Performance Analysis and Stochastic Stability of Congestion Control Protocols

E. Altman K. E. Avrachenkov A. A. Kherani B. J. Prabhu
INRIA-Sophia Antipolis
2004, route des Lucioles, 06902 Sophia-Antipolis, France
E-mail:{altman, k.avrachenkov, alam, bprabhu}@sop.inria.fr

Abstract— **We study an Adaptive Window Protocol (AWP) with a general increase and decrease profile in the presence of window dependent random losses. We derive a steady-state Kolmogorov equation and obtain its solution in analytic form. We obtain some stochastic ordering relations for a protocol with different bounds on window. A closed form necessary and sufficient stability condition using the stochastic ordering for the window process is established. Finally, we apply the general results to particular TCP versions such as New Reno TCP, Scalable TCP and HighSpeed TCP. We observe that HighSpeed TCP can be used to approximate almost any kind of window behavior by varying only one design parameter.**

I. INTRODUCTION

Most of the performance studies of Adaptive Window Protocols (AWP) consider specific instances of the problem (for example [3], [2], [6] study Additive Increase Multiplicative Decrease (AIMD) protocols). However, various modifications to TCP are frequently proposed to address specific problems arising in various types of networks; recent examples include HighSpeed TCP [5] and Scalable TCP [4] proposed for very high bandwidth-delay product networks. These new proposals can also be viewed in the framework of Additive Increase protocols so that now the additive increase in a round-trip time is function of the current window size (it is constant in the case of standard TCP). Performance related analysis of any such protocol has always been an important issue. It is thus desirable to have a general framework (and its solution) for performance analysis of an AWP.

The loss process seen by a TCP sender may have its origin in deliberate marking/dropping owing to some active queue management (AQM) scheme employed in the network, or could be due to congestion losses or link errors, in general the rate of receiving a loss signal will depend on the window process itself (see [6] for related discussion). In this study, we consider a general state dependent loss rate.

It is clear that the stability of window process of a general AWP will depend on the rate at which it receives loss signals. For example, an aggressive protocol may result in very high windows for moderate loss rates and vice versa. Stability of the window process is thus interesting to study. We address the problem of finding conditions for stability of a general AWP controlled window evolution under a general state-dependent loss rate. The contributions (and organization) of this work is as follows:

Section II: We give a characterization of a general AWP and identify the various quantities that determine the performance of such protocols. The window evolution under a general AWP is mapped to that under an AWP with a linear increase profile (like in standard TCP). Kolmogorov equations satisfied by the stationary probability measure is then derived.

Section III: Gives conditions under which two AWPs have related stationary distribution. Furthermore, we demonstrate that the window process under multiplicative decrease protocol is also related to the workload process in a queueing system.

Section IV: We give a general methodology for analysis of any such protocol while allowing for a general window dependent drop rate. The stationary distribution of a general AWP with a general loss rate is related to that of an AWP with linear increase profile and a constant loss rate. This observation is important as the latter system is easier to analyse. *An analytic expression for the stationary distribution of the system with an AWP with linear increase profile and a constant loss rate is provided.*

Section V and VI: We apply the results of Section IV to study the performance of recently proposed TCP modifications (Scalable TCP [4] and HighSpeed TCP [5]). We also refine an existing result on the standard AIMD protocol of TCP. We obtain some results of independent interest of queueing systems theory by relating the window process under a multiplicative increase multiplicative decrease (MIMD) protocol to the workload process in an M/D/1 queue. This provides us with a closed form expression for the workload process in an M/D/1 queue with bounded workload process. We also obtain a *duality* relation between the customer averages in a D/M/1 queue and time average in an M/D/1 queue (both queues with bounded workload capacity).

Section VII: We obtain some stochastic ordering relations for a protocol with different bounds on window. A closed form necessary and sufficient stability condition using the stochastic ordering for the window process is established.

The proofs of all the results of this paper can be found in [13]. The report [13] also contains some additional results. Since the paper addresses many issues, for sake of making clear the context of discussion, we decided to spread the discussion on related literature across the paper instead of mentioning them together. An extensive literature survey on TCP modeling can be found in [6].

II. THE MODEL

We consider an AWP controlled persistent file transfer over an Internet (bottleneck) link. For applications using HighSpeed and Scalable TCP, this link will typically be a very high bandwidth-delay product link. We assume that the connection is long enough to see a stationary regime and that its throughput performance is governed by the steady state regime (see [3] for justification of this assumption). Required conditions for existence of such a regime are provided in a later section. Recent applications using HighSpeed TCP and Scalable TCP typically transfer very large volume files and hence studying persistent transfers is justified in such cases. We model the process of losses as a Poisson process with a time varying intensity that depends on the instantaneous window size of the AWP [6]. These losses could be owing to congestion losses, random link losses or some deliberate packet marking/dropping by the router buffer using an AQM.

As is common in related studies ([2], [3], [6]), we consider the evolution of window as an infinitely divisible fluid. Let x_t denote the window size of the AWP at time instant t (note that we are not specifying the initial window size x_0, thus assuming a stationary window process). Incase of no loss in the time interval $[t, t+\Delta]$, the window increase is given by,

$$x_{t+\Delta} = x_t + f(x_t)\Delta + o(\Delta), \tag{1}$$

where $f(\cdot)$ is a Lipschitz continuous function bounded below by some positive quantity. We also assume that there is a lower bound on the window size, denoted x_{min}.

The increase in window size can not continue for ever because drops owing to congestion or channel losses or AQM marking can occur at random instants in time[1]. Let $N(t)$ be the counting process corresponding to the loss events, i.e., $N(t) - N(t-u)$ is the number of losses in time interval $(t-u, t]$. In what follows, we assume that $N(t)$ is a Poisson process with time varying intensity. Further, we assume that the instantaneous rate of the $N(t)$ process depends only on the current window size x_t of the connection. Let $\lambda(x)$ be the rate of $N(t)$ process when window size $x_t = x$. The assumptions imply that $P\{N(t+\Delta)-N(t)=1\}=1-P\{N(t+\Delta)-N(t)=0\}=\lambda(x_t)\Delta+o(\Delta)$. Each loss results in a window reduction (this is because TCP assumes that each packet drop/mark corresponds to a congestion event in the network). Under the fluid model, it is standard to assume that this window reduction is reflected as an instantaneous jump in the x_t process. Thus for small Δ, if $N(t+\Delta) - N(t) = 1$, the window is instantaneously reduced as

$$x_{t+\Delta} = g(x_t) + o(\Delta), \tag{2}$$

for some function $g(\cdot)$ such that $g(x) < x$ and $g(x_{min}) = x_{min}$. Additionally, $g(\cdot)$ is such that if $x_1 < x_2$ then either $g(x_1) < g(x_2)$ or $g(x_1) = g(x_2) = x_{min}$. This implies that the set $s(x) = \{u \geq x : g(u) \leq x\}$ is connected. Define

$$h(x) = \sup\{u \geq x : g(u) \leq x\} = \sup s(x).$$

We will use the notation $g^{-1}(x)$ to mean $h(x)$.

A. *Transformation to AWP with Linear Increase Profile*

For a function $F(x)$ such that $\frac{dF(x)}{dx} = \frac{1}{f(x)}$, let us define a new process

$$y_t = F(x_t).$$

Then the transformed process $\{y_t\}$ is such that, when the window $\{x_t\}$ is increasing, we have

$$y_{t+\Delta} - y_t = \frac{x_{t+\Delta} - x_t}{f(x_t)} + o(\Delta) = \Delta + o(\Delta).$$

The reason for introducing this transformation is that it simplifies the analysis and visualisation of the window evolution process since now the transformed process has a linear increase profile ($y_{t+\Delta} - y_t = \Delta + o(\Delta)$). Since $f(\cdot) > 0$, it is seen that $F(\cdot)$ is strictly increasing and hence invertible. Thus there is a one to one correspondence between an AWP and its linearly increasing counterpart. A detailed justification of the above transformation is given in [13]. Under the above transformation, the loss process has an intensity $\tilde{\lambda}(y) \stackrel{\Delta}{=} \lambda(F^{-1}(y))$ when $y_t = y$. In case of a loss event in interval $[t, t+\Delta]$ the decrease profile of this transformed protocol will be determined by $F(\cdot)$ and $g(\cdot)$ as,

$$y_{t+\Delta} = G(y_t) + o(\Delta),$$

where $G(\cdot) \stackrel{\Delta}{=} F(g(F^{-1}(\cdot)))$ is assumed to have same properties as $g(\cdot)$.

The map $F: \ W \mapsto Y$ is actually a transformation from a general increase protocol to an additive increase protocol (like the standard TCP's congestion avoidance algorithm). Thus, it is enough to study protocols following an additive increase general decrease algorithm. In the rest of this section we will only work with an AWP that has a linear increase profile and a general decrease profile in the presence of a general window dependent loss rate $\lambda(\cdot)$.

Let $y_{min} \stackrel{\Delta}{=} F(x_{min})$ be the lower bound on the transformed window size. Then $G(y_{min}) = y_{min}$ Let

$$\begin{aligned} S(y) &\stackrel{\Delta}{=} \{u \geq y : G(u) \leq y\}, \\ H(y) &\stackrel{\Delta}{=} \sup S(y). \end{aligned}$$

Since the set $s(x)$ is connected and compact for each x with $\inf s(x) = x$, the set $S(y)$ is also connected and compact for any given y and $\inf S(y) = y$. Note that the above definitions imply that $G(H(y)) = y$. The interpretation of these quantities are as follows: $S(y)$ is the set of all possible window sizes, greater than or equal to y, such that an occurrence of a loss event at these window sizes results in a window size of at most y, and $H(y)$ is the maximum such window size.

We now give the quantities defining an AWP, and the transformation introduced above for some standard examples.

[1]Congestion losses occur also when the window size reaches the practical limit of the total round trip pipe size (sum of the link bandwidth-delay product and the router buffer). This aspect of congestion losses will be addressed later in this section. For presentation of the basic model, we assume here that there is no upper bound on x_t.

1) For the case of an additive increase multiplicative decrease protocol like the congestion avoidance algorithm of standard TCP,

$$f(x) = 1, \qquad g(x) = \frac{x}{2},$$
$$F(x) = x, \qquad G(y) = \frac{y}{2},$$
$$S(y) = [y, 2y], \qquad H(y) = 2y.$$

2) For the case of an MIMD protocol like the slow start algorithm of standard TCP or Scalable TCP [4],

$$f(x) = \alpha x, \qquad g(x) = \beta x,$$
$$F(x) = \frac{\log(x)}{\alpha}, \qquad G(y) = y - \theta,$$
$$S(y) = [y, y + \theta], \qquad H(y) = y + \theta,$$

for some $\alpha > 0$, $\beta < 1$ and $\theta = -\frac{\log \beta}{\alpha}$.

B. Incorporating a Bound on the Window Size

The window evolution process described above does not incorporate any bound on the maximum allowed window size. In practice, however, there will be an upper bound M on the window size that the AWP is allowed to use. This bound usually is either the receiver's advertised window (which is the maximum number of packets that the receiving entity's receive buffer can accommodate) or the total round trip pipe size. The behavior of the AWP under these two bounds are very different. In the first case where the window is restricted by the receiver's advertised window M, the window size stays at this value till another loss event takes place. While in the case where M represents the round trip pipe size, reaching this limit results in an instantaneous congestion loss and the window size is reduced. However, since the loss rate is assumed to be function of window size alone, it follows that we can study the second case via the first case (for details, see [3] which also addresses this issue for a constant loss rate). Hence in what follows we will restrict ourselves to the case where M represents the window limitation owing to the receiver's advertised window.

Assume that the range of the values of the window process is divided into the intervals between points $[H^j(y_{min}), H^{j+1}(y_{min})]$ where H^j is $j-$fold composition of $H(\cdot)$ with $H^0(y_{min}) \triangleq y_{min}$. Consider an M such that $M = H^m(y_{min})$ for some $m \geq 1$. Note that, under our choice of M, $H^j(y_{min}) = G^{m-j}(M)$ with $G^0 \triangleq M$ and $G^i = G(G^{i-1})$. Let $l_i \triangleq G^{i-1} - G^i$. Under the above definitions, $y \in [G^i, G^{i-1}] \Rightarrow H(y) \in [G^{i-1}, G^{i-2}]$. The case where such an m does not exist, i.e., $H^{m-1} < M < H^m$ for some m, is not possible since the definition of $G(\cdot)$ depends on y_{min} and M implicitly, and it ensures that $G(G^{m-1}) = y_{min}$ so that $H^m = M$.

We consider a further modification in the evolution of the window process y_t. For this modified process, the window size is unbounded. However, when $y > G^0$, we assume that the loss rate is constant and equal to $\lambda(G^0)$. We also assume that if a loss event takes place when $y \geq G^0$, the window is dropped to $G^1 = G(M) = H^{m-1}(y_{min})$. The evolution of the modified process for $y < G^0$ is unchanged, i.e., a loss event occurs with rate $\lambda(y_t)$ and the window is dropped to $G(y_t)$ in case of a loss event when $y_t < G^0$. Note that we are assuming a linear increase of y_t for any value of y. Thus, the modified process has the following evolution: the increase profile is given by $y_{t+\Delta} = y_t + \Delta + o(\Delta)$. Losses occur according to a Poisson process of rate $\lambda(y_t \wedge G^0)$ and the window reduction in case of a loss event in time interval $[t, t+\Delta)$ is $y_{t+\Delta} = G(y_t \wedge G^0)$.

C. The Kolmogorov Equations

Let $\pi(x)$ be the density function and $\Pi(x)$ be the distribution function of the x_t process with an increase profile $f(\cdot)$, decrease profile $G(\cdot)$ and loss rate $\lambda(\cdot)$. The Kolmogorov equations satisfied by $\pi(\cdot)$ is derived in [13] as $f(x)\pi(x) =$

$$\begin{cases} \int_{u=x}^{\infty} \pi(u)\lambda(G^0)du = \lambda(G^0)\Pi^c(x), & x \geq G^0, \\ \int_{u=x}^{G^0} \pi(u)\lambda(u)du + \lambda(G^0)\Pi^c(G^0), & G^1 < x \leq G^0, \\ \int_{u=x}^{H(x)} \pi(u)\lambda(u)du, & x_{min} \leq x < G^1. \end{cases}$$

III. Relations Between Window Evolutions of Two Systems

We now consider two systems, 1 and 2, having their own increase and decrease profiles and loss rates, denoted by $f_i(\cdot), g_i(\cdot), \lambda_i(\cdot),\ i \in \{1, 2\}$. We provide a condition under which these two systems have related stationary probability distribution. Assuming that $g_1(x) = g_2(x) = g(x),\ \forall\ x$, and that in both the systems the upper bound on the window is the same (and is equal to M), the Kolmogorov equations for the two systems are

$$f_i(x)\pi_i(x) = \int_{u=x}^{g^{-1}(x)} \lambda_i(u)\pi_i(u)du,$$

i.e.,

$$\frac{f_i(x)}{\lambda_i(x)} \frac{\lambda_i(x)\pi_i(x)}{E[\lambda_i(X)]} = \int_{u=x}^{g^{-1}(x)} \frac{\lambda_i(u)\pi_i(u)}{E[\lambda_i(X)]} du$$

where

$$E[\lambda_i(X)] = \int_x \lambda_i(x)\pi_i(x)dx$$

is the expected loss rate in the i^{th} system. It is clear from the above set of equations that if $\frac{f_1(x)}{\lambda_1(x)} = \frac{f_2(x)}{\lambda_2(x)},\quad \forall x$, the functions $\frac{\lambda_1(x)\pi_1(x)}{E[\lambda_1(X)]}$ and $\frac{\lambda_2(x)\pi_2(x)}{E[\lambda_2(X)]}$, both being probability density functions integrating to unity, are equal for each x. Thus,

Theorem 1: If two AWP controlled window evolution are such that (i) both have same drop profile, and (ii) both have the same ratio of increase profile to loss rate for each x, then

$$\frac{\pi_1(x)}{\pi_2(x)} = C\frac{\lambda_2(x)}{\lambda_1(x)} = C\frac{f_2(x)}{f_1(x)}$$

where $C = \frac{E[\lambda_1(X)]}{E[\lambda_2(X)]}$.

This result is important as it gives us a way to analyse one system using the analysis of the other related system. We use this result in Section VI-B where we use the observation that an AIMD protocol with constant loss rate and an MIMD protocol with linear loss rate satisfy the requirement of Theorem 1 as for the first (AIMD) system $f(x) = \alpha$ and $\lambda(x) = \lambda$ while for the second (MIMD) system $f(x) = \alpha x$ and $\lambda(x) = \lambda x$, and both have same multiplicative decrease factor. Since the analysis for the first system is known from [3], we use it to find stationary distribution for the MIMD protocol with linear loss rate.

In the special case where both the system use multiplicative decrease profile with a constant decrease factor β, we can get some more detailed equivalence between the two systems. This is done next.

A. A Queueing Model for Multiplicative Decrease

Consider an AWP with a constant multiplicative decrease factor β. Introduce the transformation $z_t = \ln M - \ln x_t$. We are assuming that $x_{min} = 0$, i.e., z_t is unbounded. We can do this since we can use standard approach ([1, Chapter 14]) to analyse the case where z_t is bounded by $\ln M - \ln x_{min}$ from that where z_t is unbounded. From the transformation, it is evident that the multiplicative decrease of the process x_t transforms to a *constant* increase of $\ln \beta$ in the evolution of z_t process (see [13] for a pictorial representation of the transformation). The evolution of z_t process suggests that z_t can be thought of as workload process of a queue for which the service requirement of the customers is constant ($-\ln \beta$). If the increase profile and loss rate for x_t process are $f(\cdot)$ and $\lambda(\cdot)$, then in the z_t process, the customer arrival rate is $\lambda(Me^{-z_t})$ and service rate is $\frac{f(Me^{-z_t})}{Me^{-z_t}}$, both depending on the workload process z_t. Thus we get a queueing system with constant service requirements and state dependent service rates and arrival rates. We get

Theorem 2: Consider window evolutions in the two systems 1 and 2 introduced above, both with same multiplicative decrease profile. If $\frac{f_1(x)}{\lambda_1(x)} = \frac{f_2(x)}{\lambda_2(x)}$ then the distribution of window size just before loss instants is *same* in both the systems.

Applying this result to the two systems satisfying the above condition where the first one is AIMD with constant loss rate and the second one is MIMD with linear loss rate, we see that the stationary distribution of the window process just before (and hence just after) loss instants is same. Thus, the standard AIMD protocol with constant loss rate is same as MIMD protocol with linear loss rate in the sense that the distribution of the window sizes just before losses are the *same* for the two.

Further, since Theorem 2 is valid for any two AWPs satisfying the required conditions, it is seen that if for one of the AWPs the loss rate is constant, the PASTA property implies that the stationary (time average) distribution of the window size in the system with constant loss rate is same as the window size distribution just before losses in either of the system.

Now we specialize Theorem 1 to the case of multiplicative decrease protocols and provide a stronger result.

Theorem 3: Consider window evolutions in the two systems 1 and 2 introduced above, both with same multiplicative decrease profile. If $\frac{f_1(x)}{\lambda_1(x)} = \frac{f_2(x)}{\lambda_2(x)}$ then the time average distribution of window size $\pi_i(\cdot)$ in the two systems is related by

$$\frac{\pi_1(x)}{\pi_2(x)} = C\frac{f_2(x)}{f_1(x)} = C\frac{\lambda_2(x)}{\lambda_1(x)}$$

where $C = \frac{\lambda_1(M)\Pi_1^c(M)}{\lambda_2(M)\Pi_2^c(M)}$ with $\Pi_i^c(\cdot)$ denoting the complementary distribution function.

Corollary 1: Consider the scenario of Theorem 3. Then

$$\frac{E[\lambda_1(X)]}{E[\lambda_2(X)]} = \frac{\lambda_1(M)\Pi_1^c(M)}{\lambda_2(M)\Pi_2^c(M)}.$$

Proof Follows from Theorem 1 and Theorem 3. •

Application of this result to the running example of the two systems where the first one is AIMD with constant loss rate and other is MIMD with linear loss rate, we see that the expected window size in the MIMD case is

$$E[X] = \frac{M\Pi_2^c(M)}{\Pi_1^c(M)}.$$

Remark It is important to note that the window process with a lower bound of 1 and an upper bound of $M < \infty$ is always ergodic in the case of multiplicative decrease algorithm. This is because for any bounded loss rate and positive increase profile, the window process $\{x_t\}$ is irreducible. However, if we assume $x_{min} = 0$, then the corresponding unbounded transformed queueing process need not always be ergodic. Thus, we can not always use the truncation method of [1] mentioned above. Hence it becomes necessary to solve the detailed Kolmogorov equations for each case. This remark is, in particular, relevant for the case where the AWP is MIMD and the loss rate is constant. For this case the transformed process z_t is just the workload process of an M/D/1 queue. However we can not use this approach for $\lambda > -\ln \beta$ owing to the above mentioned reason.

Remark The process $z_t \triangleq M - x_t$ always represents the workload process in a queue with state dependent arrival rate, service rate and service requirement.

Remark The results of this section indicate that if the losses come from an AQM scheme, then there are many AWP-AQM pairs (i.e., $f(\cdot)$ and $\lambda(\cdot)$) which have the same drop profile ($g(\cdot)$) and have similar performance (in the sense of Theorem 1). Moreover, if the decrease profile is fixed to be a multiplicative one, we see that all these AWP-AQM pairs have *same* window distribution before drop instants (Theorem 2).

IV. Solution to the Kolmogorov Equations

The general Kolmogorov equation for the stationary distribution of an AWP with increase profile $f(\cdot)$, decrease profile $g(\cdot)$ and loss rate $\lambda(\cdot)$ is rewritten as

$$\frac{f(x)}{\lambda(x)}\frac{\lambda(x)\pi(x)}{E[\lambda(X)]} = \int_{u=x}^{g^{-1}(x)} \frac{\lambda(u)\pi(u)}{E[\lambda(X)]}du,$$

where $E[\lambda(X)] = \int_{x=x_{min}}^{\infty} \lambda(x)\pi(x)dx$. Note that $E[\lambda(X)]$ always exists if the window process is bounded by a quantity

M. Introducing the transformation $\tilde{\pi}(x) = \frac{\lambda(x)\pi(x)}{E[\lambda(X)]}$, we get

$$\frac{f(x)}{\lambda(x)}\tilde{\pi}(x) = \int_{u=x}^{g^{-1}(x)} \tilde{\pi}(u)du,$$

which is the Kolmogorov equation of an AWP whose increase profile is $\frac{f(x)}{\lambda(x)}$ and decrease profile is $g(x)$ while the loss rate now is a state-independent constant, equal to unity; We will see an example of such an approach in Section VI-B. We can further introduce a transformation of the new protocol (having increase profile $\frac{f(x)}{\lambda(x)}$) to another AWP with a linear increase profile as indicated in Section II-A. Thus, without loss of generality, we can assume that the protocol under consideration has a linear increase profile and the loss rate is unity. Now we provide an expression for the stationary probability distribution for the bounded process with linear increase, a unit loss rate and a general decrease profile $G(\cdot)$.

Theorem 4: For $M > x > G^1$, $\Pi^c(x) = c_1 e^{-x}$. For $x \in I_k = [G^k, G^{k-1}], k \geq 2$,

$$e^x \Pi^c(x) = \sum_{j=1}^{k} c_j J_{k,k-j}(x),$$

where

$$c_k = e^{G^k}\Pi^c(G^k),$$

with

$$c_1 = [\sum_{j_1=1}^{m-1}\sum_{j_2=1}^{j_1-1}\cdots\sum_{j_{m-2}=1}^{j_{m-3}-1} q_{m,j_1}q_{j_1,j_2}\ldots q_{j_{m-2},1}]^{-1}$$

and the constants $q_{k,j}$ are defined as

$$q_{k,j} = [J_{k-1,k-1-j}(G^{k-1}) - J_{k,k-j}(G^{k-1})],$$

with

$$\begin{aligned} J_{2,1}(x) &= \int_{u=G^2}^{x} e^{u-H(u)}du, \\ J_{k,l}(x) &= \int_{G^k}^{x} e^{u-H(x)}J_{k-1,l-1}(H(u))du \text{ (for } x \in I_k) \\ M &= H^m(x_{min}). \end{aligned}$$

Remark It is worth noting that Theorem 4 gives the workload process distribution in a queue with Poisson arrival process, state dependent deterministic service requirements and bounded workload process (this is because the process $M - x_t$ corresponds to the workload process in the mentioned queueing system). In Proposition 2 we give details for the standard M/D/1 queue with finite workload capacity.

Remark As mentioned in the beginning of this section, the stationary distribution for the transformed system of AWP with linear increase profile and constant loss rate with a bounded window also gives the distribution of the original window process with state dependent loss rate and general increase profile upto a multiplicative constant. Hence *Theorem 4 gives the solution to the Kolmogorov equation for a general AWP with a general loss rate* (upto a multiplicative constant of $E[\lambda(X)]$).

Till now the development did not consider exact form of loss rate $\lambda(\cdot)$ for the original process. In the following sections we consider specific forms of $\lambda(\cdot)$ to find the stationary window size distribution and work out the solution of Kolmogorov equation for several available TCP versions. We start with the case where $\lambda(x) \equiv \lambda$, independent of the current window size in Section V. We then consider the situation of a linearly increasing loss rate, i.e., $\lambda(x) = \lambda x$ in Section VI.

V. Constant Loss Rates: $\lambda(x) = \lambda$

We give a method of solving the Kolmogorov equations for a general AWP with constant loss rate. This method has been used in Section IV and we briefly mention it here for sake of completeness. First observe that any transformation applied to the window size does not affect the loss rate. Thus for any given AWP, we can always apply the transformation introduced in Section II-A to get a linear increase profile. For the evolution of this transformed process, we see that the jump rate (loss rate) is still λ, independent of anything else. Thus we need only study the case of linear increase general decrease protocols. In this section we first identify the special structure of the Kolmogorov equation for window evolution with constant loss rate with a general decrease profile. We then work out the details for Scalable TCP [4].

Here we do not dwell into the issue of lower bound x_{min} on the window size of the original process. This is because the lower bound on the transformed process is $y_{min} = F(x_{min})$ can take very different values depending on $F(\cdot)$. For example, if the original AWP is MIMD, the function $F(\cdot)$ turns out to be logarithmic and hence the lower bound y_{min} can be $-\infty$ or 0 depending on whether $x_{min} = 0$ or 1, respectively.

For this case the Kolmogorov equations are

$$\begin{aligned} \pi(y) &= \int_{u=y}^{\infty} \pi(u)\lambda du, \quad G^1 < y, \\ \pi(y) &= \int_{u=y}^{H(y)} \pi(u)\lambda du, \quad y_{min} \leq y < G^1. \end{aligned}$$

Using integrating factor method for the Kolmogorov equation for $y \geq G^1$, we get

$$\Pi^c(y) = \Pi^c(G^1)e^{-\lambda(y-G^1)}, \quad y \geq G^1.$$

For $x \in [G^k, G^{k-1}]$, $k \geq 1$, let $\Pi_k(x) \triangleq \Pi(x)$. Thus,

$$\frac{d}{dy}\Pi_k(y) = \Pi(H(y))\lambda - \Pi(y)\lambda, \quad k \geq 2.$$

Assuming that $H(\cdot)$ are such that $\Pi(\cdot)$ is continuous at $G^i, \forall i$, we have $\Pi_k(G^{k-1}) = \Pi_{k-1}(G^{k-1}), k \geq 2$. For $k \geq 2$, this gives us $\Pi_k(\cdot)$ recursively as

$$\Pi_k(x) = \Pi_{k-1}(G^{k-1})e^{\lambda(G^{k-1}-x)} - \lambda e^{-\lambda x}\int_{u=x}^{G^{k-1}} e^{\lambda u}\Pi_{k-1}(H(u))du.$$

Similar approach has also been used in [3] which considers an AIMD protocol with constant loss rate.

A. Application to MIMD Protocols with Bounded Window

Once again our approach will be to transform the MIMD window evolution to the case of a linear increase profile. For the case of MIMD protocols, the window evolution is described as follows. In case of no loss in interval $[t, t+\Delta]$, the window increases to $x_{t+\Delta} = (x_t + \alpha x_t \Delta + o(\Delta)) \wedge M$, for some $\alpha > 0$ and an upper bound M on the window size. In case of a loss in interval $[t, t+\Delta]$, the window decreases to $x_{t+\Delta} = (\beta x_t) \vee 1 + o(\Delta)$, where $1 > \beta > 0$ is the multiplicative decrease constant. The natural lower bound of $x_t \geq 1$ packet applies. It is clear now that the transformation $x_t \mapsto \frac{\log x_t}{\alpha} \triangleq y_t$ results in the process $\{y_t\}$ having linear increase profile. The transformed window after a loss event in interval $[t, t+\Delta]$ is given by $y_{t+\Delta} = (y_t - \theta)^+ + o(\Delta)$, where $\theta \triangleq \frac{-\log(\beta)}{\alpha} > 0$. Thus for this case, $y_{min} = 0$ and $H(y) = y + \theta$. Hence, $G^0 = M = m\theta$ and $G^l = (m-l)\theta$, $G(u) = (u-\theta)^+$. Let $\Pi(y)$ be the probability distribution of the process $\{y_t\}$. Defining, $\Pi_k(y) = \Pi(k\theta + y)$ for $0 \leq y \leq \theta$, we have

Proposition 1: For $0 \leq k \leq m-1$ and $0 \leq x \leq \theta$,

$$\Pi_k^c(x) = \sum_{j=0}^{m-k-1} \Pi_{k+j}^c(0) \frac{(\lambda x)^j}{j!}.$$

Proposition 2: The constants $\Pi_k^c(0)$ are given by

$$\Pi_{m-1}^c(0) = [(a^{m-1} - \phi_1(m-1)) + \sum_{s=1}^{m-3} (-1)^s \sum_{l=s}^{m-2} \phi_s(l) \\ (a^{m-l-1} - \phi_1(m-l-1)) + (-1)^{m-2}(a-b)\phi_{m-2}(m-2)]^{-1}$$

and for $0 \leq k \leq m-2$,

$$\begin{aligned}\Pi_k^c(0) &= \Pi_{m-1}^c(0)[(a^{m-k-1} - \phi_1(m-k-1)) \\ &\quad +(-1)^{m-k-2}(a-b)\phi_{m-k-2}(m-k-2) \\ &+ \sum_{s=1}^{m-k-3} (-1)^s \sum_{l=s}^{m-k-2} \phi_s(l)(a^{m-k-l-1} - \phi_1(m-k-l-1))],\end{aligned}$$

with $a = e^{\lambda\theta}$ and $\phi_j(l)$ defined recursively as, $\phi_0(0) = 0$ and

$$\phi_{j+1}(l) = \sum_{s=1}^{l-j} \phi_1(s)\phi_j(l-s), \quad j \geq 1.$$

Proof: See [13]. •

Remark For the above case where $x_{min} = 1$, the evolution of process $\frac{\log M}{\alpha} - \frac{\log x_t}{\alpha} = \frac{\log M}{\alpha} - y_t$ corresponds to the workload process of an M/D/1 queue with a bounded workload capacity of $\frac{\log M}{\alpha}$ and service requirement of θ for each customer. This is a system similar to that of [7] with a difference that the model of [7] assumes that the customer that can make the workload to exceed a certain fixed threshold is lost. While in our case such a customer is not completely lost but is admitted with a service that makes the workload process equal to the threshold. Our result is thus of independent interest in queueing theory.

Remark We can also easily incorporate another value of $0 < x_{min} \neq 1$ in the above analysis. As mentioned in Section III-A, if we assume that $x_{min} = 0$, the transformation $\frac{\log M}{\alpha} - \frac{\log x_t}{\alpha}$ corresponds to the workload process of a classical M/D/1 queue. For this case the moments and the stationary window size distribution are well known.

B. MIMD with Unbounded Window: A D/M/1 Queue

Assuming that $M = \infty$, i.e., there is no bound on the window size, we can not use the results from above directly in this case. Another approach to obtain the stationary distribution $\Pi(\cdot)$ is to look at the process $\{y_n, n \geq 0\}$ embedded just after the loss instants of the transformed process with linear increase profile, $\{y_t\}$. Let $\{a_n, n \geq 0\}$ denote the time between two successive losses. Then, $\{y_n\}$ is a continuous state space Markov chain which is given by the recursive equation

$$y_{n+1} = (y_n + a_n - \theta)^+. \tag{3}$$

We note that the loss process a_n is exponentially distributed with rate λ. Equation 3 is the same as the recursive equation for the workload in a $D/M/1$ queue with interarrival time θ and mean service time $\frac{1}{\lambda}$. The steady state distribution of y, $P(y_n \leq y)$ can be obtained as [10]

$$P(y_n > y) = \left(1 - \frac{s_1}{\lambda}\right)e^{-s_1 y}, \tag{4}$$

where s_1 is the root of the equation $s + \lambda = \lambda e^{s/\theta}$ in $Re(s) < 0$. The stability condition for the workload process of this D/M/1 queue (and, equivalently, for the window size process $\{y_t\}$) is $\theta > \frac{1}{\lambda}$.

In order to obtain the distribution at a random arrival instant, we note that the window size just before loss instant, y_{n+1}^-, is given by $y_{n+1}^- = y_n + a_n$. Since a_ns are exponentially distributed with parameter λ,

$$\begin{aligned}P(y_{n+1}^- > y) &= \lambda \int_0^\infty P(y_n > y - a)e^{-\lambda a} da \\ &= \lambda \int_y^\infty e^{-\lambda a} da + \lambda \int_0^y P(y_n > y - a)e^{-\lambda a} da \\ &= e^{-\lambda y} + \lambda(1 - \frac{s_1}{\lambda})e^{-s_1 y} \int_0^y e^{-(\lambda - s_1)a} da = e^{-s_1 y}.\end{aligned}$$

Using PASTA property, the window size distribution at a random time is the same as that seen by the loss arrivals. Since $y = \frac{\log x}{\alpha}$, the window distribution at any random time is

$$P(x_t > x) = x^{-\frac{s_1}{\alpha}} \tag{5}$$

Remark This approach can also be used for bounded window process when loss rate is large enough so that the bound is attained with negligible probability.

Remark If the window size in the original process $\{x_t\}$ is bounded by a value of M then the evolution of the process $\{y_n\}$ (now embedded just before loss instants in the process $\{y_t\}$) is

$$y_{n+1} = \min\left((y_n - \theta)^+ + a_n, \frac{\log M}{\alpha}\right),$$

which is the workload just after an arrival in a D/M/1 queue with a bounded workload capacity of $\frac{\log M}{\alpha}$. The connection to M/D/1 queue implies that this is also the residual workload seen by arriving customers in an M/D/1 queue with finite workload capacity. We have, using the PASTA property in the M/D/1 system,

Theorem 5: The distribution of workload process just after arrivals in a D/M/1 queue with a finite workload capacity is same as that of the residual workload in an M/D/1 queue with same bound on the workload capacity.

This phenomenon can be viewed as a duality between the time averages in an M/D/1 queue and the customer averages in a D/M/1 system.

VI. Linear Loss Rate: $\lambda(x) = \lambda x$

In this section we give a general expression for the stationary distribution of the window size process with a linear increase profile under a linear loss rate assumption for general window decrease profile. We then provide the stationary distribution for Scalable TCP and HighSpeed TCP under linearly increasing loss rates. This is of practical interest as a linear loss rate is seen by the connection when each packet is dropped with a fixed probability p (see [6]).

A. Additive Increase General Decrease AWP

We now consider an AWP with a linear increase profile and assume that the loss rate is linearly increasing with the window size, i.e., $\lambda(u) = \lambda u$ for some $\lambda > 0$. This is the case of practical interest because the standard congestion avoidance phase of TCP is linearly increasing. Moreover, recently proposed HighSpeed TCP [5] opens up a possibility of wide range of protocols where the window increase is approximately linear (with a larger additive increase constant as compared to the standard TCP) and the decrease is given by some window dependent factor. As mentioned already, loss rates in cases where each packet is dropped with a fixed probability and TCP drops its window at most once in a round-trip time indeed increase linearly with the window size of the AWP. In the following we assume that the increase profile is same as that in standard TCP, i.e., window increases by one unit per unit time; this can be assumed because an increase profile with a different (constant) slope can be mapped to that of unit slope while keeping the loss rate linear using the transformation introduced in Section II-A.

Proposition 3: For x such that $G^1 < x < G^0$, the stationary distribution is, for $c_1 = \lambda G^0[1 - \Pi(G^0)]e^{\frac{\lambda {G^0}^2}{2}}$,

$$\pi(x) = c_1 e^{\frac{-\lambda x^2}{2}} \tag{6}$$

For $x \in (G^l, G^{l-1})$, $m \geq l > 1$,

$$\pi(x)e^{\lambda\frac{x^2}{2}} = \sum_{j=0}^{l-1} c_{l-j}\lambda^j \int_{u_1=H(x)} \dots \int_{u_j=H(u_{j-1})} u_1 e^{\lambda\frac{G(u_1)^2 - {u_1}^2}{2}} \dots u_j e^{\lambda\frac{G(u_j)^2 - {u_j}^2}{2}} du_j \dots du_1, \tag{7}$$

where c_j are some constants to be computed using the exact form of $G(\cdot)$.

For numerical computations, we can use continuity of $\Pi(\cdot)$ at the boundaries G^i to compute c_j's like done in Section V-A. Now we work out the above expression for the case of AIMD protocol.

1) The Case of Standard TCP: AIMD: For the case of standard TCP with linear window dependent loss rate, [6] has obtained an expression for the stationary window size distribution. Their method however requires *guessing* the expression for the stationary distribution and then proving it inductively. Our approach is to directly determine the distribution without need for guessing. This is a considerable amount of simplification for the case of a general AWP as we will see in section VI-B that the distribution can in general be not straightforward to guess. We will also see in the present section that the very nature of AIMD makes it (relatively) easier to predict the structure of the stationary distribution.

Proposition 4: For $x \in [G^l, G^{l-1}]$,

$$\pi(x) = \sum_{j=0}^{l-1} c_{l-j} b_j e^{a_j x^2}$$

where

$$\begin{aligned} b_j &= \Pi_{n=1}^{j} \frac{\lambda}{2K\sum_{\kappa=0}^{n}\beta^{-2\kappa}}, \\ K &= \lambda\frac{\beta^2 - 1}{2}, \\ a_j &= (\frac{K}{\beta^2}(\sum_{\kappa=0}^{j-1}\beta^{-2\kappa}) - 0.5\lambda), \end{aligned}$$

with $b_0 = 1$. Here c_l are integration constants.

Proof See [13] for proof and closed form expression for c_js in terms of Gamma functions. •

A similar expression has been obtained in [6]. However, [6] provides only the recursion for the integration constants appearing in their expressions. They need to compute the value of these constants using numerical integration at the end, whereas we have a closed form expression for these constants. The model of [6] allows for the window size of 0 packets (during time-out periods) and also allows multiplicative decrease while window size is less than one packet, this makes their recursion of infinite length. This also results in a large discrepancy in the distributions for small window sizes. As we pointed out in Section III-A, allowing for a window size of less than one packet may result in a model that is stable in only a restricted set of parameter values. Further, [6] also accounts for timeout periods and also distinguishes between triple dupack losses and timeout based loss recovery. It is seen that our model is easily extended to consider these possibilities (though we believe that these phenomenon are rare, hence not of significant importance, when the TCP-SACK version is used).

B. MIMD Protocols with Linear Loss Rates

Recall the evolution of window process $\{x_t\}$ for MIMD protocol from Section V-A. The window is bounded below by a constraint of $x_{min} = 1$ packet. The window evolution under such scenario is depicted in Figure 1. The figure shows that the

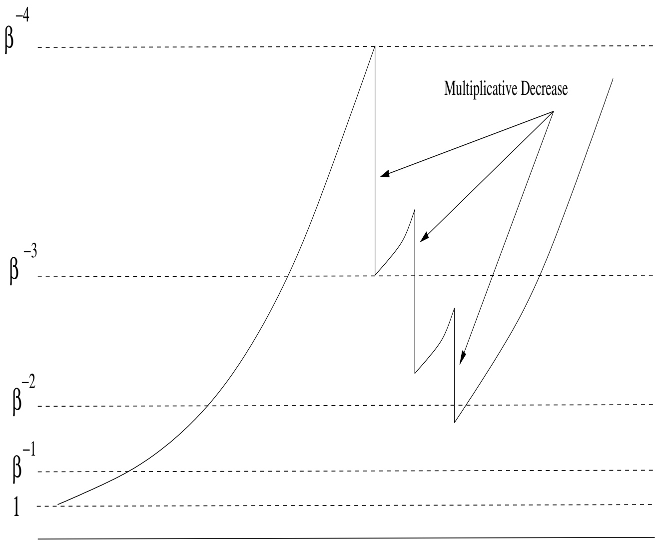

Fig. 1. Window evolution under MIMD protocol like Scalable TCP with a lower bound on window size.

window starts evolving from an initial value of 1 packet. There are some multiplicative decrease of window owing to random losses. The vertical axis is shown to be divided into various intervals $I_k \stackrel{\Delta}{=} (\beta^{-k}, \beta^{-k-1}]$. Here $\beta < 1$ is the multiplicative decrease factor. The significance of these regions is that if a loss event occurs when the window size is in interval I_{k+1} then the reduced window is in region I_k. We remark here that we are not working with the transformed window having a linear increase profile as introduced in Section II-A. The upper bound on x is $M = \beta^{-m}$ for some m.

For this case the Kolmogorov equations can be obtained for $x < \beta^{-m+1}$, as

$$\pi(x)\alpha x = \int_{u=x}^{u=\frac{x}{\beta}} \pi(u)\lambda u du.$$

Denote now, by an abuse of notation, $\lambda = \frac{\lambda}{\alpha}$. The above Kolmogorov equation is then

$$\pi(x)x = \int_{u=x}^{u=\frac{x}{\beta}} \pi(u)\lambda u du.$$

Proposition 5: The steady state probability density function of the window size under linear loss rate is given by, if $x \in I_{m-k}, k \geq 2$,

$$\pi(x) = MP_M \sum_{j=1}^{k} c_i^{(k)} \frac{\lambda}{x\beta^{i-1}} e^{\frac{\lambda x}{\beta^{i-1}}}.$$

Here $c_i^{(k)}$ are constants obtained by normalising $\pi(\cdot)$ to get probability measure and P_M is probability mass at M.

Proof: See [13] for expressions for P_M and $c_i^{(k)}$. •

One is often interested in finding the moments of the window process. This can be obtained easily without need to compute the coefficients $c_i^{(k)}$ as follows. We assume here that $x_{min} = 0$ and $M = \infty$; this is expected to approximate the case when the upper and lower bounds are not attained frequently. The Kolmogorov equation obtained above is multiplied by $x^{j-1},\ j \geq 0$ to obtain

$$\begin{aligned}
\pi(x)x^j &= x^{j-1}\int_{u=x}^{u=\frac{x}{\beta}} \pi(u)\lambda u du \\
\int_{x=0}^{\infty} \pi(x)x^j dx &= \int_{x=0}^{\infty} x^{j-1}\int_{u=x}^{u=\frac{x}{\beta}} \pi(u)\lambda u du dx \\
E[X]^j &= \int_{u=0}^{\infty}\int_{x=\beta u}^{u} x^{j-1} dx \pi(u)\lambda u du \\
\Rightarrow E[X] &= \frac{1}{-\lambda\ln(\beta)}
\end{aligned}$$

$E[X]^{j+1} = \frac{j}{\lambda(1-\beta^j)}E[X]^j = \frac{j!}{\lambda^j \Pi_{i=1}^{j}(1-\beta^i)}E[X],\ j \geq 1$, thus we get all the moments of the window size distribution. We see from the above that the tail of the window size distribution is exponentially decaying and that all the moments exist.

C. HighSpeed TCP

HighSpeed TCP (HSTCP, [5]) updates the window in a round-trip time according to the following rules: In case of no loss in a round-trip time during which the window size was w, the window is incremented by a window dependent quantity, denoted $a(w)$, so that the new window size is $w + a(w)$, and in case of a packet drop on a round-trip time, the window is decremented by a window dependent factor $b(w)$ so that the new window size is $(1-b(w))w$. The window size is bounded by two values w_l and w_h and

$$\begin{aligned}
a(w) &= \frac{2w^2 b(w)p(w)}{2-b(w)}, \\
b(w) &= \frac{\log(\frac{w}{w_l})}{\log(\frac{w_h}{w_l})}(b_h - b_l) + b_l, \\
p(w) &= \exp\left(\frac{\log(\frac{w}{w_l})}{\log(\frac{w_h}{w_l})}\log(\frac{p_h}{p_l}) + \log(p_l)\right),
\end{aligned}$$

where $b_h = b(w_h)$, $b_l = b(w_l)$, $p_l = p(w_l)$ and $p_h = p(w_h)$ are design parameters.

It is suggested in [5] to set $w_l = 31$ and $p_l = \frac{1.5}{w_l^2}$. Note that

$$p(w) = \nu w^{\mu}$$

where

$$\mu = \frac{\log(\frac{p_h}{p_l})}{\log(\frac{w_h}{w_l})}, \quad \text{and} \quad \nu = \frac{p_l}{w_l^{\mu}}$$

and $b(w) = A\log(w) + B$ with

$$A = \frac{b_h - b_l}{\log(\frac{w_h}{w_l})} \quad \text{and} \quad B = b_l - A\log(w_l).$$

Since $b_h < b_l$, $A < 0$ and since $w_h \geq w_l,\ p_h \leq p_l \Rightarrow \mu < 0$.

We observe that, if R represents the round-trip time, then

$$w(t+R) = w(t) + a(w(t)) = w(t) + \frac{2w^{2+\mu}b(w)\nu}{2-b(w)}.$$

This equation shows the importance of parameter μ in understanding the behavior of HSTCP. For example, $\mu = -2$ implies that HSTCP is similar to the standard AIMD algorithm of

TCP where in each round-trip time, the window is incremented by a small value (in this case $\frac{2b(w)\nu}{2-b(w)} \approx \nu b(w)$). If we take $\mu > -2$, then we get a protocol whose window increment increases with the window, for example, taking $\mu = -1$ implies that HSTCP is similar to Scalable TCP in behavior since now the increment is approximately linear in window size. This observation suggests need for care in tuning the HSTCP parameters. It also implies the possibility of existence of a choice of $\mu \in (-2, -1)$ which is neither as aggressive as Scalable TCP nor conservative as standard TCP. Now we analyse HSTCP assuming that $A \approx 0$ so that the decrease factor is constant. Since the form of function $b(w)$ is a design choice (see [5]), this form of $b(w)$ can be chosen for simplicity of implementation. Further, for this choice of $b(w)$ we can find the stationary window size distribution for the protocol for different values of μ as follows: First observe that for $b(w) = B$, the increase profile of the protocol is

$$f(w) = \frac{2\nu B w^{2+\mu}}{2-B}$$

and assuming a linear loss rate $\lambda(w) = \lambda w$, the Kolmogorov equation can be transformed to the case of unit loss rate as in Section III to get,

$$\frac{2\nu B w^{1+\mu}}{\lambda(2-B)}\pi(w) = \int_{u=w}^{\frac{w}{1-B}} \pi(u)du.$$

Now, this Kolmogorov equation, when transformed to the case of AWP with linear increase profile, becomes

$$\tilde{\pi}(y) = \int_{u=y}^{\frac{y}{(1-B)^{-\mu}}} \tilde{\pi}(u)du.$$

The closed form solution for this equation is known from [3] as this corresponds to the case of AIMD protocol with constant loss rate (here we have used the fact that $-\mu > 0$ so that $(1-B)^{-\mu} < 1$).

VII. Stability Results

An important problem now is to study the stability of the process $\{x_t\}$ for a given AQM or loss profile ($\lambda(\cdot)$) and a given AWP increase/decrease profile (the functions $f(\cdot)$ and $g(\cdot)$). Alternatively, for a given AWP, one would like to *design* an AQM profile; this design process must obviously address the issue of the stability of the window process under the chosen AQM profile. By *stability* here we mean that the window size (or the buffer occupancy) should, with large probability, take values in compact sets. In the following we give necessary and sufficient conditions for stability of the $\{x_t\}$ process; these condition can then be used in the design of AQM profile.

We first provide a stochastic ordering result which says that the steady-state window process with a larger upper bound is stochastically larger than the process with a smaller upper bound. We then give a necessary and sufficient condition for existence and uniqueness of an invariant measure for the window process such that this measure has most of its mass concentrated on compact sets. We then provide a transformation from a process with state-dependent loss rate to one with state-independent loss rate. The necessary and sufficient stability condition for state-independent loss rate is then seen to apply to a general AWP with a general loss rate. Since the loss rate $\lambda(\cdot)$ can take very different forms if an Active Queue Management (AQM) scheme is used, the study of this section also applies to the interaction of an AQM and AWP.

A. Construction of Bounded Processes

Throughout in this section we will assume that the deterministic increase of the $\{x_t\}$ process is linear. We have already seen that an AWP with a general increase profile can be *continuously* transformed to the one with linear increase. The assumption on $\{x_t\}$ process made above is then justified by the fact that a continuous transformation preserves compactness of sets and hence will also preserve stability property.

Consider the sequence $\{x_t^M\}$ of window processes bounded above by a constant M. Between jumps, the process $\{x_t^M\}$ increases linearly. However, if the process achieves the level M, it stays there until next jump (which occurs at rate $\lambda(M)$). For each such M, let $\pi_M(\cdot)$ and $\Pi_M(\cdot)$ denote, respectively, the stationary density and distribution for the bounded process $\{x_t^M\}$ (we assume existence of these). Let P_M denote the point mass at M of the stationary probability for $\{x_t^M\}$. The steady state Kolmogorov equation satisfied by $\pi_M(\cdot)$ is

$$\pi_M(y) = P_M \lambda(M) B(M, y) + \int_{x=y}^{M-} \lambda(x) B(x, y) d\Pi_M(x), \tag{8}$$

here $B(x, y)$ denotes the probability that a jump is to point less than or equal to y given that a downward jump occurred when $x_t = x$. For evolution of the window process $\{x_t\}$, we see that $B(x, \cdot)$ is a unit step function since the jumps are deterministic, i.e., $B(x, y) = I\{y > g(x)\}$.

B. Limit of the Bounded Processes

It is to be noted here that the convergence of the *process* $\{x_t^M\}$ to $\{x_t\}$ as $M \to \infty$ follows from arguments similar to those in [8]. Now, if a stationary distribution $\Pi(\cdot)$ exists for the original process $\{x_t\}$, it must satisfy the steady state Kolmogorov equation

$$\pi(y) = \int_{x=y}^{\infty} \lambda(x) B(x, y) d\Pi(x) \tag{9}$$

Proposition 6: The process $\{x_t\}$ is stable and has an invariant measure $\pi(\cdot)$ if $\Pi_M(\cdot)$ forms a tight family of probability measures.

Proof: See [13]. (See [9] for definition of tight family of probability measures.) •

Above we showed, via a weakly convergent subsequence, that an invariant probability measure exists if the sequence $\{\Pi_M(\cdot)\}$ is *tight*. However, it *may be* possible that there are many subsequences of $\{\Pi_M(\cdot)\}$ converging to different weak limits. In that case each of these weak limits is an invariant measure. Below we give sufficient conditions under which such a situation does not arise. This condition is satisfied by

AWP controlled window evolution with state-independent loss rate.

Lemma 1: If $B(\cdot, y)$ is a unit step function that is strictly decreasing for each y and if $\lambda(x)$ is independent of state x, then the sequence $\{\Pi_M(x)\}$ is monotone non-increasing in M for each fixed x, i.e., $x^{M_1} \leq_{st} x^{M_2}$ for $M_1 \leq M_2$.

Remark The monotonicity property obtained above is not an intuitive result. This is because for the bounded processes, though the solution to the Kolmogorov equations can be monotone, the normalization required to make them probability measures can have an unpredictable effect in general. In our case, however, it turns out that monotonicity is preserved by the required normalisation.

Remark Since the proof of Lemma 1 is sample-path wise and does not use the exact form of the increase profile, we see that it applies also to the system with a general increase profile and a constant loss rate.

We have shown above that the sequence $\{\Pi_M(x)\}$ is monotone non-increasing for specific structure of $B(\cdot, \cdot)$. Monotonicity of $\{\Pi_M(x)\}$ for each x implies that there is a unique pointwise limit of the sequence $\{\Pi_M(x)\}$ for each x. This remains valid whether or not $\{\Pi_M(\cdot)\}$ is *tight*. It is now easy to see that if $\{\Pi_M(\cdot)\}$ is *tight* then there exists a *unique* weak limit of $\{\Pi_M(\cdot)\}$. If, however, $\{\Pi_M(\cdot)\}$ is not *tight* then it follows that there exists a value $0 < r < 1$ such that $\lim_{M\to\infty} \Pi_M(x) > r$ for all x. We have thus

Proposition 7: Under the conditions of Lemma 1, the process $\{x_t\}$ is stable and has a unique invariant measure iff $\{\Pi_M(\cdot)\}$ forms a tight family of probability measures.

Now we state an important result which will be used in study of stability of an AWP under a general loss rate.

Theorem 6: The process $\{x_t\}$ is unstable iff $\lim_{M\to\infty} P_M > 0$.

Proof: Since if $\lim_{M\to\infty} P_M > 0$, the sequence of probability measures $\Pi_M(\cdot)$ can not be *tight*, the proof follows from Proposition 7.

If the sequence $\Pi_M(\cdot)$ is tight then it is easily seen, using monotonicity of $\Pi_M(\cdot)$, that $\lim_{M\to\infty} P_M = 0$. Thus the reverse implication also follows. •

Remark Results relating stability and tightness of probability measures are known in context of Markov chains also (see [11]). The results of this section are for Markov processes of a specific kind and the criteria for checking the tightness as in Theorem 6, obtained from establishing monotonicity probability measure over constrained state spaces is new.

We now make the following conjecture,

Conjecture 1: For the case of state dependent loss rate $\lambda(\cdot)$, the if part of Theorem 6 remains valid, i.e., if $\lim_{M\to\infty} P_M > 0$ then the process $\{x_t\}$ is unstable.

C. Application of the Stability Result

We now establish a *necessary and sufficient* criteria for stability of a general AWP (general increase and decrease profile) under a general loss rate. Using a transformation introduced in Section IV brings us in the framework of Theorem 6 which assumes that the loss rate is constant and the AWP has a linear increase profile. Thus, without loss of generality, we can assume that the protocol under consideration has a linear increase profile and the loss rate is unity. It is clear that the original system is stable if this transformed system is stable. Now we use Theorem 4 that provides an expression for the probability mass at the bound M for the bounded process with linear increase and unit loss rate and obtain (recall the notation of Theorem 4)

Theorem 7: A general AWP controlled window evolution is stable under a general loss rate iff

$$\frac{e^{-M}}{\sum_{j_1=1}^{m-1}\sum_{j_2=1}^{j_1-1}\cdots\sum_{j_{m-2}=1}^{j_{m-3}-1} q_{m,j_1} q_{j_1,j_2}\cdots q_{j_{m-2},1}} \underset{m\to\infty}{\longrightarrow} 0.$$

Proof Follows from Theorem 4 since we know that the AWP is stable *if and only if* $P_M \to 0$ as $M \to \infty$. •

Remark Recall that Lemma 1 applies also to the system with a general increase profile and constant loss rate. Since, for the window evolution under a general increase profile $f(\cdot)$ and a general loss rate $\lambda(\cdot)$, its stationary probability measure $\pi(\cdot)$ is such that $\frac{\pi(x)\lambda(x)}{E[\lambda(X)]}$ satisfies the Kolmogorov equation for a system with increase profile $\frac{f(x)}{\lambda(x)}$ (see Section IV), the function

$$\Pi_M(x) = \int_{u=x_{min}}^{x} \frac{\pi(u)\lambda(u)}{E[\lambda(X)]} du$$

is monotone in M for each x. This result carries over to the corresponding queueing system with finite workload capacity in a natural way.

VIII. Numerical Results

We obtained time average density of the window process from *ns-2* [14] simulations for AIMD protocol with constant loss rate and MIMD protocol with linear loss rate. The multiplicative decrease factor $\beta = 0.5$ for both the protocols and the loss rate, λ_a, for AIMD protocol was set to either 0.005 or 0.008. The MIMD protocol had an increase profile of $f_m(x) = 1.01x$ as in Scalable TCP while the AIMD protocol had $f_a(x) = 1$. The loss rate for MIMD protocol was $\lambda(x) = \lambda_m x$ where λ_m was chosen so that the conditions of Theorem 1 were satisfied. This requirement is satisfied if $\lambda_m = 0.01\lambda_a$, i.e., $\lambda_m = 0.00005$ or 0.00008. Figure 2 gives the function $\pi_m(x)$ for MIMD and $\frac{Cf_a(x)\pi_a(x)}{f_m(x)}$ where C is $\frac{\lambda_m E_m[X]}{\lambda_a}$ with $E_m[X]$ being the expected window size for MIMD protocol obtained from simulation. The results are as predicted by Theorem 1, i.e., $\pi_m(x) = \frac{\lambda_m E_m[X]}{\lambda_a}$, $\forall x$. For the same experimental setup, we also obtained the distribution of window sizes just before losses. The results are plotted in Figure 3 which shows that, in agreement with Theorem 2, this distribution is same for the two systems. Now, we compute the numerical values from our analysis of Section VI-B and compare it with simulation results of Figure 2 for MIMD with linear loss rate. Figure 4 gives the comparison between analysis and simulations. Since the density function is already plotted in Figure 2, here we plot the $(E[X^n])^{\frac{1}{n}}$ vs. n for $1 \leq n \leq 10$. The analysis and simulations are seen to match well for smaller values of n (≤ 6); the small discrepancy for

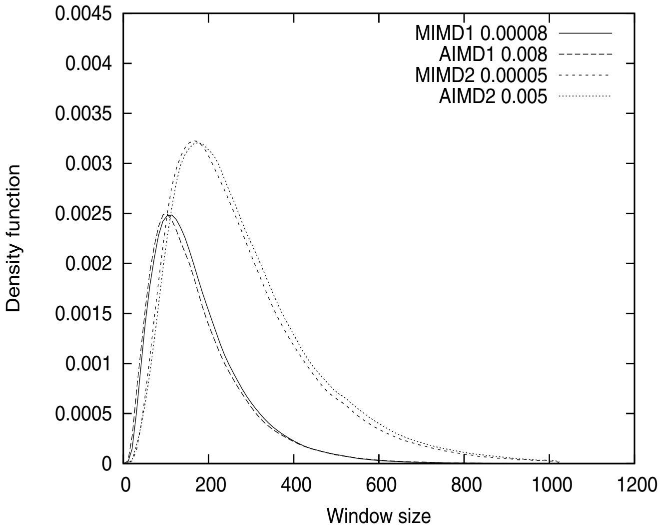

Fig. 2. Comparison of time average distribution for MIMD with linear loss rate and for AIMD with constant loss rate.

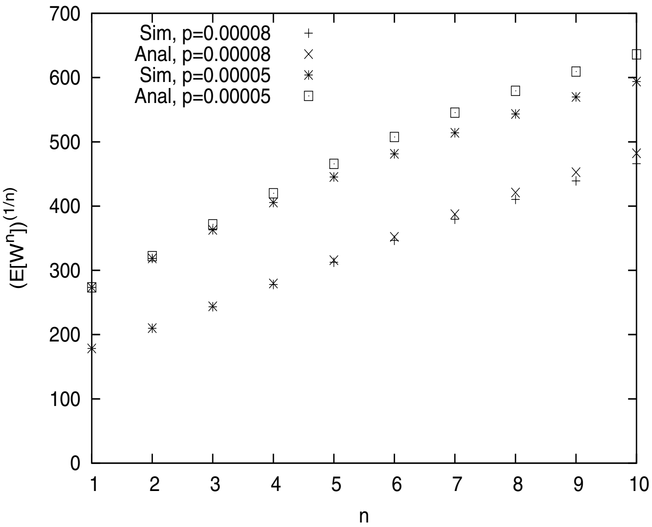

Fig. 4. First 10 moments for MIMD with linear loss rate.

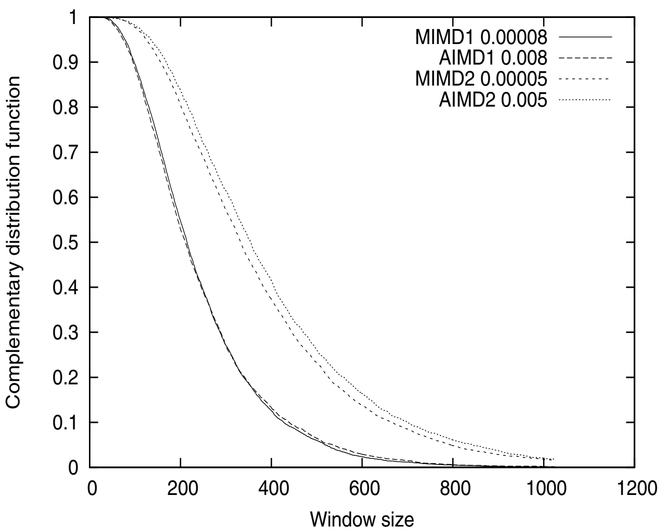

Fig. 3. Window size distribution just before loss instants for MIMD with linear loss rate and for AIMD with constant loss rate.

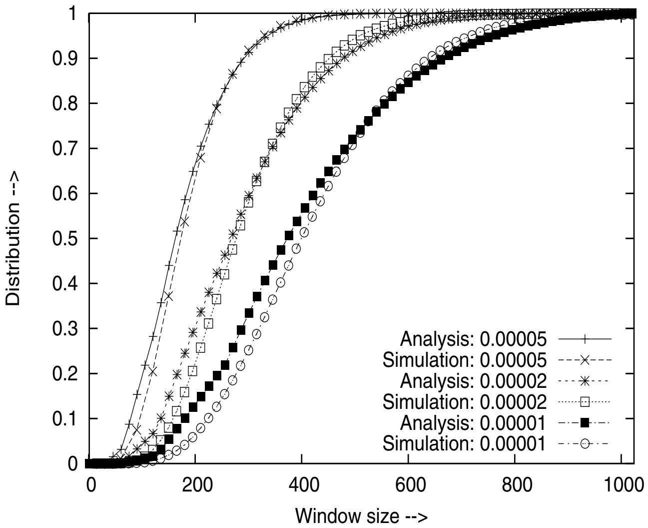

Fig. 5. Window distribution for AIMD protocol with linear loss rate.

large values of n could be owing to finite simulation run-length.

Figure 5 gives results from simulation and numerical computation of analysis of Section VI-A.1 for TCP's standard AIMD protocol with linear loss rate for different values of $\lambda = 1e-5, 2e-5$ and $5e-5$. The slight discrepancy between simulation and analytical results could be owing to numerical problems involved in solving the required recursions (see [3] for discussion on similar lines).

Figure 6 gives complementary distribution function of the stationary window process for HSTCP assuming that the multiplicative decrease factor $b(w)$ is fixed to a constant value B. Recall the parameters A, B, μ and ν of Section VI-C. We fix $A = 0$, $B = 0.125$ and ν so that $\frac{2B\nu}{2-B} = 0.01$ so that the case of $\mu = -1$ corresponds to the Scalable TCP [4]. The plot shows results for values of the parameter $\mu = -0.9, -1.0, -1.2$. In order to do this, we varied the parameters p_l and p_h accordingly. The figure also gives numerical results from the analysis of Section VI-C. It is observed from the figure that one can approximate any increase function only by varying μ while keeping the multiplicative drop factor $b(w)$ constant. In particular, we note that the distribution is very sensitive to the value of the parameter μ. This simplifies the algorithm as now there are not many independent design choices and, moreover, the analysis of Section VI-C combined with that of [3] provides closed form result for the stationary distribution.

IX. Conclusion

We considered a general congestion control protocol with a state dependent loss probability. We obtained closed for

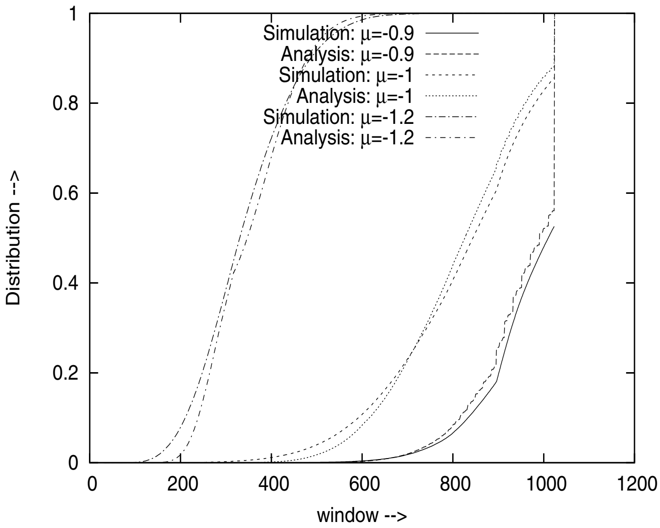

Fig. 6. Window size distribution for HSTCP with linear loss rate.

expression for the stationary window size distribution of a general AWP with a general loss rate. Various transformations introduced provided us with many equivalence relations. Most significant being that of the relation between window evolution and the workload process in a finite capacity queueing system with state dependent service and arrival rates and a state dependent deterministic service requirement. Several results of independent interest in queueing theory were obtained. Some monotonicity properties of the stationary window distribution as well as a necessary and sufficient condition for stability of the window size process were proved.

We have assumed that the loss $\lambda(\cdot)$ is a given function. This may be the case in the applications using AQM schemes and where congestion losses are rare. However, when most of the losses are owing to congestion losses, it appears to be more realistic that the form of $\lambda(\cdot)$ will itself be determined by the AWP. Also, it is possible that, like in model of [12], the loss process $\lambda(\cdot)$ may itself be a stochastic process. These considerations are topic of further research.

Theorem 7 may not be easily verifiable for a general AWP decrease profile (since this involves finding the functions $J_l(\cdot)$). A simpler condition to establish the convergence or divergence of the involved series is yet another further possible direction.

In the analysis of HSTCP we have chosen a multiplicative decrease algorithm with window independent decrease factor. We now aim at using some approximations for the evolution of the window process using the drop profile suggested in [5]. It is also important to study an optimal choice of the parameter μ.

Acknowledgement This work was supported by grant from the *Centre Franco-Indien pour la Promotion de la Recherche Avancee* (CEFIPRA) under project no. 2900-IT-1.

REFERENCES

[1] S. Asmussen, “Applied probability and queues,” *Springer*, 2003.

[2] A. A. Kherani and A. Kumar, “Stochastic Models for Throughput Analysis of Randomly Arriving Elastic Flows in the Internet,” *in Proceedings of IEEE INFOCOM*, New York, June, 2002.

[3] E. Altman, K. Avratchenkov, C. Barakat and R. Nunez Queija, “State-dependent M/G/1 Type Queueing Analysis for Congestion Control in Data Networks,” *in Proceedings of IEEE INFOCOM*, Anchorage, April, 2001.

[4] Tom Kelly, “Scalable TCP: Improving Performance in Highspeed Wide Area Networks,” Submitted for publication, December 2002. Available at `http://www-lce.eng.cam.ac.uk/~ctk2`/scalable/

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

[6] A. Budhiraja, F. Hernandez-campos, V. G. Kulkarni and F. D. Smith, “Stochastic Differential Equation for TCP Window size: Analysis and Experimental Validation,” *Probability in the Engineering and Informational Sciences*, Vol. 18, 2004.

[7] D. Perry, W. Stadje and S. Zacks, “The M/G/1 Queue with Finite Workload Capacity”, *Queueing Systems*, Vol 39, 2001.

[8] D. P. Heyman and W. Whitt: *Limits for Queues as the Waiting Room Grows*, 1988.

[9] P. Billingsley: *Convergence of probability measures*, Wiley, New York-London-Sydney, 1968.

[10] L. Kleinrock, “Queueing systems, Vol 1 : Theory”, J. Wiley and sons , 1975.

[11] A. A. Borovkov, “Ergodicity and stability of stochastic processes”, J. Wiley and sons , 1998.

[12] E. Altman, K. E. Avrachenkov and C. Barakat, “A stochastic model of TCP/IP with stationary random losses”, ACM SIGCOMM 2000, Stockholm, Sweden, also in Computer Communication Review, v.30, no.4, October 2000, pp.231-242.

[13] E. Altman, K. E. Avrachenkov, A. A. Kherani and B. J. Prabhu, “Performance Analysis and Stochastic Stability of Congestion Control Protocols” *INRIA Report No. RR-5262, Sophia-Antipolis, France, July 2004.*

[14] NS-2 Network Simulator, *available at* `http://www.isi.edu/nsnam/ns/`