure 8 shows the throughput in (TCP packets)/RTT as a function of the loss rate, p, for the model in which the maximum window at the sender is limited by the receiver's advertised window. The receiver buffer is assumed to be limited to 500 packets. The error bars are the 99% confidence intervals. A good match is observed between the simulations and the analysis.

In Figure 9, the throughput is plotted as a function of loss rate. The throughputs as obtained from models (i), (ii), and the exact model (model (iv)) are compared with the simulations. The vertical line $p = 1/(k+1)$ separates the

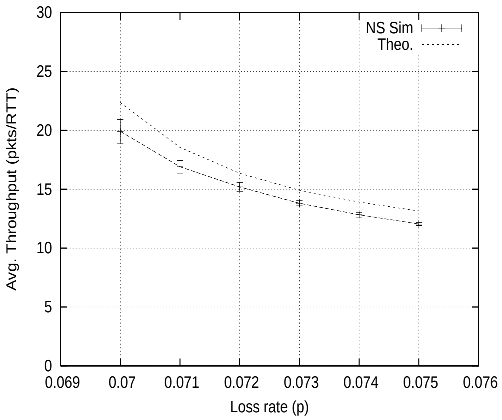


Figure 7. Throughput (pkts/RTT) versus loss rate, p.

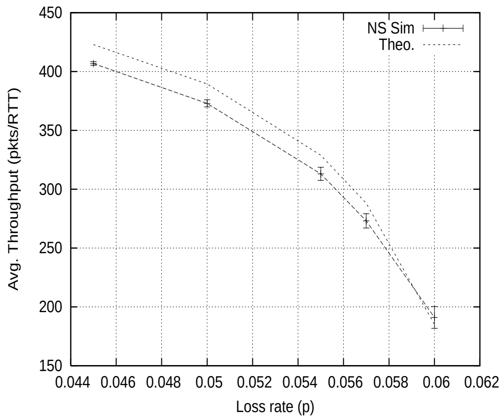


Figure 8. Throughput (pkts/RTT) versus loss rate, p.

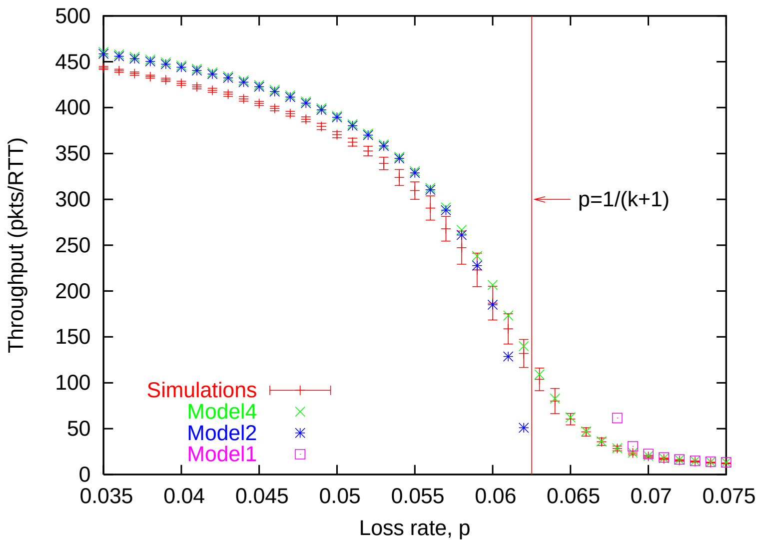
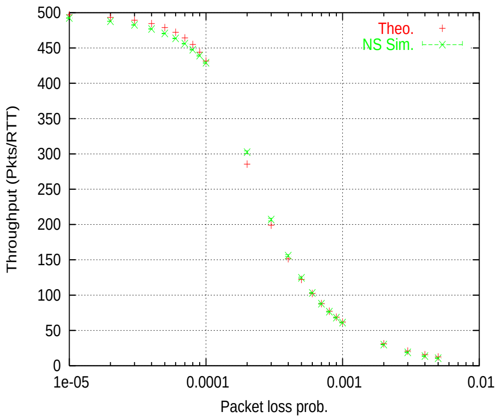


Figure 9. Throughput (pkts/RTT) versus loss rate, p.

two regions where model (i) and model (ii) are valid, respectively. As p approaches $1/(k+1)$ from either direction, the approximate models (i) and (ii) diverge from the simulation results. However, model (i) gives a good estimate when $(k+1)p >> 1$, i.e., $p >> 0.625$ ($k = 15$ in the simulations). Similarly, model (ii) gives a good approximation of the system when $p << 0.625$. The exact model fits well throughout the range of p. The throughput for model (i) is plotted for $p \geq 0.068$ because a (in equation (18)) is > 1 for $p \geq 0.0673$.

Approximation : Next, we compare the approximate throughput formula for the window dependent loss model as obtained using (36), with simulations. The simulation setup is as before. In Figures 10 and 11, the throughput is plotted as a function of the packet loss probability, q, for two different maximum window sizes, $B_u = 500$ and $B_u = 2000$. The approximation gives a good match over a large range of values of q. Although we had assumed $qB_u << 1$,



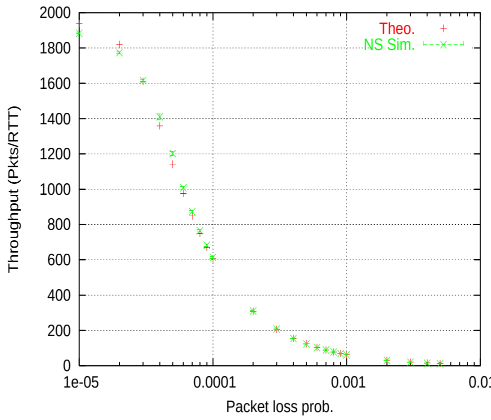


Figure 10. Throughput versus packet loss probability. $B_u = 500$.

Figure 11. Throughput versus packet loss probability. $B_u = 2000$.

the approximation seems to give a good match for larger values of q, too. As q increases, the expected window size, $E[W]$, and the probability of being near B_u decreases. The inequality, $qE[W] << 1$, still holds for larger values of q, and therefore the approximation seems to give a good match.

9 Conclusions

The logarithm of the window size process of a connection using the MIMD congestion control algorithm is equivalent to the workload process in a $G/G/1$ queue. The throughput of the connection and the higher moments of the window size process can be computed using the Laplace-Stieltjes transform of the equivalent workload process. For window independent losses, an exact expression can be obtained for the steady state probability distribution of the window size, and the throughput of the connection. For window dependent losses, an approximate expression, analogus to the square root formula for standard TCP, can be used to compute the throughput. This approximation is observed to be close to the actual throughput obtained from simulations.

10 Acknowledgments

We would to thank the an anonymous referee for helpful comments.

References

[1] S. Floyd. HighSpeed TCP for Large Congestion Windows. RFC 3649, Experimental, December 2003. Available at www.icir.org/floyd/hstcp.html.

[2] C. Jin, D. X. Wei, and S. H. Low. FAST TCP: Motivation, Architecture, Algorithms, Performance. In *Proceedings of the IEEE INFOCOM*, March 2004.

[3] L. Xu, K. Harfoush, and I. Rhee. Binary Increase Congestion Control (BIC) for Fast Long-Distance Network. In *Proceedings of the IEEE INFOCOM*, March 2004.

[4] T. Kelly. Scalable TCP: Improving Performance in Highspeed Wide Area Networks. *Computer Communication Review*, 33(2):83–91, April 2003.

[5] R. El Khoury and E. Altman. Analysis of Scalable TCP. In *Proceedings of the International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETS)*, July 2004.

[6] E. Altman, K. Avrachenkov, and C. Barakat. A Stochastic Model of TCP/IP with Stationary Random Losses. In *Proceedings of the ACM SIGCOMM*, September 2000.

[7] M. Vojnovic and J. Y. Le-Boudec. Some Observations on Equation-based Rate Control. In *Proceedings of the 17th International Teletraffic Congress*, September 2001.

[8] F. Baccelli and D. Hong. AIMD, Fairness and Fractal Scaling of TCP Traffic. In *Proceedings of the IEEE INFOCOM*, April 2002.

[9] D. Chiu and R. Jain. Analysis of the Increase/Decrease Algorithms for Congestion Avoidance in Computer Networks. *Journal of Computer Networks and ISDN*, 17(1):1–14, 1989.

[10] E. Altman, K. Avrachenkov, and B. J. Prabhu. Fairness in MIMD Congestion Control Algorithms. In *Proceedings of the IEEE INFOCOM*, 2005. Also available as a INRIA Technical Report RR-5312 at http://www.inria.fr/rrrt/rr-5312.html.

[11] L. Kleinrock. *Queueing Systems Volume I: Theory*. Wiley & sons, 1975.

[12] P. Glasserman and D.D. Yao. Stochastic Vector Difference Equations with Stationary Coefficients. *J. Appl. Prob.*, 1995.

[13] F. Baccelli and P. Bremaud. *Palm Probabilities and Stationary Queues*. Springer, 1987.

[14] S. G. Krantz. *Handbook of Complex Variables*. Birkhauser, 1999.

[15] S. McCanne and S. Floyd. ns: Network Simulator. Available at http://www.isi.edu/nsnam/ns/.

A Proof of Proposition 5

We first note that for $m = 1$, (26) gives $P_1 = P_0 \frac{p}{1-p}$,which is also obtained from (25). From (25), for $m \geq 1$, P_m's follow the recursion

$$P_{m+1} = \frac{P_m - pP_{m-k}}{1-p}. \tag{A.1}$$

We show that the probabilities given by (26) satisfy the above recursion. Let $m = n(k+1) + j$, where $j = m \mod (k+1)$. Let n denote the level of m. There are two cases : a) $j = k$, and b) $j < k$. When $j = k$, P_m and P_{m-k} are on the same level n. When $j < k$, P_m is on level n and P_{m-k} is on level $n-1$.

Case a): From (26),

$$\begin{aligned} P_{n(k+1)+k} = &\frac{P_k}{(1-p)^{n(k+1)}} \\ &\cdot \sum_{i=0}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+k} + p^i S_{i,(n-i)(k+1)+k-1}\right)(1-p)^{ik}, \\ P_{n(k+1)} = &\frac{P_k}{(1-p)^{n(k+1)-k}} \\ &\cdot \sum_{i=0}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)} + p^i S_{i,(n-i)(k+1)-1}\right)(1-p)^{ik}, \end{aligned}$$

Substituting in the RHS of (A.1), we get

$$
\begin{aligned}
P_{n(k+1)+k} - pP_{n(k+1)} = &\frac{P_k}{(1-p)^{n(k+1)}} \cdot \\
&\left[\sum_{i=0}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+k} + p^i S_{i,(n-i)(k+1)+k-1} \right) (1-p)^{ik} \right. \\
&\left. + \sum_{i=0}^{n} (-1)^{i+1} \left(p^i S_{i-1,(n-i)(k+1)} + p^{i+1} S_{i,(n-i)(k+1)-1} \right) (1-p)^{(i+1)k} \right], \\
= &\frac{P_k}{(1-p)^{n(k+1)}} \cdot \\
&\left[\left(p^{i-1} S_{i-1,(n-i)(k+1)+k} + p^i S_{i,(n-i)(k+1)+k-1} \right) (1-p)^{ik} |_{i=0} \right. \\
&+ \sum_{i=1}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+k} + p^i S_{i,(n-i)(k+1)+k-1} \right) (1-p)^{ik} \\
&+ \sum_{i=1}^{n} (-1)^i \left(p^{i-1} S_{i-2,(n-i+1)(k+1)} + p^i S_{i-1,(n-i+1)(k+1)-1} \right) (1-p)^{ik} \\
&\left. + (-1)^{n+1} \left(p^n S_{n-1,0} + p^{n+1} S_{n,-1} \right) (1-p)^{(n+1)k} \right],
\end{aligned}
$$

$$
\begin{aligned}
= &\frac{P_k}{(1-p)^{n(k+1)}} \cdot \left[\left(p^{i-1} S_{i-1,(n-i)(k+1)+k} + p^i S_{i,(n-i)(k+1)+k-1} \right) (1-p)^{ik} |_{i=0} \right. \\
&+ \sum_{i=1}^{n} (-1)^i \Big[p^{i-1} (S_{i-1,(n-i)(k+1)+k} + S_{i-2,(n-i)(k+1)+k+1}) \\
&+ p^i (S_{i,(n-i)(k+1)+k-1} + S_{i-1,(n-i)(k+1)+k}) \Big] (1-p)^{ik} \\
&\left. + (-1)^{n+1} \left(p^n S_{n-1,0} + p^{n+1} S_{n,-1} \right) (1-p)^{(n+1)k} \right],
\end{aligned}
$$

The coefficients $S_{i,j}$ follow the recursion

$$
S_{i,j} = S_{i,j-1} + S_{i-1,j}. \tag{A.2}
$$

Also, from (27), $S_{-1,j} = \delta_{0,j}, \forall j$, and $S_{0,j} = 1, \forall j \geq 0$. Therefore, for $i = 0$, we can substitute $S_{i-1,(n-i)(k+1)+k}$ by $S_{i-1,(n-i)(k+1)+k+1}$, and $S_{i,(n-i)(k+1)+k-1}$ by $S_{i,(n-i)(k+1)+k}$. Similarly, $S_{i,0} = 1, \forall i$, and $S_{i,-1} = 0, \forall i$. Therefore, we can substitute $S_{n,-1}$ by $S_{n+1,-1}$, and $S_{n-1,0}$ by $S_{n,0}$. The coefficients of p^{i-1} and p^i can be substituted by the recursion given in (A.2).

$$= \frac{P_k}{(1-p)^{n(k+1)}} \cdot$$

$$\Bigg[\left(p^{i-1} S_{i-1,(n-i)(k+1)+k+1} + p^i S_{i,(n-i)(k+1)+k}\right) (1-p)^{ik}|_{i=0}$$

$$+ \sum_{i=1}^{n} (-1)^i \left[p^{i-1} S_{i-1,(n-i)(k+1)+k+1} + p^i (S_{i,(n-i)(k+1)+k}\right] (1-p)^{ik}$$

$$+(-1)^{n+1} \left(p^n S_{n,0} + p^{n+1} S_{n+1,-1}\right) (1-p)^{(n+1)k} \Bigg],$$

$$= \frac{P_k}{(1-p)^{n(k+1)}} \cdot \sum_{i=0}^{n+1} (-1)^i \left[p^{i-1} S_{i-1,(n+1-i)(k+1)} + p^i S_{i,(n+1-i)(k+1)-1}\right] (1-p)^{ik}$$

$$= (1-p) P_{(n+1)(k+1)} = (1-p) P_{m+1}.$$

This proves case (a).

Case (b) : Since $j < k$, we can write $m = n(k+1)+j-k$ as $(n-1)(k+1)+j+1$. From (26),

$$P_m = P_{n(k+1)+j} = \frac{P_k}{(1-p)^{n(k+1)+j-k}}$$

$$\cdot \sum_{i=0}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1}\right) (1-p)^{ik},$$

$$P_{m-k} = P_{(n-1)(k+1)+j+1} = \frac{P_k}{(1-p)^{n(k+1)+j-k-k}} \cdot$$

$$\sum_{i=0}^{n-1} (-1)^i \left(p^{i-1} S_{i-1,(n-1-i)(k+1)+j+1} + p^i S_{i,(n-1-i)(k+1)+j}\right) (1-p)^{ik},$$

Therefore,

$$P_m - pP_{m-k} = \frac{P_k}{(1-p)^{n(k+1)+j-k}} \cdot$$

$$\Bigg[\sum_{i=0}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1}\right) (1-p)^{ik}$$

$$+ \sum_{i=0}^{n-1} (-1)^{i+1} \left(p^i S_{i-1,(n-1-i)(k+1)+j+1} + p^{i+1} S_{i,(n-1-i)(k+1)+j}\right) (1-p)^{(i+1)k} \Bigg],$$

$$
\begin{aligned}
&= \frac{P_k}{(1-p)^{n(k+1)+j-k}} \cdot \Bigg[\left(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1}\right) (1-p)^{ik}|_{i=0} \\
&+ \sum_{i=1}^{n} (-1)^i \left(p^{i-1} S_{i-1,(n-i)(k+1)+j} + p^i S_{i,(n-i)(k+1)+j-1}\right) (1-p)^{ik} \\
&+ \sum_{i=1}^{n} (-1)^i \left(p^{i-1} S_{i-2,(n-i)(k+1)+j+1} + p^i S_{i-1,(n-i)(k+1)+j}\right) (1-p)^{ik} \Bigg].
\end{aligned}
$$

Using substitutions similar to that in case (a) we get

$$
\begin{aligned}
&= \frac{P_k}{(1-p)^{n(k+1)+j-k}} \cdot \Bigg[\left(p^{i-1} S_{i-1,(n-i)(k+1)+j+1} + p^i S_{i,(n-i)(k+1)+j}\right) (1-p)^{ik}|_{i=0} \\
&+ \sum_{i=1}^{n} (-1)^i \left[p^{i-1} S_{i-1,(n-i)(k+1)+j+1} + p^i (S_{i,(n-i)(k+1)+j}\right] (1-p)^{ik} \Bigg] \\
&= \frac{P_k}{(1-p)^{n(k+1)+j-k}} \cdot \sum_{i=0}^{n} (-1)^i \left[p^{i-1} S_{i-1,(n-i)(k+1)+j+1} + p^i S_{i,(n-i)(k+1)+j}\right] (1-p)^{ik} \\
&= (1-p) P_{n(k+1)+j+1} = (1-p) P_{m+1}.
\end{aligned}
$$

This completes the proof. $\square$

B Proof of Proposition 7

Let $\theta = \frac{1-\alpha^{-(k+1)}}{1-\alpha^{-1}}$. Note that θ is less than 1, since α is greater than 1.

The discriminant is given by

$$
\begin{aligned}
c_1^2 - 4c_0c_2 &= (1 + (k+1)qB)^2 - 4Bq\theta \\
&= 1 + ((k+1)qB)^2 + (2(k+1) - 4\theta)qB \\
&> 0.
\end{aligned}
$$

The last inequality is obtained using the fact $k \geq 1$ and $\theta < 1$. Since the discriminant is positive, both the roots, $E[W]_{1,2}$, are real. Using the fact $c_1 < 0$ and $c_0, c_2 > 0$ along with Descartes' sign rule, we obtain that both the roots are real and positive.

We first show that $\frac{-c_1}{2c_2} > B_u$.

$$
\begin{aligned}
\frac{-c_1}{2c_2} &= \frac{1+(k+1)qB_u}{2q\theta} \\
&= \frac{1}{2q\theta} + \frac{(k+1)qB_u}{2q\theta}.
\end{aligned}
$$

The first term in the above equality is positive. In the second term, we use the fact $\theta < 1$ together with $k \geq 1$, to note that the coefficient of B_u is greater than 1. Therefore, we obtain $\frac{-c_1}{2c_2} > B_u$. Hence, $E[W]_1$ in (35) does not satisfy the constraint $E[W] \leq B_u$.

Next, we show that $E[W]_2$ satisfies the given constraint. We denote the dependence of $E[W]_2$ on q by $f(q)$.

$$
\begin{aligned}
\lim_{q\downarrow 0} f(q) &= \lim_{q\downarrow 0}\left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta}\right) - \sqrt{\left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta}\right)^2 - \frac{B_u}{q\theta}} \\
&= \lim_{q\downarrow 0}\left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta}\right) \\
&\quad - \sqrt{\left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta} - B_u\right)^2 + \left(\frac{(k+1)}{\theta} - 1\right)B_u^2} \\
&= \lim_{q\downarrow 0}\left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta}\right) - \left(\frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta} - B_u\right) \\
&= B_u. \qquad \text{(B.1)}
\end{aligned}
$$

Let $g(q)$ be defined as

$$
g(q) = \frac{1}{2q\theta} + \frac{(k+1)B_u}{2\theta}.
$$

The function $g(q)$ is a continuous and decreasing function of $q \in (0, 1]$. Let $h(q)$ be defined as

$$
h(q) = \sqrt{(g(q) - B_u)^2 + d_0},
$$

where $d_0 = \left(\frac{(k+1)}{\theta} - 1\right) B_u^2 > 0$. Since $d_0 > 0$, $h(q)$ is a positive and continuous function of $q \in (0, 1]$.

We can rewrite $f(q)$ as

$$
f(q) = g(q) - h(q). \qquad \text{(B.2)}
$$

Now consider

$$h'(q) = \frac{(g(q) - B_u)}{\sqrt{(g(q) - B_u)^2 + d_0}} g'(q), \qquad q \in (0, 1]$$

$$|h'(q)| = \left| \frac{(g(q) - B_u)}{\sqrt{(g(q) - B_u)^2 + d_0}} \right| |g'(q)|, \qquad q \in (0, 1].$$

Since $d_0 > 0$, we obtain

$$|h'(q)| < |g'(q)|. \tag{B.3}$$

Since $g(q)$ is a decreasing function of q, $g'(q) < 0$. We can rewrite (B.2) as

$$f'(q) = -(|g'(q)| + h'(q)). \tag{B.4}$$

Using inequality (B.3) in (B.4), we obtain

$$f'(q) < 0.$$

Therefore $f(q)$ is a decreasing function of $q \in (0, 1]$. Also, from (B.1), we have $\lim_{q \downarrow 0} f(q) = B_u$. Therefore, $f(q) < B_u, q \in (0, 1]$. Hence, $E[W]_2$ satisfies the constraint. □