Информационные процессы, Том 2, № 2, 2002, стр. 149–154.

KALASHNIKOV MEMORIAL SEMINAR

Markovian Model of Internetworking Flow Control

O. Bogoiavlenskaia

Petrozavodsk State University
Lenin St., 33, Petrozavodsk, 185640, Russia
email: olbgvl@cs.karelia.ru

Received October 14, 2002

1. GENERAL OVERVIEW

The Internet transport layer presently plays a key role in the internetworking traffic control. The transport layer protocols form the network connections behavior at the end-to-end level. Therefore, the transport layer provides a decisive contribution in internetworking reliability, stability and performance. The family of transport layer protocols presently use a wide selection of flow and congestion control algorithms. Nevertheless, the Additive Increase Multiplicative Decrease (AIMD) algorithm plays a key role among them. We investigate a throughput provided by the AIMD algorithm. Our evaluation comes from a sender's point of view to understand the rate that a sender injects segments to the network under AIMD control.

As a result of our investigation, we have developed a Markovian model of AIMD sliding window size, cwnd, and throughput, T_{CA}. Our new model yields new important results concerning AIMD behavior. We obtained the steady state distributions of the AIMD cwnd and T_{CA}. This provides complete information of AIMD behavior in a particular networking environment.

Several recent publications (the most highly referenced among them are [1], [2]) present estimations of T_{CA} expectation or its bounds. All of them have restrictions on their applicability that will vary depending on the properties of the networking environment. Our model can establish bounds for using of algebraic estimations.

Our model incorporates the probabilistic nature of the networking environment. It provides a new, powerful tool for planning and designing of the AIMD based transport layer, including issues of QoS. Our approach relaxes several important restrictions accepted in most of analytical AIMD models. Round trip time (RTT) is considered to be a random variable described by its distribution function. The model of cwnd size and throughput never exceeds their natural limits, which are parameters of the model. The model is valid for high and low bandwidth links, since it does not use "round" modeling (see details in [2]). RTT and segment loss probability may (or may not) depend on the congestion window size.

This features make the approach highly flexible and applicable to a wide range of networking environments. All results, including the distributions mentioned above, are obtained in explicit analytical form. We have developed an algorithm that optimizes calculation of the cwnd distribution and computes it with linear complexity. Our analytical results are validated by using observations of experimental connections in an emulated network environment. In addition, we compare our model with other models of AIMD [1,2]. T_{CA} expectation provided by the model demonstrates high stability and good agreement with the experimental data.

2. MODELING ASSUMPTIONS

We only consider the AIMD portion of transport layer flow control. That is, the cwnd increases by one segment per RTT if no segment is lost. When the loss of a segment is detected at the sender, then cwnd is reduced by one-half.

Segment losses occur according to a segment loss pattern. The segment loss pattern is defined by the distribution of the number of segments sent in succession between two consequent loss indications. Our assumptions allow any memoryless segment loss pattern. For example, the Bernoulli scheme (or Bernoulli segment loss pattern) is memoryless.

We also include in our model the sender's link capacity as a parameter. Since the configuration of the end-to-end path as well as the capacity available to the connection are assigned randomly, they are difficult to be predicted. Throughput cannot be higher than the sender's link capacity, which is a natural limit. This is a very rough estimation. If one can derive the bottleneck capacity of the end-to-end path, then it may replace the sender's link capacity.

The maximal window size is restricted by some given value. It may be the receiver advertised window size (rwin) or any other restriction, as well.

From this set of assumptions we derive the distribution of cwnd and AIMD throughput as functions of the segment loss pattern, the RTT distribution function, the maximal window size and the sender's link capacity.

There are several restrictions accepted in most T_{CA} models that *are not included* in our model. The most important one is "round" modeling [1, 2] since it affects the limiting behavior of the resulting formulae (i.e., the T_{CA} thus computed tends to infinity for small loss probabilities even for a deterministic segment loss pattern). This means that under "round" model protocol sends one window during one RTT. Let us suppose that each RTT is equal to a time unit. If there is no loses cwnd increases by 1 each round. If it starts from $cwnd = 1$ then during n rounds protocol sends

$$1 + 2 + \cdots + n = \frac{1}{2}\left(n^2 + n\right)$$

segments. Thus the number of data sent becomes a quadratic function of time if there is no upper limits of cwnd and throughput. This relation requires infinite network capacity. Actual number of data sent must be linear or smaller then some linear function of time.

3. THE MODEL

Let us consider the pair $\xi = (w, n)$, where w is a current cwnd size and n is the number of the segments sent under this cwnd. For each w, $n = 1, \ldots, w$. Let us denote $w(t)$ and $n(t)$ the values that define ξ at the moment $t > 0$. Then according to our assumptions $\nu(t) = \{(w(t), n(t))\}_{t>0}$ is a semi-Markov random process (SMP).

Let τ_i be the time when ith segment is sent. The segments are numbered according to the following rule: $i_1 > i_2$ if $\tau_{i_1} > \tau_{i_2}$. Then sequence $\xi(\tau_0), \xi(\tau_1), \ldots, \xi(\tau_n), \ldots$ forms a Markov chain embedded in the process $\nu(t)$. The space of states of the chain $\{\xi_i\} = \xi(\tau_i)$ is finite. Let us denote the space X. We also denote $p^k_{\xi\eta}$ the probability of the transition from state ξ into state η by k steps for any $\xi, \eta \in X$. Since set X is finite then

$$p^k_{\xi\eta} \xrightarrow[k\to\infty]{} \pi_\eta$$

and

$$\sum_{\xi\in X} \pi_\xi = 1.$$

Let $P_\xi(t)$ denote the probability of $\nu(t) = \xi$ and α_ξ be the expectation of the segments inter-departure time. The following theorem takes place .

Theorem 1. *If RTT has a finite expectation, then*

$$P_\xi(t) \xrightarrow[t\to\infty]{} \frac{\alpha_\xi \pi_\xi}{\sum\limits_{\xi\in X} \alpha_\xi \pi_\xi}. \tag{1}$$

The proof is based on the Smith renewal theorem.

Values π_ξ obviously are solution of the Kolmogorov equations for the Markov chain $\{\xi_i\}$. The equations are complicated. Therefore, we consider the time moments when the sender assigns the new cwnd. Thus, the chosen sequence of cwnd sizes ξ' also forms a Markov chain.

The chain $\{\xi'_i\}$ is embedded in the chain $\{\xi_i\}$. The solution of the Kolmogorov equations for the chain $\{\xi'_i\}$ gives us the distribution π'_ξ for all pairs $\xi' = (w, 1)$, i.e. $\pi'_\xi = \pi_w$. The following theorem on π_w takes place.

Theorem 2. *The distribution π_w satisfies following relations*

$$\pi_i = \pi_j K_i,$$

where

$$K_{w_{max}} = \frac{F_{w_{max}-1}}{1 - f_{w_{max}}}, \tag{2}$$

$$K_i = F_{i-1}, \qquad j < i < w_{max}, \tag{3}$$

and

$$K_{i-1} = \frac{1}{f_{i-1}} \left(K_i - (K_{2i}(1 - f_{2i}) + K_{2i+1}(1 - f_{2i+1}))\right) \qquad i < j. \tag{4}$$

Here $j = \lfloor \frac{w_{max}}{2} \rfloor$, $f_{w_{max}}$ and $F_{w_{max}}$ are functions of segment loss probability.

This recurrent presentation for the distribution π_w is the base for the numerical algorithm we have developed. It optimizes calculation of π_w and π_ξ. The algorithm has a linear complexity. The values α_ξ are estimated using normal distribution according to the central limit theorem.

Note that random process $\mu(t) = \{w(t)\}_{t>0}$ also is SMP, and ξ'_i is a Markov chain embedded in it. Let U_w be the duration of the round. Then

$$U_w = \begin{cases} RTT, & \text{if} \quad wt_0 < RTT \\ wt_0, & \text{otherwise} \end{cases} \tag{5}$$

Let us denote β_w the expectation of U_w. Also let $P_w(t)$ be the probability of $\mu(t) = w$. Then the following theorem on the SMP $\mu(t)$ takes place.

Theorem 3. *If RTT has a finite expectation, then*

$$P_w(t) \xrightarrow[t\to\infty]{} \frac{\beta_w \pi_w}{\sum\limits_{w=2}^{w_{max}} \beta_w \pi_w}. \tag{6}$$

Let us denote $p_{w,n} = p_\xi = \lim\limits_{t\to\infty} P_\xi(t)$. The distribution of the AIMD window size is

$$\omega_i = \sum_{n=1}^{i} p_{i,n} \tag{7}$$

Obviously, $\omega_w = \lim\limits_{t\to\infty} P_w(t)$.

Let t_0 be the time that the sender needs to inject one segment to the network. T_{CA} depends on the relation between wt_0 and RTT. If $wt_0 > RTT$ then T_{CA} is equal to the sender's link capacity. If $wt_0 \leq RTT$ then T_{CA} is $T = w/RTT$.

Let us consider the second case in detail. We denote $R_w(x)$ as the distribution function of RTT (RTT depends on cwnd). Since $T = w/RTT$, we must consider all those w and RTT such that their relation gives T_0 and sum their joint probabilities to calculate the probability of T_{CA} at the particular value. Therefore the throughput distribution function is defined as

$$T(x) = \sum_{i=2}^{w_{max}} \omega_i \left(1 - R_w\left(\frac{i}{x}\right)\right) = 1 - \sum_{i=2}^{w_{max}} \omega_i R_w\left(\frac{i}{x}\right). \tag{8}$$

Formula (8) is true if $x < L$, where $L = 1/t_0$ is the sender's link capacity and $T(L) = 1$.

4. NUMERICAL EXAMPLES.

Several numerical examples here demonstrate the scope of the model. Figure 1 presents the cwnd frequencies for receiver advertised window $w_{max} = 120$ segments and segment loss probability p. The frequencies are obtained using the algorithm based on the Theorem 2.

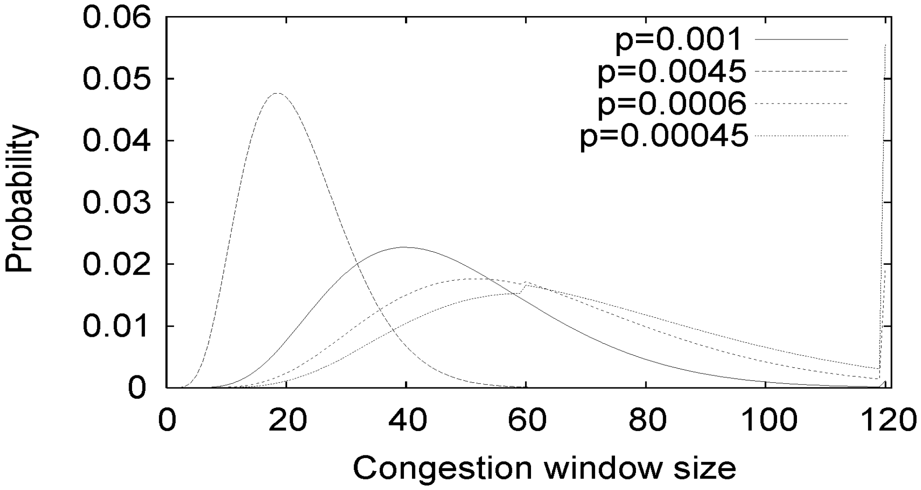

Figure 1. Frequencies of the congestion window sizes.

The figure demonstrates three basic types of the cwnd behavior. For small values of p, cwnd is likely to reach w_{max} and stay there for a long time. This creates a peak at w_{max}. If a segment is lost at cwnd=w_{max}, then cwnd is halved and this creates another smaller peak of the distribution at $w_{max}/2$. With larger values of p, cwnd demonstrates unstable behavior and fluctuates around a large range of sizes. Therefore the distribution loses its maximum (local or global) at w_{max}.

Figure 2 shows the T_{CA} expectation for receiver advertised window $w_{max} = 70$ with different sender's link capacities. Note that for $p > 0.03$, $rwin$ (parameter w_{max}) does not significantly affect T_{CA} since large windows are never reached. The presented model of T_{CA} never exceeds the given link capacity. Our results show whether T_{CA} utilizes the link capacity for the given combination of w_{max} and RTT distribution.

The deviation of T_{CA} plays a significant role as it shows the stability of a connection and may be crucial for its performance. Figure 3 presents standard deviation of T_{CA} as a function of the segment loss probability. It has a maximum between $p = 0$ and $p = 0.05$. Obviously the standard deviation of cwnd must be zero for $p \approx 1$ and $p = 0$ and it is nonzero between these two points. Therefore, it has at least one maximum as a function of the segment loss probability. The maximum corresponds to the case of cwnd fluctuating around a wide range of values. The standard deviation of T_{CA} in many cases inherits this property. Our model provides not only qualitative but quantitative analysis, as well. Figure 3 demonstrates how does RTT deviation affect T_{CA} deviation.

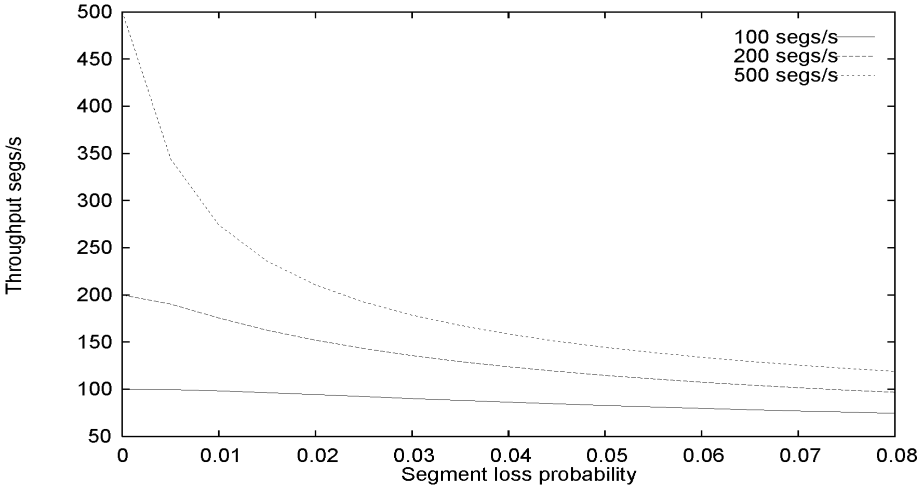

Figure 2. Expectation of T_{CA} for different sender's link capacities. $w_{max} = 70$ segments

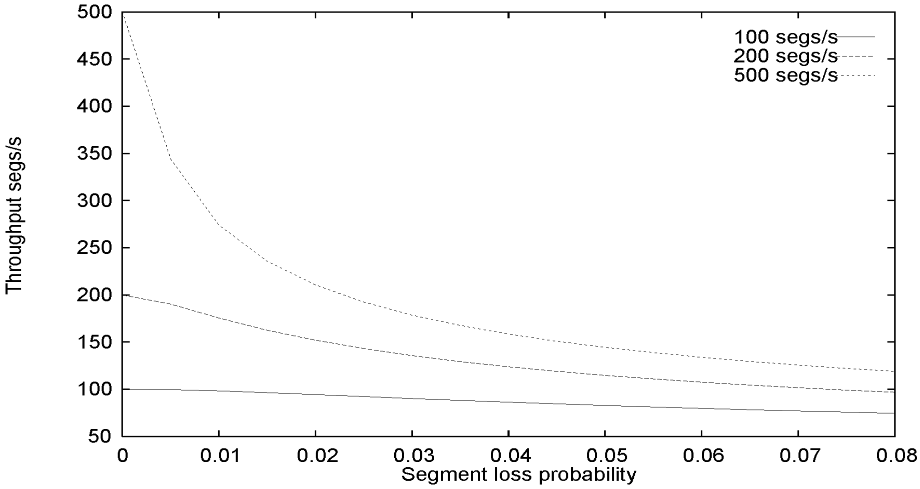

Figure 3. Standard deviation of T_{CA} for different RTT standard deviation. $w_{max} = 30$ segments, sender's link capacity 300 segs/s

Table 1 provides comparison between estimation of T_{CA} expectation presented in [1] (FF) and our model. ET in the table marks values of T_{CA} expectation defined by distribution (8). The table contains T_{CA} expectation values for receiver advertised window sizes 10, 40 and 80 segments and maximum link capacity 115 segs/sec. RTT satisfies normal pdf. Estimation FF grows unrestrictedly while segment loss probability tends to zero. It exceeds link capacity if segment loss probability is smaller than 0.0001. Hence in the case the maximal capacity is better estimation than FF.

5. ACKNOWLEDGMENTS

I express my gratitude to Prof. T. Alanko, Prof. M. Mutka, M. Kojo and P. Salorahti for fruitful multiple discussions and friendly cooperation.

p	FF	ET, w_{max}=10	ET, w_{max}=40	ET, w_{max}=80
0.9	1.90	3.72	3.72	3.72
0.5	2.55	3.73	3.73	3.73
0.1	5.70	7.21	7.22	7.22
0.05	8.07	9.99	10.33	10.33
0.01	18.04	14.79	24.01	24.01
$5 * 10^{-3}$	25.51	15.56	34.06	34.31
10^{-3}	57.04	16.17	55.49	71.50
$5 * 10^{-4}$	80.67	16.26	59.60	87.55
10^{-4}	**180.38**	16.33	63.00	103.79
10^{-5}	**570.41**	16.34	63.76	107.46
10^{-6}	**1804.00**	16.34	63.84	107.76
10^{-7}	**5704.00**	16.34	63.85	107.79
10^{-8}	**18040.00**	16.34	63.85	107.80

Table 1. Models comparison (Absolute values). Link rate is **115** segs/sec. Values exceeding the bandwidth capacity are marked by bold font.

REFERENCES

1. Floyd S., Fall F. Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Transactions on Networking, 1999, 7(4), pp. 458-472.

2. Padhey J., Firoiu V., Towsley D., Kurose J. Modeling TCP Throughput: A Simple Model and its Empirical Validation, IEEE/ACM Transactions on Networking 2001, 8(2), pp. 133-145.