Stochastic Modeling of TCP/IP over Random Loss Channels

Alhussein A. Abouzeid, Murat Azizoglu, Sumit Roy *

Department of Electrical Engineering, University of Washington,
Box 352500
Seattle, WA 98195-2500

Abstract. An analytical framework for modeling the performance of a single TCP session in the presence of random packet loss is presented that is based on a semi-Markov model for the window size evolution. The model predicts the throughput for LANs/WANs (low and high bandwidth-delay products) with good accuracy, as compared against simulation results with *ns* simulator. Generally, higher speed channels are found to be more vulnerable to random loss than slower channels, especially for moderate to high loss rates.

1 Introduction

TCP/IP has been designed for reliable (wired) networks in which packet losses occur primarily due to network congestion. TCP employs window-based end-to-end congestion avoidance [6] by sending an acknowledgment (ACK) back to the source for each successful packet. At all times, the source keeps a record of the number of unacknowledged packets that it has released into the network, called the *the window size*. The source detects a packet loss by either the non-arrival of a packet ACK within a certain time (maintained by a *timer*), or by the arrival of multiple ACKs with the same next expected packet number. A packet loss is interpreted as an indication of congestion, and the source responds by reducing its window size so as not to overload the network with packets. Modeling this dynamic behavior of congestion window size is key to analyzing TCP/IP throughput performance.

In some circumstances (e.g. networks with wireless links), packet losses occur randomly due to link effects than due to network congestion. While random packet loss on the Internet has been reported in [7], it was not taken into consideration in the design of TCP/IP congestion control. Previous research [2–4] has shown that random packet loss (which is not due to congestion) may severely decrease the throughput of TCP because TCP interprets random packet loss to be due to congestion and hence lowers the input data rate into the network, and consequently the throughput. In [2, 3], a discrete-time model for random packet loss was used in which any given packet is lost with probability q independent

* Author for all correspondence - Ph/FAX : (206) 221-5261/543-3842; e-mail: roy@ee.washington.edu

of all other packets implying a geometric distribution on the number of successful packets between consecutive loss events. In [4], packet loss is characterized by an inhomogeneous Poisson process and the steady-state distribution of the window size obtained under the assumption of infinite buffer size. In this work, we assume a continuous-time packet loss model governed by a general renewal process and incorporate (finite) buffer sizing impact on TCP performance.

Our basic system model assumes an infinite source that releases packets into a buffer of size B upon receiving ACKs from the destination. Packets are sent over a link with capacity is μ packets/second and a net delay of τ (propagation delay plus any other processing delays etc.). Define $T = \tau + 1/\mu$ to be the time between the start of transmission of a packet and the reception of an ACK for this packet. Then μT is the bandwidth-delay product and the ratio $\beta = \frac{B}{\mu T}$ is the buffer size normalized by the bandwidth-delay product.

2 Ideal Channels without Random Packet Loss

We first briefly review the operation of TCP-Reno (TCP-R) for ideal channels, and summarize the key results in [1, 3].

Denote $w_p = \mu T + B = \mu\tau + B + 1$, and note that when the window size reaches w_p, the bit pipe (the combination of the channel and the transmit buffer) is fully utilized. A further increase in window size at this stage causes buffer overflow, at which point the window size is halved and W_{th} is set to $w_p/2$. Let $t' = 0$ denote the time of establishment of the TCP session under consideration, and let $W(t')$ denote the congestion window size at time t'. Let n denote the number of packets acknowledged during a time interval t. The *deterministic* window size $W(t')$ evolution during a TCP session has been analyzed in [1, 3] and yield useful expressions that are summarized below (see Figure 1 and the original sources for details).

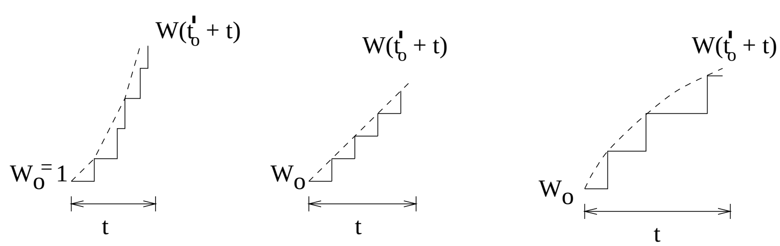

Fig. 1. Sketch of the exponential, linear and sub-linear $O(\sqrt{t})$ phases for window evolution. Solid lines indicate the actual window size evolution while dotted lines indicate the envelope.

1. **Slow Start** $(1 < W(t') < W_{th})$. Consider two instants t'_0, $t'_0 + t$ in a slow start phase of any of the TCP cycles. Choose t'_0 such that $W(t'_0) = 1$.Then,

$$W(t'_0 + t) = 2^{t/T} \tag{1}$$
$$n = W(t'_0 + t) - 1 \tag{2}$$

2. **Congestion Avoidance - Phase I** $(W_{th} < W(t') < \mu T)$. Consider two instants t'_0, $t'_0 + t$ in a congestion avoidance phase of any of the TCP cycles. Choose t'_0 such that $W(t'_0) = W_0$.Then,

$$W(t'_0 + t) = W_0 + t/T \tag{3}$$
$$n = \frac{1}{T}(W_0 t + t^2/(2T)) \tag{4}$$

3. **Congestion Avoidance - Phase II** $(\mu T < W(t') < w_p)$. Consider two instants t'_0, $t'_0 + t$ in a congestion avoidance phase of any of the TCP cycles. Choose t'_0 such that $W(t'_0) = W_0$.Then,

$$W(t'_0 + t) = \sqrt{W_0^2 + 2\mu t} \tag{5}$$
$$n = \mu t \tag{6}$$

It is apparent that the TCP window size evolution is periodic, i.e., consists of TCP 'cycles'. Using (1) - (6), the average packet transmission rate R is the ratio of the number of packets sent in one cycle of the TCP session to the time duration of the cycle, i.e.,

$$\beta < 1 : R = \tfrac{n_A + n_B}{t_A + t_B} \tag{7}$$
$$\beta > 1 : R \simeq \mu \tag{8}$$

and the corresponding average throughput is

$$\rho = \frac{R}{\mu} \tag{9}$$

The values for n_A, n_B, t_A and t_B are obtained by substituting for W_0 and $W(t')$ (see Figure 1) in (3)-(6) by the initial and final values of the slow start and congestion avoidance phases. Note the difference in the expressions for $\beta < 1$ and $\beta > 1$ - for $\beta < 1$, the window size evolution contains a linear growth phase described by (3) during congestion avoidance, which doesn't exist in the latter.

3 Channels with Random Packet Loss

3.1 Random Loss Model

Let S_i denote the time of the i^{th} packet loss, for $i = 1, 2, \ldots$ and $X_i = S_i - S_{i-1}$ the time between $(i-1)^{\text{th}}$ and i^{th} loss events with $X_1 = S_1$. We consider

$\{X_1, X_2, \ldots\}$ to be a set of IID random variables with probability density function $f(x)$ and distribution function $F(x)$. Thus, the process (pdf) defined by the loss occurrence times $\{S_1, S_2, \ldots\}$ is a renewal process with inter renewal pdf $f(x)$.

Now, suppose that at a certain time instant X_1 ($=S_1$), the first *random* packet loss event occurs. Denote the window size at that instant by W_1. When the source detects this loss (by the arrival of duplicate ACKs for the case of TCP-R), the window size is halved. The window size now increases as depicted earlier (the window size starts from $W_1/2$ and increases till w_p, at which time a buffer overflow takes place and $W(t')$ is set to $w_p/2$, and so on) until another random packet loss takes place at a random time instant $S_2 = X_1 + X_2$. Denote the window size at this time (the time of the second loss) by W_2.

In what follows, we call one period from $w_p/2$ till w_p the free-running period or the 'typical' cycle (i.e. free from random loss effects). Note that the second random loss event can happen before the occurrence of any 'typical' cycles. The window size $W(t')$ is a semi-Markovian stochastic process, because the window size evolution after a random loss (except for its starting value which is half of that just before the random loss) is statistically independent from the window size evolution before the random loss. Further, since $\{X_1, X_2, \ldots\}$ are independent and identically distributed (IID), the window sizes $\{W_1, W_2, \ldots\}$ (window sizes just before the random loss) form a finite state Markov Chain (i.e. the embedded Markov Chain of the semi-Markov process $W(t')$) [8].

3.2 Analysis

For the above model, we wish to compute the following quantities for the embedded Markov Chain
(1) $E[N|W_1 = w_1]$, the expected number of packets successfully transmitted before another random packet loss occurs, given that the most recent random loss took place at w_1; (2) The conditional probability $P[W_2 = w_2|W_1 = w_1]$ (denoted for convenience by P ; the probability that the next *random* loss takes place at $W_2 = w_2$ given that the previous *random* loss took place at $W_1 = w_1$.

To do this, we will ignore the first cycle of TCP-R and assume that the TCP session starts with window size $w_p/2$ (instead of 1) - this approximation should have a negligible effect on the average throughput since (i) a source with an infinite number of packets was assumed, hence the transient behavior (slow start) at the beginning of the connection is expected to be negligible, even for the case of random loss; and (ii) the duration as well as the number of packets sent during this slow start phase is small.

Two ranges of β are considered separately, $\beta < 1$ and $\beta > 1$, and expressions for $E[N|W_1]$ and $P[W_2|W_1]$ are found for each of the two ranges.

Define,
N_a, N_b: the number of packet transmissions during Congestion Avoidance Phase I and II, respectively, of the atypical cycle following a random packet loss at a window size $W_1 = w_1$.
N_A, N_B: the number of packet transmissions during Congestion Avoidance Phase

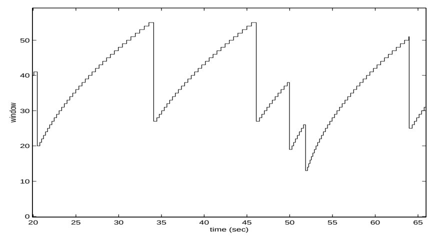

Fig. 2. Sample function of the window size evolution with random packet loss for a TCP-R session. $\mu = 100$, $\tau = 0.1$, $\beta = 4.0$ and $E[X] = 10$.

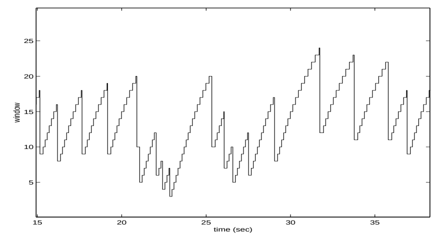

Fig. 3. Sample function of the window size evolution with random packet loss for a TCP-R session. $\mu = 100$, $\tau = 0.1$, $\beta = 4.0$ and $E[X] = 1$.

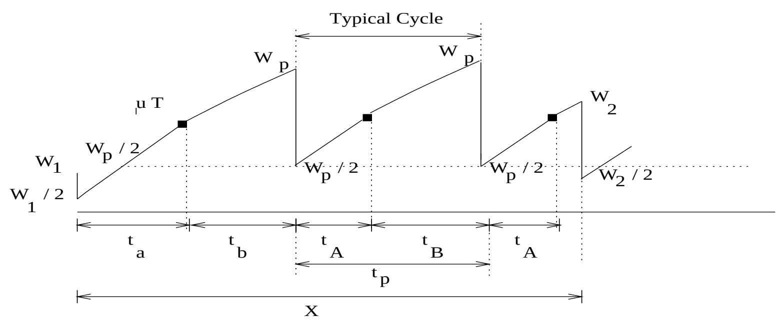

Fig. 4. A sketch of a sample function of window size with random loss for TCP-R ($\beta < 1$).

I and II, respectively, of a typical cycle.
$N_p = N_A + N_B$ is the number of packets sent in a typical cycle ($N_A = 0$ for $\beta > 1$).
The corresponding durations of time where the above number of packets is transmitted (time is counted since the beginning of the phase referenced) are t_a, t_b, t_A, and t_B respectively. Thus, $t_1 = t_a + t_b$, $t_p = t_A + t_B$ are the durations of the atypical and a typical cycle, respectively. For $\beta > 1$, $N_A = 0$ and $t_A = 0$.

Further details of the derivation and the results for a general inter-loss distribution $F_X(x)$ are contained in [10] and are ommitted due to space constraints. The analysis results for the case of $\beta < 1$ and $F_X(x) = e^{-\lambda x}$ are:

$$E[N|W_1 = w_1] = \sum_{n=0}^{n_a-1} e^{-\lambda(T/2)\sqrt{w_1^2+8(n+1)}-w_1} + \frac{e^{-\lambda t_1}}{1-e^{-\lambda t_p}} \sum_{n=0}^{n_A-1} e^{-\lambda(T/2)\sqrt{w_p^2+8(n+1)}-w_p} + e^{-\lambda/\mu}\frac{1-e^{-(\lambda/\mu)n_B}}{1-e^{-(\lambda/\mu)}}\left(e^{-\lambda t_a} + \frac{e^{-\lambda t_1}}{1-e^{-\lambda t_p}} e^{-\lambda t_A}\right) \quad (10)$$

$$P = \begin{cases} 0 & 0 < w_2 < w_1/2 \\ e^{-\lambda T(w_2 - w_1/2)}(1 - e^{-\lambda T}) & w_1/2 < w_2 < w_p/2 \\ e^{-\lambda T w_2}(1 - e^{-\lambda T})(e^{\lambda T w_1/2} + \frac{e^{-\lambda(t_1 - w_p/2)}}{1 - e^{-\lambda t_p}}) & w_p/2 < w_2 < \mu T \\ e^{\lambda \frac{(\mu T)^2}{2\mu}}(e^{-\lambda w_2^2} - e^{-\lambda(w_2+1)^2})(e^{-\lambda t_a} + \frac{e^{-\lambda(t_1 + t_A)}}{1 - e^{-\lambda t_p}}) & \mu T < w_2 < w_p \end{cases} \quad (11)$$

Finally, the average packet transmission rate R is computed from

$$R = \frac{E[N]}{E[X]} \quad (12)$$

where $E[N]$ is the average number of packets successfully sent in an inter-loss duration, and E[X] is the average time between two random losses ($= 1/\lambda$). $E[N]$ is given by

$$E[N] = \sum_{W=0}^{w_p} E[N|W]\pi(W) \quad (13)$$

where π is the steady state distribution of the MC. π is numerically computed using the eigensolver routines in MATLAB^{TM} for $P[W_2|W_1]$ for different values of λ, μ, τ and B.

4 Simulations Results and Concluding Remarks

In the packet level C code simulations, we considered the same set-up described in the system model in Section 1. The results from the analysis match closely the results from the simulations (Figure 5). Neglecting the slow start phase at the beginning of a TCP-R session in the analysis contributes in some deviation between the simulation and analysis results. For a given channel (i.e. bandwidth-delay

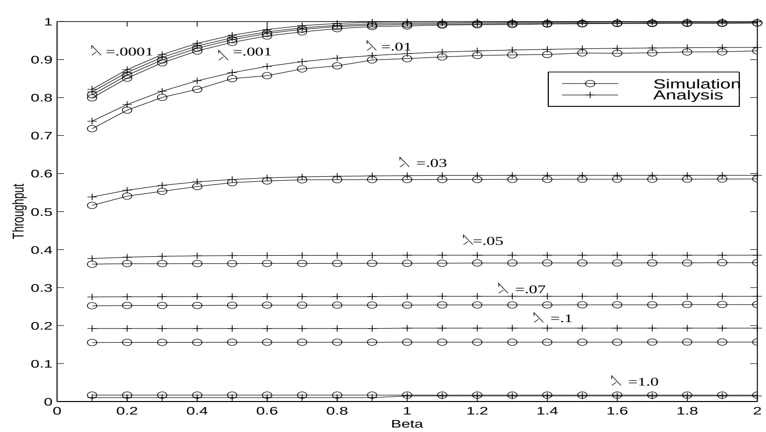

Fig. 5. $\mu = 100, \tau = 1.0$

product), the deviation (between the simulation and the analysis results) for low loss rates is small. This is because the slow start phase duration is sufficiently small such that the window size reaches $w_p/2$ in a very short time (compared to the average time to the first random loss) corroborating our approximation. in the analysis. As λ increases, so does the deviation since it becomes increasingly probable that the first random loss takes place early in the slow start phase, thereby precipitating a congestion avoidance phase with an initial window size that is considerably smaller than $\frac{w_p}{2}$ as assumed in the approximation. Consequently, the simulated throughput (on the average) is lower than that predicted by analysis, most noticeably for moderate values of random loss. For heavy loss rates, the deviation decreases again since the approximate window size quickly decreases from its starting value of $w_p/2$ to that (i.e., the true) in the simulations.

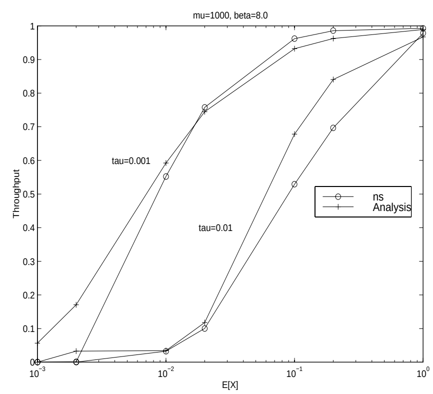

Fig. 6. Comparison between analysis and *ns* results for memoryless channels.

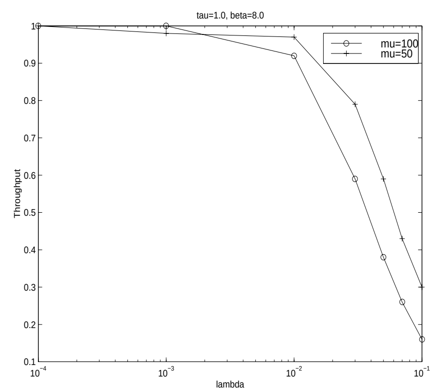

Fig. 7. Throughput comparison for two links with different speeds.

The analytical results based on the proposed random loss model matches with the *ns* simulation results as shown in Figure 6. The *ns* simulations are done using a two node topology and default TCP-R parameters (packet size = 1000 bytes, unlimited receiver's advertised window and $Tcptick = 0.01$).

The main conclusions that can be deduced from the throughput behavior in Figures 5, 6 and 7 are summarized below:

(1) For a link with a loss rate λ and a bandwidth-delay product $\mu\tau$, the results show that increasing the buffer size (i.e. increasing β) does *not* always increase the throughput. For channels with high loss rate, increasing the buffer size has no positive effect on the throughput; however for channels with low loss rates, increasing the buffer size increases the throughput considerably.

(2) For low loss rates, faster channels (higher μ) have higher throughput. However (contrary to what may be expected) for moderate to high loss rates, slower channels have higher throughput. The explanation for this is simple

though perhaps not transparent. Recall that for channels without random loss, the throughput is given by $\frac{n_p/t_p}{\mu}$. For channels with random loss, the throughput is given by $\frac{\lambda E[N]}{\mu}$. The expression in the numerator is the average transmission rate. Now, for the case of no random loss, increasing μ increases n_p significantly and hence the average transmission rate as well as the throughput increase. Similarly, for low random loss rates, increasing μ increases $E[N]$ significantly and hence both average rate and throughput increase. On the other hand, for moderate to high loss rates, increasing μ does not increase the number of packets successfully transmitted proportionately (due to the effect of random loss); hence the average transmi ssion rate increases but the throughput actually decreases.

One practical interpretation of this result for the Internet relates to a user's dial-up modem connection to a server. Purchasing a faster modem would increase the average transmission rate, but may not be economically justifiable in the case of moderate-to-high loss rate channels since the proportion of the used bandwidth (i.e, throughput) for the new faster modem is less than that for the slower (and hence, less expensive) one.

References

1. S. Shenker, L. Zhang, and D. D. Clark, "Some observations on the dynamics of a congestion control algorithm," *Computer Communications Review*, pp. 30-39, Oct 1990.
2. T. J. Ott, J. H. B. Kemperman, and M. Mathis, "The stationary behavior of ideal TCP congestion avoidance," 1996.
3. T. V. Lakshman, and U. Madhow, "The performance of TCP/IP for networks with high bandwidth-delay products and Random loss," *IEEE/ACM Transactions on Networking*, vol. 5, no. 3, June 1997.
4. S. Savari, and E. Telatar, "The behavior of certain stochastic processes arising in window protocols," *preprint*, July 1998.
5. A. Kumar, "Comparative Performance Analysis of Versions of TCP in a Local Network with a Lossy Link," *IEE/ACM Transactions on Networking*, vol. 6, no. 4, August 1998.
6. V. Jacobson, "Congestion avoidance and control," *Proceedings of ACM SIGCOMM ' 88*, August 1988.
7. J. Bolot, "End-to-end packet delay and loss behavior in the internet," *Proc. ACM SIGCOMM'93*.
8. R. G. Gallager, "Discrete stochastic processes," Boston : Kluwer, 1996.
9. S. Floyd, "TCP and successive fast retransmits," February 1995, *URL ftp://ftp.ee.lbl.gov/ email/vanj.90apr30.txt*.
10. A. Abouzeid, S. Roy and M. Azizoglu, "Markovian Analysis of TCP/IP Congestion Control for Channels with Random Loss," *submitted to ACM/IEEE Trans. on Networking*, 1999.