

Time-Averaging of High-Speed Data Transfer Protocols

Richard Marquez, Eitan Altman, and Solazver Solé-Álvarez

Abstract—We propose two modeling approaches of additive-increase/multiplicative-decrease (AIMD) congestion control mechanisms. The first separates the increase and decrease parts whereas the second describes the rate evolution of the congestion window as a continuous process governed by a differential equation. We relate these approaches and show that the second (fluid-flow) model results from deterministic time averaging of the first (discontinuous) model. A generalized class of nonlinear protocols, which includes Floyd's HighSpeed TCP, is then proposed and analyzed. Our findings are validated by simulation.

Index Terms—Averaging, discontinuous differential equations, high-speed transfer data protocols, modeling, transmission control protocol.

I. INTRODUCTION

The transmission control protocol (TCP), during congestion avoidance phase, regulates its congestion window by an additive-increase/multiplicative-decrease (AIMD) mechanism [1]. Increase and decrease TCP parameters are represented, respectively, by constants a and b . Usually, $a = 1, b = 1/2$ [2], [3]. Mathematical modeling of TCP and, especially, fluid-flow modeling has received considerable attention in last few years; see, e.g., [4].

Many argue that TCP performance is poor for fast long-distance networks and have proposed different high speed protocols (HSTCP, STCP, XCP, etc.). For instance, HighSpeed TCP (HSTCP) [5] and Scalable TCP (STCP) [6] consider increase/decrease parameters $a(w)$ and $b(w)$ which depend on window size w . A large class of generalized AIMD models, which includes TCP, HSTCP, and STCP as particular cases, is proposed in this note. We consider two modeling approaches: 1) A *discontinuous* modeling approach, which separates the increase and decrease parts; it uses smooth increase of the rate with sharp smoothed decrease at loss instants; see, e.g., [7], [8] for different approaches on discontinuous models. 2) A fluid-flow continuous model describing the window evolution as a *continuous process* governed by a differential equation [4], [9].

Studying a single TCP flow, we represent increase and decrease TCP behaviors by two continuous differential equations and combine them through a *discontinuous binary feedback* signal u into a single discontinuous model. We assume that the impact of the rest of the network can be modeled as periodic packet losses. We show that a continuous (fluid-flow) model can then be obtained by *deterministic time averaging*. We give conditions for the existence of a unique solution process to which the congestion window size converges.

This note is organized as follows. In Section II, we briefly review time averaging of differential equations with periodic right-hand side. Section III proposes a discontinuous model to represent the behavior

of a single TCP flow. This model is extended to generalized congestion control mechanisms, that include TCP, HSTCP, and STCP. Section IV uses periodic in time packet losses to derive generalized fluid-flow models. The proposed (discontinuous and averaged) models are implemented in MATLAB and compared to *ns-2* traces [10]. We conclude with some final remarks. A preliminary version of this work was presented in [11].

II. CLASSICAL TIME AVERAGING

Deterministic time-averaging has been utilized frequently to obtain simpler models which retain the important properties of a system. For instance, this approach is standard for pulse-modulated systems [12]. The following results and definitions are taken from [13], [14].

Let $x = x^*$ be an equilibrium point for the nonlinear system $\dot{x} = f(x)$, i.e., $f(x^*) \equiv 0$, where $f : D \rightarrow \mathbb{R}^n$ is continuously differentiable and D is a neighborhood of x^* . Let the Jacobian matrix of $f(x)$ at $x = x^*$ be $A = (\partial f / \partial x)(x)|_{x=x^*}$. Then, 1) x^* is asymptotically stable if the real part $\Re(z_i) < 0$ for all eigenvalues z_i of matrix A , and 2) x^* is unstable if $\Re(z_i) > 0$ for one or more z_i of A , see [13, Th. 4.7]. An equilibrium point is called *hyperbolic* if $\Re(z_i) \neq 0$. *Asymptotic stability* means the solution $x(t)$ converges to x^* as time t tends to ∞ . A *periodic solution* (or *orbit*) corresponds to a solution $x(t)$ of $\dot{x} = f(x)$, such that $x(t+T) = x(t)$ for a constant $0 < T < \infty$. Roughly speaking, a periodic solution is asymptotically stable if every solution tends to it.

Let x, y, x_0 belong to an open subset $D \subset \mathbb{R}$, let $t \in \mathbb{R}^+ = [0, \infty)$, and let the parameter ϵ vary in the range $(0, \epsilon_0]$ with $\epsilon_0 \ll 1$. Let $f : \mathbb{R}^+ \times D \rightarrow \mathbb{R}$ be a piecewise continuous function.¹ Consider the problem of finding the solution $x(t)$ of

$$\frac{dx}{dt} = \epsilon f(t, x), \quad x(0) = x_0 \quad (1)$$

If $f(t, x)$ is a T -periodic function in its first argument, we let the *averaged system* be

$$\frac{dy}{dt} = \epsilon f^0(y), \quad y(0) = x_0 \quad (2)$$

where $f^0(y) = T^{-1} \int_0^T f(t, y) dt$. The *slow motion* of the periodic solution $x(t)$ of (1) corresponds to the solution $y(t)$ of (2). When $f(t, x) = f(x)$, i.e., independent of t , we have moreover $y = \bar{x} = T^{-1} \int_0^T x(t) dt$, see [13]. The periodic or oscillatory component of $x(t)$ around this slow motion constitutes the *fast behavior*.

Theorem 1 [13, p. 430]: There exists a positive ϵ_0 such that, for all $0 < \epsilon \leq \epsilon_0$, 1) $x(t) - y(t) = O(\epsilon)$ as $\epsilon \rightarrow 0$ on the time scale $1/\epsilon$, and 2) if x^* is a hyperbolic (respectively, asymptotically) stable equilibrium point for f^0 , then $x(t) - y(t) = O(\epsilon)$ as $\epsilon \rightarrow 0$ for all $t \in \mathbb{R}^+$, and (1) possesses a unique periodic orbit (periodic solution) which is hyperbolic (respectively, asymptotically) stable and belongs to an $O(\epsilon)$ neighborhood of x^* .

This theorem provides a relation between slow and fast behaviors of (1). Roughly speaking, if the eigenvalues of the Jacobian matrix of $f^0(y)$ around the equilibrium point of the averaged (2) all have negative real parts, the corresponding periodic solution $\phi(t, \epsilon)$ of (1) is then asymptotically stable for ϵ sufficiently small; moreover, $\phi(t, \epsilon)$ lies in an $O(\epsilon)$ neighborhood of x^* . If one of the eigenvalues has positive real part, $\phi(t, \epsilon)$ is unstable.

¹Here, the classical "smooth" assumption is replaced by a "piecewise continuous" assumption [15, App. C].

Manuscript received January 3, 2005; revised July 20, 2005. Recommended by Associate Editor R. Srikant. This work was supported in part by a French-Venezuelan ECOS-Nord Grant V04M01 (PI-2003000207), by the EuroNGI Network of Excellence, and by FONACIT Grant S1-2002000472.

R. Marquez and S. Solé-Álvarez are with the Departamento de Sistemas de Control, Facultad de Ingeniería, Universidad de Los Andes, Mérida 5101, Venezuela (e-mail: marquez@ula.ve; solazver@ula.ve).

E. Altman is with INRIA, BP93, 06902 Sophia Antipolis, France (e-mail: altman@sophia.inria.fr).

Digital Object Identifier 10.1109/TAC.2005.860259

III. DISCONTINUOUS GENERALIZED TCP MODELS

A. Discontinuous TCP Model

Consider a single TCP source. A *round-trip time* RTT is the interval between the time instant a packet is sent and the moment its acknowledgment (ACK) is received. The basic *congestion avoidance* algorithm [2], which defines the behavior of congestion window cwnd , can be described as follows: If there is no congestion, cwnd is increased by an amount a every time a full window is acknowledged (i.e., at every RTT). This phase is known as *additive increase*

$$\text{ACK: } \text{cwnd}[n+1] = \text{cwnd}[n] + a. \quad (3)$$

Dividing (3) by RTT, we obtain the *rate of change* of the congestion window size, as follows:

$$\frac{\text{cwnd}[n+1] - \text{cwnd}[n]}{\text{RTT}} = \frac{a}{\text{RTT}} \approx \frac{d}{dt} \text{cwnd}.$$

Replacing cwnd by the continuous variable w yields the differential equation for the cwnd increase dynamics [4], [7]

$$\frac{dw}{dt} = \frac{a}{\text{RTT}}. \quad (4)$$

After a congestion is detected, i.e., by duplicate ACKs or marked packets, cwnd is reduced by a factor $(1-b)$, $0 < 1-b < 1$. We have *multiplicative decrease*

$$\text{Drop: } \text{cwnd}[n+1] = (1-b) \times \text{cwnd}[n]. \quad (5)$$

We assume that the time T between reduction of the window are much larger than RTT. This is indeed the case if packet losses are separated by a time much larger than RTT. When using a *NewReno* or *SACK* implementation [16] of TCP, this assumption still holds even in the case of several losses within the same RTT; we call these *loss events*. This assumption is then equivalent to having the time between loss events much larger than RTT. This is the basis of the following (smooth) modeling approach of TCP decreasing behavior in the presence of packet losses. Instead of considering downward jumps at loss instants, we prefer to *smooth the jumps* over the RTT interval. This reflects the fact that, in practice, the throughput does not decrease immediately. This is also in line with modeling of the decrease part in other congestion control protocols as in [17].² Thus, we propose to replace (5) by

$$\frac{dw}{dt} = -\frac{k}{\text{RTT}}w \quad (6)$$

which holds valid during a time interval of length RTT. The solution $w(t)$ to (6) is given by $w(t) = w_0 \exp(-kt/\text{RTT})$ where $w_0 = w(0) > 1$. Starting with a window size w_0 , the congestion window decreases after an RTT to $w(t = \text{RTT}) = (1-b)w_0$, see (5). This is, $(1-b)w_0 = w_0 \exp(-k)$ which results in the value $k = -\ln(1-b)$ ("ln" stands for the natural logarithm). Equation (6) then becomes

$$\frac{dw}{dt} = \ln(1-b) \frac{w}{\text{RTT}}. \quad (7)$$

Later, for the study of HSTCP, we use the approximation $-\ln(1-b) \approx 2b/(2-b)$, $0 \leq b \leq 0.5$.

Let u denote a congestion indication signal, taking values in the set $\{0, 1\}$. This variable represents the *binary feedback* of [1], where

²Nonetheless, our approach relies on the basic reasonable assumption $\text{RTT} \ll T$ (where T is the interval of time between two consecutive reduction of the window), providing a reliable approximation to instantaneous jumps.

$u = 0$ represents the noncongestion (increase load) phase, and $u = 1$ indicates the decrease part due to packet losses. By using u , (4) and (7) lead to a differential equation with *discontinuous right-hand side*, see [18], which we call *discontinuous TCP model*

$$\text{RTT} \frac{dw}{dt} = a - (a - \ln(1-b)w)u, \quad w(0) = w_0 \geq 1. \quad (8)$$

In order to demonstrate existence, uniqueness of solutions, or even averaging results of discontinuous model (8), it is enough to point out that signal u corresponds to a piecewise continuous function; see, e.g., [15, App. C]. Let $t_k \in \mathbb{R}$, $k = 1, 2, \dots$, be the time instants at which a packet loss events occur, such that $0 \leq t_1 < t_1 + \text{RTT} \leq t_2 < t_2 + \text{RTT} \leq t_3 \dots$. Signal u can be written as a train of pulses

$$u(t) = \begin{cases} 1, & \text{if } t_k \leq t < t_k + \text{RTT} \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

B. A Class of Generalized Models

A *discontinuous generalized TCP (GTCP) model* results then

$$\text{RTT} \frac{dw}{dt} = g(w) - (g(w) + wh(w))u \quad (10)$$

where w is the congestion window size, and u is given as before by (9); $g(w)$ and $h(w)$ are appropriate smooth functions defining, respectively, increase and decrease behaviors. In the case of TCP, it is clear we obtain $g(w) = a$ and $h(w) = -\ln(1-b)$.

C. HighSpeed TCP

HSTCP is described as follows [5]. If window size cwnd is smaller than a given value W , i.e., $\text{cwnd} \leq W$, the increase/decrease parameter functions are $a(\text{cwnd}) = a = 1$, $b(\text{cwnd}) = b = 1/2$, as in standard TCP. For $\text{cwnd} > W$, parameters $\pi(\text{cwnd})$, which is reminiscent of the per-packet drop rate p , $a(\text{cwnd})$, $b(\text{cwnd})$ are

$$\begin{aligned} \pi(\text{cwnd}) &= e^{[\ln(P_1) - \ln(P)] \frac{\ln(\text{cwnd}) - \ln(W)}{\ln(W_1) - \ln(W)} + \ln(P)} \\ b(\text{cwnd}) &= (B-b) \frac{\ln(\text{cwnd}) - \ln(W)}{\ln(W_1) - \ln(W)} + b \\ a(\text{cwnd}) &= \text{cwnd}^2 \pi(\text{cwnd}) \frac{2b(\text{cwnd})}{2-b(\text{cwnd})}. \end{aligned} \quad (11)$$

Default values: $B = 0.1$, $W = 31$, $P = 3/(2W^2)$, $P_1 = 10^{-7}$, $W_1 = 83000$. We assume $b(\text{cwnd})$ remains constant ($=B$), for $\text{cwnd} > W_1$. In practice parameter b is lower bounded, i.e., $0 < 0.09 \leq b \leq 0.5$, see [5]. The discontinuous HighSpeed TCP model is then given by (10)

$$\text{RTT} \frac{dw}{dt} = wh(w)(w\pi(w) - (w\pi(w) + 1)u) \quad (12)$$

where cwnd has been replaced by w , parameters $a(w)$, $b(w)$ are defined by (11), and $h(w)$ and $g(w) = a(w)$ are given by

$$h(w) = \frac{2b(w)}{2-b(w)} \quad \text{and} \quad g(w) = w^2 \pi(w)h(w). \quad (13)$$

D. Scalable TCP

In STCP, congestion window increases nonlinearly in much the same way as *slow start* [2] does. Let C be the link capacity (in packets/s). For the increasing part, the initial value of the congestion window is given by $(1-b)C \times \text{RTT}$ packets. In the decrease behavior, after $T - \text{RTT}$ seconds, we have $C \times \text{RTT}$ packets. In order to model the increase part,

we have $C \times \text{RTT} = (1 - b)C \times \text{RTT} \exp((k/\text{RTT})(T - \text{RTT}))$, cf. (6). Thus $k = -\ln(1 - b)(\text{RTT}/T - \text{RTT})$. The discontinuous STCP model results then

$$\text{RTT} \frac{dw}{dt} = g(w) - \left(g(w) + w \frac{\text{RTT} \ln(1 - b)}{T - \text{RTT}} \right) u \quad (14)$$

where $g(w) = \ln(1 - b)w$. Note that this model depends explicitly on T . This is related to the fact that this protocol possesses a ‘‘fixed recovering time,’’ one of the announced features of STCP. See [6] for details. We will see this generates a kind of unfairness when periodic losses are considered, see paragraph IV.D.

IV. FLUID-FLOW (AVERAGED) MODELS

Throughout this section, we assume the following.

- 1) Constant T (between successive window reductions).
- 2) Fixed round trip time RTT much smaller than T , i.e., $\text{RTT} \ll T$, such that the quotient RTT/T is constant.

Remark 1: In practice T is often random (e.g., when TCP operates over wireless channels). We prefer to approximate it by a constant, e.g., the expected inter-loss time. If random independent interloss times are replaced by their expectation, then the steady state throughput decreases; see [19, eq. (9)]. Therefore, the model we study can be used as a *lower bound* for TCP throughput. This is a standard approximation; see, e.g., [20]. Our model gives a relation between the throughput and T which is valid even when T is unknown. Yet there are many cases where the distribution of T does not depend on the protocol of the analyzed connection. This is the case in TCP over wireless channels and in wide area networks where huge number of other connections may share a common bottleneck link. In that case, T may be estimated independently and used to obtain the throughput.

Assumption 1 leads to the following control function $u = u(t)$:

$$u(t) = \begin{cases} 1, & \text{if } t_k \leq t < t_k + \text{RTT} \\ 0, & \text{if } t_k + \text{RTT} \leq t < t_{k+1} = t_k + T \end{cases} \quad (15)$$

where $t_k, k = 1, 2, \dots$, represents a time instant at which a packet loss occurs. A period T has a length $T = t_{k+1} - t_k$. Function $u(t)$ is then T -periodic, i.e., $u(t + T) = u(t)$ for all t .

Time-average congestion window \bar{w} is defined by $\bar{w} = T^{-1} \int_0^T w(\sigma) d\sigma$. We use \bar{w} to denote the average of w over a period T ; *averaged control* μ is given by $\mu = T^{-1} \int_0^T u(\sigma) d\sigma$. Note μ corresponds to a sort of *packet-loss time rate*. Expression (15) yields a constant μ , independent of time scaling, given by $\mu = \text{RTT}/T$.

A. Averaged TCP Model

The averaging method of (1) can also be applied to system given by (8) and (15). Changing the time variable t to $t = \epsilon s$, we have

$$\text{RTT} \frac{dw}{ds} = \epsilon [a - (a - \ln(1 - b)w(s))u(s)] \quad (16)$$

$$\text{where } u(s) = \begin{cases} 1 & \text{if } s_k \leq s < s_k + \frac{\text{RTT}}{\epsilon} \\ 0 & \text{if } s_k + \frac{\text{RTT}}{\epsilon} \leq s < s_{k+1} = s_k + \frac{T}{\epsilon} \end{cases}$$

and $s_k = t_k/\epsilon$. Input $u(s)$ is periodic in s of period T/ϵ . A reasonable choice of ϵ is $\epsilon = \alpha \text{RTT}$, $0 < \alpha < 1$. As in (1) and (2), we associate with (16) the autonomous averaged system

$$\text{RTT} \frac{d\bar{w}}{ds} = \epsilon [a - (a - \ln(1 - b)\bar{w})\mu] \quad (17)$$

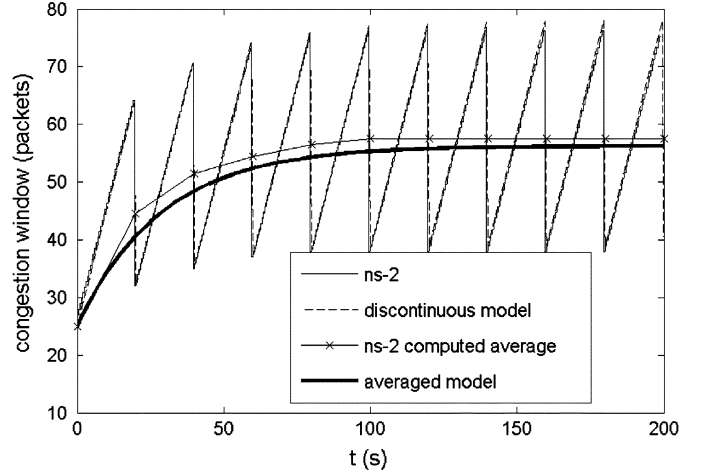


Fig. 1. *ns-2* behavior (solid line) coincides with the discontinuous model (dashed line). Averaged model behavior is plotted with a thick solid line. The average computed from *ns* trace is plotted with x 's.

where $u(s)$ has been replaced by its average μ . Note that μ must be limited to $0 \leq \mu = \text{RTT}/T < 1$.

Theorem 1 justifies approximating the solutions $w(t)$ of the controlled system (16) by the solutions $\bar{w}(t)$ of the averaged system (17). System (17) has a unique equilibrium point $\bar{w}^* = -(a/\ln(1 - b))((1/\mu) - 1) > 1$ which is *asymptotically stable*. Since (17) is linear, it directly follows that \bar{w}^* is globally stable, as long as $\epsilon \ln(1 - b)\mu < 0$, which holds in practice. By Theorem 1, the periodic discontinuous system (16) possesses then an *asymptotically stable periodic solution* of period T/ϵ . Going back to the original time coordinates, (17) yields the *averaged TCP model*

$$\text{RTT} \frac{d\bar{w}}{dt} = [a - (a - \ln(1 - b)\bar{w})\mu]. \quad (18)$$

In Fig. 1, we compare a *ns-2* trace, the averaging of this simulation over periods of T , and the behaviors of (8) and (18), obtained in MATLAB. We consider $w(0) = \bar{w}(0) = 25$ packets, $T = 20$ s, $\text{RTT} = 0.5$ s. The averaged model (18) seems to be a good approximation of the *ns* trace averaged over each period T .

Let us recall some results of [3]. Let S be the average sending rate in packets per RTT . The total number of packets between losses is $n = S \times T/\text{RTT}$. As drop rate p is defined by $n = 1/p$, i.e., $pS = \text{RTT}/T = \mu$, then $\mu = \bar{w}p$. Approximating $S \approx \bar{w}$, and $-\ln(1 - b) \approx 2b/(2 - b)$, we obtain $\text{RTT}(d\bar{w}/dt) = a - (a + (2b/2 - b)\bar{w})\bar{w}p$ where p is a new control. Equilibrium point $(\bar{w}^*, p^*(\bar{w}^*))$ yields

$$p^* = \frac{a}{\left(a + \frac{2b}{2-b}\bar{w}^*\right)\bar{w}^*}. \quad (19)$$

When \bar{w}^* is large, $(2b\bar{w}^*/2 - b) \gg a$, (19) translates into

$$p^* = \frac{a(2 - b)}{2b} \frac{1}{(\bar{w}^*)^2} \quad (20)$$

i.e., the ubiquitous *TCP response function* [3].

B. Main Result

Previous discussion yields our *averaged (fluid-flow) GTCP model*

$$\text{RTT} \frac{d\bar{w}}{dt} = \phi(\bar{w}, \mu) = g(\bar{w}) - (g(\bar{w}) + \bar{w}h(\bar{w}))\mu \quad (21)$$

where congestion signal u has been replaced by the packet-loss time rate μ . At steady-state, i.e., $\phi(\bar{w}^*, \mu) = 0$, i.e.,

$$\mu = \frac{g(\bar{w}^*)}{g(\bar{w}^*) + \bar{w}^* h(\bar{w}^*)} = \lambda(\bar{w}^*)$$

where $\lambda : (1, \infty) \rightarrow (0, 1)$. Equilibrium point \bar{w}^* is unique if $\lambda(\bar{w}^*)$ is invertible. A (locally) asymptotically stable \bar{w}^* of averaged system (21) exists if the Jacobian of ϕ satisfies $(\partial\phi/\partial\bar{w})(\bar{w}^*, \mu) < 0$.

We obtain then the following.

Lemma 1: The T -periodic discontinuous GTCP model, (10) and (15), possesses a unique hyperbolic periodic orbit on $\mathcal{W} \subset (1, \infty)$ if, for all $w = w(t) > 1, t \geq 0, \bar{w}^* \in \mathcal{W}$, functions g and h satisfy.

1) $g(w) > 0$; **2)** $h(w) > 0$; **3)** the invertible function $\lambda^{-1} : (0, 1) \rightarrow \mathcal{W}$ exists; and **4)** $-h(w)g(w) - w[g(x)(\partial h/\partial w) - h(x)(\partial g/\partial w)] < 0$. The largest interval \mathcal{W} given by $\mathcal{W} \equiv (1, \infty)$.

Proof: Straightforward by applying Theorem 1. Replace x^* by w^* , x by w , y by \bar{w} , x_0 by w_0 .³ ■

C. Fluid-Flow HSTCP Model

The averaged model of HSTCP⁴ is given by, see (12)

$$\begin{aligned} \text{RTT} \frac{dw}{dt} &= a(\bar{w}) - a(\bar{w})\mu - \frac{2b(\bar{w})}{2 - b(\bar{w})} \bar{w}\mu \\ &= wh(w)(w\pi(w) - (w\pi(w) + 1)\mu). \end{aligned} \quad (23)$$

The following proposition is not at all obvious for nonlinear extensions of AIMD algorithms.⁵

Corollary 1: The T -periodic HSTCP model, defined by (12) and (15), possesses a unique hyperbolic periodic orbit for initial conditions $w(0) = w_0 > W$ satisfying $b(w_0) > 0$.

Proof: Conditions 1)–4) of Lemma 1 apply for initial condition $w_0 > W$. Function $b(w)$ is strictly decreasing on $W < w < W_1$. In this case, there exist a unique equilibrium point $\bar{w}^* \in \mathcal{W}$. Parameters (11) implies $g(\bar{w}) > 0, h(\bar{w}) > 0$, for $W < \bar{w} < \bar{W}$. Thus, $\bar{w}^* \pi(\bar{w}^*) - (\bar{w}^* \pi(\bar{w}^*) + 1)\mu = 0$. Function λ^{-1} , for $W < \bar{w}^* < \bar{W}$

$$\bar{w}^* = k_1 \left(\frac{1 - \mu}{\mu} \right)^{k_2} = \lambda^{-1}(\mu) \quad (24)$$

where $k_1 = \frac{\exp((\ln(P_1) \ln(W) - \ln(P) \ln(W_1)/\ln(W_1) - \ln(W) + \ln(P_1) - \ln(P)))}{\log_{((W_1 P_1/W P))}(W/W_1)}$, i.e., $k_1 \simeq .401 \cdot 10^{-4}$, $k_2 \simeq 4.47686$. ■

D. Scalable TCP

The averaged model of (14) corresponds to the *trivial* one $\text{RTT}(d\bar{w}/dt) = 0$. This trivial averaged model means STCP possesses an infinite number of equilibrium points: In an environment under time-periodic packet losses, two identical STCP sources can

³Conditions 1)–4) of Lemma 1 are similar to those implied by *sliding regimes* of variable structure systems, cf. [18]: there exist two kinds of system trajectories approaching a (hypothetic) sliding surface, those approaching from below (increase behavior) and the other from above (decrease behavior).

⁴Note the similarities of this model to that of [21] (setting $\mu = \bar{w}\mu$)

$$\text{RTT} \frac{dw}{dt} = a(\bar{w}) - \frac{a(w)b(w)}{2} \mu - b(\bar{w}) \bar{w} \mu. \quad (22)$$

For large values of \bar{w} , (23) and (22) are similar, $b(\bar{w}) \approx 2b(\bar{w})/(2 - b(\bar{w}))$.

⁵Several limiting regimes may occur when a and b depend on w (in particular when a increases as a function of w); see, e.g., [22, Rem. 2].

share *unequally* the available bandwidth. For further modeling of this unfairness, see [23].

E. Link Sharing

1) TCP Link Sharing: Consider n TCP flows, $(\bar{w}_i, \text{RTT}_i)$, for $i = 1, \dots, n$, sharing a common bottleneck link. In steady state, the sum of averaged throughputs is given by

$$\frac{\bar{w}_1^*}{\text{RTT}_1} + \dots + \frac{\bar{w}_n^*}{\text{RTT}_n} \quad (\text{packets/s}).$$

Replacing the expressions for equilibrium points and $\mu_i = \text{RTT}_i/T_i$, it yields

$$-\frac{a}{\ln(1-b)} \left(\frac{T_1 - \text{RTT}_1}{\text{RTT}_1^2} + \dots + \frac{T_n - \text{RTT}_n}{\text{RTT}_n^2} \right).$$

The fraction of link utilization by connection 1 is then given by

$$\%L_1 = \frac{\frac{T_1 - \text{RTT}_1}{\text{RTT}_1^2}}{\frac{T_1 - \text{RTT}_1}{\text{RTT}_1^2} + \dots + \frac{T_n - \text{RTT}_n}{\text{RTT}_n^2}}.$$

Assume $T_i = T$ for all the flows. As $T \gg \text{RTT}_i$, we obtain

$$\%L_1 = \frac{1/\text{RTT}_1^2}{1/\text{RTT}_1^2 + \dots + 1/\text{RTT}_n^2} \quad (25)$$

i.e., the percentage of the bandwidth shared by a connection is inversely proportional to the square (power 2) of the round trip time of each connection, a well known result.

2) HSTCP Link Sharing: A similar result can be obtained for the bandwidth shared by n HSTCP flows $(\bar{w}_i, \text{RTT}_i)$, for $i = 1, \dots, n$: Taking into account (24), we obtain for $\bar{w} > W$

$$\%L_1 = \frac{1/\text{RTT}_1^{k_2+1}}{1/\text{RTT}_1^{k_2+1} + \dots + 1/\text{RTT}_n^{k_2+1}}.$$

As indicated before $k_2 + 1 = 5.47686 > 2$ (for TCP, value is 2). We see therefore that HSTCP with the previous parameters is more unfair than standard TCP with respect to throughput as function of the RTT.

V. CONCLUSION

We have defined discontinuous and fluid-flow (averaged) models for a class of GTCPs protocols. Assuming constant RTT and time T between successive window reductions, we showed that fluid-flow models result in fact from the classical, deterministic principle of averaging of the discontinuous models. We have demonstrated uniqueness and stability of time periodic (packet-loss) behaviors for GTCPs and, particularly, for TCP and HighSpeed TCP. Several lines of research will be pursued including averaged modeling of other types of protocols and algorithms, and conditions of uniqueness and stability.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. Discussions with H. Mounier and H. Sira-Ramírez were very insightful.

REFERENCES

- [1] D.-M. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Comput. Networks ISDN Syst.*, vol. 17, pp. 1–14, 1989.

- [2] V. Jacobson, "Congestion avoidance and control," *ACM Comput. Commun. Rev.*, vol. 18, no. 4, pp. 314–329, 1988.
- [3] S. Floyd, M. Handley, and J. Padhye. (2000, May) A comparison of equation-based and AIMD congestion control. ICSI Center for Internet Research. [Online]. Available: <http://www.icir.org/tfrc/aimd.pdf>
- [4] F. Kelly, "Mathematical modeling of the internet," in *Mathematics Unlimited—2001 and Beyond*, B. Engquist and W. Schmid, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 685–702.
- [5] S. Floyd, "High speed TCP for large congestion windows," IETF, Experimental RFC 3649, Dec. 2003.
- [6] T. Kelly, "Scalable TCP: Improving performance in highspeed wide area networks," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 33, no. 3, pp. 83–91, 2003.
- [7] P. Brown, "Resource sharing of TCP connections with different round trip times," in *Proc. IEEE Infocom*, Tel-Aviv, Israel, Mar. 2000, pp. 151–160.
- [8] F. Baccelli and D. Hong, "AIMD, fairness and fractal scaling of TCP traffic," in *Proc. IEEE Infocom*, Jun. 2002.
- [9] R. Srikant, *The Mathematics of Internet Congestion Control*. Boston, MA: Birkhäuser, 2004.
- [10] Network Simulator ns-2 [Online]. Available: www.isi.edu/nsman/ns/
- [11] R. Marquez, E. Altman, and S. Solé-Alvarez, "Modeling TCP and high speed TCP: A nonlinear extension to AIMD mechanisms," in *Proc. HSNMC 7th IEEE Int. Conf.*, Toulouse, France, Jun. 30–Jul. 2 2004.
- [12] P. Krein, J. Bentsman, R. Bass, and B. Lesieutre, "On the use of averaging for the analysis of power electronic systems," *IEEE Trans. Power Electron.*, vol. 5, no. 2, pp. 182–190, Feb. 1990.
- [13] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [14] J. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems*. New York: Springer-Verlag, 1985.
- [15] E. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. New York: Springer-Verlag, 1990.
- [16] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow. (1996, Apr.) RFC 2018: TCP selective acknowledgment options. [Online]. Available: <http://www.ietf.org/rfc/rfc2018.txt>
- [17] M. Ritter, "Network buffer requirements of the rate-based control mechanism for ABR services," in *IEEE Infocom*, San Francisco, CA, Mar. 1996.
- [18] V. Utkin, *Sliding Modes in Control and Optimization*. Berlin, Germany: Springer-Verlag, 1992.
- [19] E. Altman, C. Barakat, and K. Avrachenkov, "A stochastic model of TCP/IP with stationary ergodic random losses," in *Proc. of ACM/SIGCOMM*, Stockholm, Sweden, Aug. 28–Sep. 1 2000.
- [20] J. Mahdavi and S. Floyd, "TCP-friendly unicast rate-based flow control," Jan. 1997.
- [21] X. Huang, C. Lin, F. Ren, and H. Yin, "Highspeed TCP modeling and analysis," in *Proc. PATHNets Workshop*, San José, CA, 2004.
- [22] E. Altman, K. Avrachenkov, C. Barakat, and R. Nunez-Queija, "TCP modeling in the presence of nonlinear window growth," in *Proc. oITC-17*, Salvador da Bahia, Brazil, Sep. 2001.
- [23] E. Altman, K. Avrachenkov, and B. Prabhu, "Fairness in MIMD congestion control algorithms," *Proc. IEEE Infocom*, Mar. 13–17, 2005.

Adaptive Variable Structure Control of a Class of Nonlinear Systems With Unknown Prandtl–Ishlinskii Hysteresis

Chun-Yi Su, Qingqing Wang, Xinkai Chen, and Subhash Rakheja

Abstract—Control of nonlinear systems preceded by unknown hysteresis nonlinearities is a challenging task and has received increasing attention in recent years due to growing industrial demands involving varied applications. In the literature, many mathematical models have been proposed to describe the hysteresis nonlinearities. The challenge addressed here is how to fuse those hysteresis models with available robust control techniques to have the basic requirement of stability of the system. The purpose of the note is to show such a possibility by using the Prandtl–Ishlinskii (PI) hysteresis model. An adaptive variable structure control approach, serving as an illustration, is fused with the PI model without necessarily constructing a hysteresis inverse. The global stability of the system and tracking a desired trajectory to a certain precision are achieved. Simulation results attained for a nonlinear system are presented to illustrate and further validate the effectiveness of the proposed approach.

Index Terms—Adaptive control, cascade systems, hysteresis, nonlinear systems, Prandtl–Ishlinskii (PI) hysteresis model, robust control.

I. INTRODUCTION

The hysteresis phenomenon occurs in all the smart material-based actuators, such as piezoceramics and shape memory alloys [1]. When a nonlinear plant is preceded by the hysteresis nonlinearity, the system usually exhibits undesirable inaccuracies or oscillations and even instability [14] due to the nondifferentiable and nonmemoryless character of the hysteresis. The development of control techniques to mitigate the effects of hystereses has been studied for decades and has recently reattracted significant attention, as can be seen in [10] and the references therein. Much of this renewed interest is a direct consequence of the importance of hysteresis in numerous current applications. Interest in studying dynamic systems with actuator hysteresis is also motivated by the fact that they are nonlinear systems with *nonsmooth* nonlinearities for which traditional control methods are insufficient and thus require development of alternate effective approaches [15]. Development of a general frame for control of a system in the presence of unknown hysteresis nonlinearities is a quite challenging task.

To address such a challenge, the thorough characterization of these nonlinearities forms the foremost task. Appropriate hysteresis models may then be applied to describe the nonsmooth nonlinearities for their potential usage in formulating the control algorithms. Hysteresis models can be roughly classified into physics based models and purely phenomenological models. Physics-based models are built on first principles of physics. Phenomenological models, on the other hand, are used to produce behaviors similar to those of the physical systems without necessarily providing physical insight into the problems [19]. The basic idea consists of the modeling of the real complex hysteresis nonlinearities by the weighted aggregate effect of all possible so-called elementary hysteresis operators. Elementary hysteresis operators are

Manuscript received July 11, 2004; revised May 7, 2004. Recommended by Associate Editor G. Chen. The work of C.-Y. Su was supported in part by the Natural Science and Engineering Research Council of Canada and the National Natural Science Foundation of China under Grant 50390063.

C.-Y. Su is with Department of Mechanical and Industrial Engineering, Concordia University, Montreal, QC H3G 1M8, Canada. (e-mail: cysu@alcor.concordia.ca).

Q. Wang and S. Rakheja are with the Department of Mechanical and Industrial Engineering, Concordia University, Montreal, QC H3G 1M8, Canada.

X. Chen is with the Department of Electronic and Information Systems, Faculty of Systems Engineering, Shibaura Institute of Technology, Saitama 337-8570, Japan.

Digital Object Identifier 10.1109/TAC.2005.860260