Window Flow Control: Macroscopic Properties from Microscopic Factors

Ao Tang*, Lachlan L. H. Andrew[†], Krister Jacobsson[‡], Karl H. Johansson[‡], Steven H. Low[†], Håkan Hjalmarsson[‡]

* Cornell University, Ithaca, NY 14853, USA
[†] California Institute of Technology, Pasadena, CA 91125, USA
[‡] ACCESS Linnaeus Centre, Electrical Engineering, KTH, Stockholm, SE-100 44, Sweden

Abstract—This paper studies window flow control focusing on bridging the gap between microscopic factors such as burstiness in sub-RTT timescales, and observable macroscopic properties such as steady state bandwidth sharing and flow level stability. Using new models, we analytically capture notable effects of microscopic behavior on macroscopic quantities. For loss-based protocols, we calculate the loss synchronization rate for different flows and use it to quantitatively explain the unfair bandwidth sharing between paced and unpaced TCP flows. For delay-based protocols, we show that the ratios of round trip delays are critical to the stability of the system. These results deepen the fundamental understanding of congestion control systems. Packet level simulations are used to verify our theoretical claims.

I. INTRODUCTION

Flow control has been studied since at least the 1970s; see [7] and the references therein. Since the rapid explosion of the Internet, Internet congestion control algorithms, which are implemented in TCP (Transmission Control Protocol) and currently control the majority of Internet traffic, have been extensively studied for two decades. Roughly speaking, in the first ten years (1988-1997) starting with [8], research focused more on packet-level "microscopic" properties for simple topology networks, and researchers usually relied on qualitative reasoning and simulations [4], [6], [10], [28]. In the last ten years (1998-2007), following the seminal paper [12], the focus shifted to fluid models and tools from optimization and control theories, seeking "macroscopic" understandings with more quantitative calculations for general networks with potentially arbitrary topology [14], [18], [20], [25], [26].

However, there is a significant gap between these two types of studies; although the microscopic studies are usually more qualitative, they are much closer to the actual protocol implementations while in order to carry out general analysis, the recent macroscopic quantitative work generally ignores many (sometimes important) details of real protocols. One major issue that most existing work misses is the traffic burstiness within a round trip time (RTT). Current TCPs use window control, and rapid changes in the window induce burstiness, like slow-start causing bursts of almost back to back packets. There are also other sources of burstiness such as cross traffic [29]. Most importantly, once such burstiness is generated, it is self-perpetuating because of *ACK-clocking*. Because a new packet is transmitted only when an acknowledgment is received, the bursty pattern of acknowledgments caused by bursty transmission will cause the transmissions in the next RTT also to be bursty. This has been verified with simulations and real Internet tests [11].

Can we bridge the gap mentioned above by reconciling these two types of work? Can we analytically capture sub-RTT phenomena, especially sustainable burstiness patterns, with a closed loop model? Will this study provide useful new insight and solve previously unsolved problems? This paper is devoted to these questions and answers "yes" to all of them.

It turns out that some microscopic factors are very important and can affect macroscopic properties including steady state bandwidth allocation and flow level stability. In order to explain these analytically, we solve two main difficulties sidestepped by existing work in this area. One is modeling window flow control with sub-RTT factors and the other is to properly capture the closed loop effects.

The paper is organized as follows. Section II describes the mathematical model, which captures sub-RTT behaviors. We then show its impact by applying it to analysis of both loss-based and delay-based protocols. In Section III, we focus on steady state analysis. By using a simple solution to our model, we analytically predict unfair bandwidth allocation between paced and unpaced TCP Reno when they share a link. In Section IV, we shift our attention to dynamics. Using our new model, we show that the stability of FAST TCP [30] is dependent on the heterogeneity of the RTTs of the flows. This leads to the first discovery of unstable patterns of FAST TCP and also stability conditions that are usually satisfied in practice. Section V discusses some possible extensions.

II. MODEL

Consider a single bottleneck link with capacity c shared by multiple TCP flows, which are indexed by $i = 1, \ldots, N$.

The following variables are used throughout the paper. All are functions of time t; when written without explicit time dependence, they denote equilibrium values.

p: queuing delay at the link.
w_i: window size[1] of flow i.
x_i: arrival rate of flow i's data at the queue.
d_i: propagation delay for flow i.
τ_i: round trip delay (RTT) for flow i, $\tau_i = d_i + p$.

A model for a congestion control system must specify two things: (a) the TCP window control algorithm (e.g., TCP Reno in Section III and FAST TCP in Section IV) which determines how the congestion signal affects the window and (b) how the congestion signal, based on the queue length, evolves in

---

[1]We assume this is also the number of packets in the network, although a sudden reduction in $w_i(t)$ will not instantly withdraw packets.

response to the window sizes. For (b), it is well accepted that the length of a FIFO queue integrates the difference between incoming traffic and the link capacity,

$$\dot{p}(t) = \frac{1}{c}\left(\sum_{i=1}^{N} x_i(t) - c\right). \tag{1}$$

It remains to find an equation to relate the window (which sources control) with the rate (which affects the queue). This is traditionally done by approximating the rate by the ratio between the corresponding window and RTT, i.e.,

$$x_i(t) = \frac{w_i(t)}{\tau(t)}. \tag{2}$$

Although (2) applies to the average over an RTT and is good for equilibrium analysis, it does not apply to the instantaneous quantities and so is not suitable for accurate analysis of dynamics and microscopic rate patterns. It has several fatal problems. First, there is ambiguity over whether to use $w_i(t-\tau)$ and/or $\tau(t-\tau)$ for $w_i(t)$ and $\tau(t)$. Second, it completely ignores ACK-clocking by letting rate directly follow the window's change. Third, it implicitly assumes the rate is almost uniform within one RTT while in reality, persistent sub-RTT burstiness is prevalent.

The solution is to capture ACK-clocking. The following equation, mentioned in passing in [19], expresses window flow control's goal of equating the packets in flight with the window. More precisely, the window at time $t + d + p(t)$ is the integral of all the data rates from t to $t + d + p(t)$. This is because for the data that is sent out at t, which experiences a queuing delay $p(t)$, its acknowledgment arrives at the source at $t + d + p(t)$, assuming without loss of generality that the delay from the source to the link is 0. Formally,

$$\int_{t}^{t+d+p(t)} x_i(s)ds = w_i(t + d + p(t)). \tag{3}$$

Note that (3) holds for every t and hence it defines a rate function $x(t)$ based on the window function.

Equations (1), (3) combined with the equation that describes the window control (e.g. equation (19) for FAST TCP) form our model. Extensive packet level simulations in a companion paper [9] show the accuracy of the model and its superiority to existing models. In this paper, we apply it to find how some microscopic factors affect certain macroscopic properties.

III. APPLICATION TO LOSS BASED PROTOCOLS: BURSTINESS AFFECTS STEADY STATE

A basic assumption of optimization flow control [18] is that all flows sharing a link see the same congestion signal at that link. However, this may not be the case, especially for loss-based protocols using droptail routers. When a loss event occurs (several packet losses within one RTT due to the overflow of the buffer), some flows may escape losing packets, and will not detect the event. Current TCP Reno halves its window once in any RTT in which packets are lost, and thus the window is determined by the frequency of detecting loss events, rather than the expected number of packets lost.

The probability that a flow will detect a given loss event, called the synchronization rate, is used to analyze this effect

in [2] and further in [16], [17], [23]. It was demonstrated in [16] that flows with lower synchronization rates observe higher throughput, even if they have the same RTT.

Early models [2], [16], [23], take the synchronization rate as a given parameter. In this section, we instead calculate it based on the instantaneous rates. This closed loop analysis shows the effect of microscopic burstiness on steady state throughputs. In particular, it is shown that flows with the same macroscopic fluid equation but different microscopic burstiness patterns (TCP Reno and Paced TCP[2]), receive not only different instantaneous rates but also different steady state rates[3]. This is in stark contrast to the prediction from standard macroscopic analysis, where these flows should have the same equilibrium rates. Similar analysis can be used to help explain the unfairness between TCP Reno and TFRC [22], [27]. To simplify the presentation, we study the case where all flows have the same RTT.

A. Basic Setting

Consider a single-bottleneck network shared by a set U of N_u unpaced TCP Reno flows and a set P of N_p paced TCP Reno flows, with identical propagation delays d. Let $N = N_u + N_p$ be the cardinality of $A = U \cup P$. Let $\Sigma_u(t) = \sum_{i\in U} w_i(t)$, $\Sigma_p(t) = \sum_{i\in P} w_i(t)$ and $\Sigma_a(t) = \Sigma_u(t) + \Sigma_p(t)$. Assume the buffer is large enough that it never empties[4]. The total RTT $\tau(t) = \Sigma_a(t)/c$ varies with changes in window sizes.

At any time, the instantaneous rate of paced flow $i \in P$ is

$$x_i(t) = w_i(t)/\tau(t) = cw_i(t)/\Sigma_a. \tag{4}$$

Unpaced flows $i \in U$ transmit in bursts. During the portion $\Delta_i(t)$ of the RTT that flow i sends, it sends at rate $X(t) = c - \sum_{k\in P} x_k(t)$, giving instantaneous rates

$$x_i(t) = \begin{cases} X(t) = c(1 - \Sigma_p(t)/\Sigma_a) & t \in \Delta_i(t) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The total time during which flow i has non-zero rate is

$$w_i(t)/X(t) = (w_i(t)/c)(\Sigma_a(t)/\Sigma_u(t)). \tag{6}$$

This time may consist of several non-contiguous bursts. At each point in time, exactly one non-paced flow is transmitting.

This is illustrated in Figure 1. When the window sizes are constant, this pattern of burstiness is a solution to the model (1), (3), see [9] for details.

Loss occurs in the round trip during which the sum of the windows reaches the bandwidth-delay product plus the buffer size. The exact time of the loss is uniformly chosen within that round trip, and D packets are dropped simultaneously. This models a brief burst of cross traffic which the almost-full buffer is unable to handle. The probability that each loss is incurred by flow i is independent, and proportional to the instantaneous rate of flow i when the loss occurs.

---

[2]Paced TCP was proposed in late 90s; see e.g., [13]. It reduces burstiness by sending packets smoothly over an RTT, at a rate given by (2).

[3]This has already been observed and intuitively explained [1]. Our contribution is to provide a closed loop analytical explanation

[4]Paced TCP seeks to improve throughput by avoiding bursts too large for the buffers. However, for moderate to large buffers, Paced TCP flows actually gets smaller throughput than unpaced flows
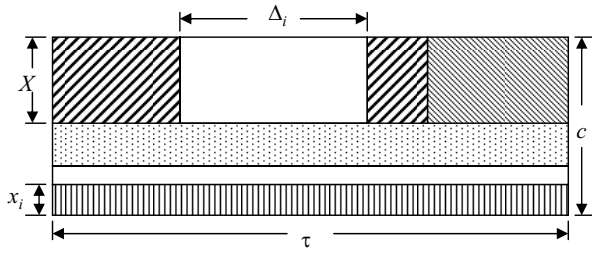
Fig. 1. Flow rates vs. time for the model of loss. Paced flows transmit at constant rate for the entire RTT, τ, while unpaced flows transmit in bursts. A single unpaced flow may have several bursts, as the diagonally shaded flow. The set of times in RTT t during which flow i is sending is $\Delta_i(t)$.

The rationale for modeling them as simultaneous is that loss occurs when the buffer is too full to absorb bursty cross traffic. A brief peak in cross traffic, such as a flow in slow-start, will cause a number of losses nearly simultaneously.

The case where there are no unpaced flows corresponds to the model of [2], [23] in which the rate is assumed to be $w_i(t)/\tau(t)$. Conversely, the case where there are no paced flows corresponds to the totally unsynchronized model in which a flow's probability of experiencing loss is directly proportional to its rate [12], [18].

This is clearly an extreme model of burstiness for unpaced flows, and can exaggerate the difference between paced and unpaced flows. However, it suffices to demonstrate the need to model certain microscopic factors in order to determine even the most basic macroscopic properties, such as the steady state.

B. Synchronization Rates

The number L_i of losses experienced by flow $i \in P$ during one loss event is binomially distributed with parameters D and $x_i(t)/c$. The number L_i of losses experienced by flow $i \in U$ will be binomially distributed with parameters D and $X(t)/c$ if its rate is $X(t) \neq 0$, and 0 otherwise. Let $a_i = 0$ if $L_i = 0$ and $a_i = 1$ if $L_i > 0$, and let $q_i = \Pr[a_i = 1]$. Then

$$
q_i = \begin{cases} \left(1 - \left(\dfrac{\Sigma_p(t)}{\Sigma_a}\right)^D\right) \dfrac{w_i(t)}{\Sigma_u(t)} & i \in U \\[2ex] \left(1 - \left(1 - \dfrac{w_i(t)}{\Sigma_a}\right)^D\right) & i \in P. \end{cases} \quad (7)
$$

When no loss is occurring, each flow increases its window by 1, and each time a flow loses one or more packets simultaneously, it reduces its window by a factor 1/2.

We measure time in units of round trip time, which is equal for all flows, but not constant. Let the times of consecutive loss events be indexed by k; denote by $v[k]$ and $v[k^+]$ the value of any variable v respectively before and after the kth loss event, and $T[k]$ be the number of RTTs between congestion events k and $k + 1$. The evolution of the windows are then given by

$$
w_i[k + 1] = b_i[k]w_i[k] + T[k] \quad (8)
$$

where $b_i[k] = (1 - a_i[k]) + 0.5a_i[k]$.

Losses occur as soon as the number of outstanding packets equals the bandwidth delay product, giving

$$
\sum_{i \in A} b_i[k]w_i[k] + NT[k] = BDP = \Sigma_a[k]. \quad (9)
$$

Using the fact that $\sum_{i \in A} w_i[k] = \Sigma_a[k]$,

$$
T[k] = \frac{\sum_{i \in A}(1 - b_i[k])w_i[k]}{N} \quad (10)
$$

whence (8) becomes

$$
w[k + 1] = A[k]w[k] \quad (11)
$$

where $w[k]$ is a vector of windows at time k, and

$$
A[k] = \text{diag}(b_i[k]) + \frac{1}{N}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}(1 - b_1[k], 1 - b_2[k], \ldots, 1 - b_N[k]) \quad (12)
$$

is a random matrix whose distribution depends on $w[k]$.

C. Steady State Analysis

As in [17], the loss indicator $b_i[k]$ in this model depends on $x_i[k]$, but now the relationship between $x_i[k]$ and $w_i[k]$ depends on whether or not i is paced, with $x_i[k]$ not proportional to $w_i[k]$ for unpaced flows. We assume the Markov chain (11) admits a unique and globally attractive invariant distribution and is ergodic. The following generalizes a result in [17], which says the expectation of the window and the synchronization rate is an invariant. Its intuitive consequence which will be proved later is that flows with higher synchronization rates will have smaller windows and therefore rates on average.

Theorem 1. *Under (11), for all $i, j \in A$,*

$$
\mathbb{E}[w_i[k]q_i[k]] = \mathbb{E}[w_j[k]q_j[k]]. \quad (13)
$$

Proof: From equations (11) and (12) and $b_i[k] = (1 - a_i[k]) + 0.5a_i[k] = 1 - 0.5a_i[k]$, we have

$$
w[k + 1] = w[k] - 0.5\left(I - \frac{1}{N}E\right)\text{diag}(a[k])w[k]
$$

where E is a square matrix with all entries 1. Taking expectations of both sides and noting $\mathbb{E}[w[k + 1]] = \mathbb{E}[w[k]]$ as they follow the unique steady state distribution, gives

$$
\left(I - \frac{1}{N}E\right)\mathbb{E}[\text{diag}(a[k])w[k]] = 0
$$

which is equivalent to

$$
\frac{1}{N}E\,\mathbb{E}[\text{diag}(a[k])w[k]] = \mathbb{E}[\text{diag}(a[k])w[k]].
$$

Hence $\mathbb{E}[\text{diag}(a[k])w[k]]$ is an eigenvector corresponding to the eigenvalue 1 for $(1/N)E$, which is $\kappa[1, 1, \ldots, 1]^T$ where κ is a constant. Thus $\mathbb{E}[w_i[k]q_i[k]] = \mathbb{E}[w_i[k]a_i[k]] = \mathbb{E}[w_j[k]a_j[k]] = \mathbb{E}[w_j[k]q_j[k]]$ for all $i, j \in A$. ∎

We now show that TCP Reno and Paced TCP don't share the link fairly.

Theorem 2. *For all $i \in U$, all $j \in P$ and all $w \in (0, \Sigma_a)$,*

$$
q_i(w)|_{i \in P} > \mathbb{E}[q_i(w)|i \in U]. \quad (14)
$$

Proof: Consider a flow sending for time $\delta \leq \tau$ per RTT at rate $w/\delta < c$, and rate 0 otherwise. The probability of this flow *not* experiencing a loss is

$$
P(w, \delta) = \left(1 - \frac{\delta}{\tau}\right) + \frac{\delta}{\tau}\left(1 - \frac{w}{c\delta}\right)^D.
$$

Note that $(r(1-(1/r))^D - r)$ is decreasing in r for all $D > 1$ and $r \in (0,1]$, by applying Taylor's theorem for $(1-u)^D$ with $u = 1/r$ to the derivative. Setting $r = c\delta/w$ gives, for any w, $P(w,\delta) > P(w,\tau)$ for all $\delta < \tau$.

An unpaced flow will always transmit with such a pattern with some $\delta < \tau$, while a paced flow corresponds to $\delta = \tau$. Thus $q_p(w) := q_i(w)|_{i \in P} > q_u(w) := \mathbb{E}[q_i(w)|i \in U]$. ∎

Theorem 2 shows that unpaced TCP Reno sees fewer loss events, and thus can be expected to achieve higher rate. Figure 2 compares the actual ratio of window sizes at loss events for paced and unpaced flows as predicted by (11), as N_u and N_p increase equally, for $D = 2, 3, 4$. It also shows the ratios of rates obtained from NS simulations with propagation delays of $50\,$ms and $200\,$ms, link capacity of $8.3\,$pkts/ms and a buffer of the product of the capacity and propagation delay, averaged over 10 experiments. It shows that this simple model roughly captures the initial decrease in fairness as N_u and N_p increase, and the leveling off after about 4 flows.

To see why Figure 2 asymptotes to a constant, note that if $N \gg D$ then $Dw_i[k]/\Sigma_a \ll 1$ for $i \in P$, giving

$$1 - (1 - w_i[k]/\Sigma_a)^D \approx Dw_i[k]/\Sigma_a \qquad (15)$$

and if N_u is large, then $\Sigma_u[k] \approx N_u\mathbb{E}[w_i]$, whence by (13)

$$\left.\frac{D}{\Sigma_a}\mathbb{E}[w_i^2]\right|_{i \in P} \approx \left.\frac{1 - (1 - N_u\mathbb{E}[w_i]/\Sigma_a)^D}{N_u\mathbb{E}[w_i]}\mathbb{E}[w_i^2]\right|_{i \in U} \qquad (16)$$

In [17] it was shown that $\mathbb{E}[w_i^2] \approx (\mathbb{E}[w_i])^2$ when synchronization is low. In the above model, this approximation applies as synchronization is limited by the burstiness of the unpaced flows. This gives an algebraic equation relating $W_u = \mathbb{E}[w_i]$ for $i \in U$ with $W_p = \mathbb{E}[w_i]$ for $i \in P$. If also $(1 - N_uW_u/\Sigma_a)^D \ll 1$ the equation is a quadratic in W_p,

$$\frac{D}{\Sigma_a}W_p^2 \approx \frac{W_u}{N_u} = \frac{\Sigma_a}{N_u}\left(\frac{\Sigma_a - N_pW_p}{N_u}\right), \qquad (17)$$

valid for $\log(1/(1 - N_uW_u/\Sigma_a)) \ll D \ll N$, with solution

$$\frac{W_p}{W_u} \approx \frac{\sqrt{1 + 4D(N_u/N_p)^2} - 1}{2DN_u/N_p - N_p/N_u - \sqrt{(N_p/N_u)^2 + 4D}}. \qquad (18)$$

This depends on N_u and N_p only through N_u/N_p.
Remark: 1. The model depends on the number of drops, D. We currently take D as a constant, but it could be drawn from a distribution.
2. In practice, drops are spread over an interval. This, and the fact that non-identical RTTs will cause bursts to spread, mean that real networks will show less bias against paced flows than this model predicts, especially when D is large. □

Until now, we have focused on mean throughput (window) behavior. The model actually also provides information on steady state throughput (window) distributions. Figure 3 exemplifies the distribution of window sizes of paced and unpaced flows, with $N_p = 3$, $N_u = 3$, total bandwidth-delay-product of $\Sigma_a = 2700$ packets and $D = 6$, based on simulation of (11). This confirms the significant sustained unfairness. It also demonstrates that the window of an unpaced flow has a larger variance, which matches intuition.
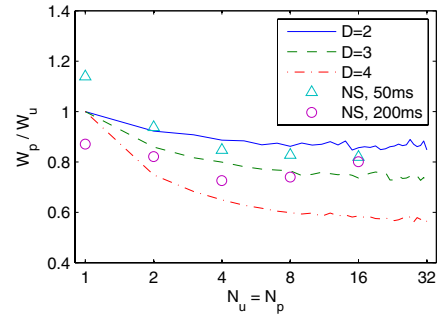

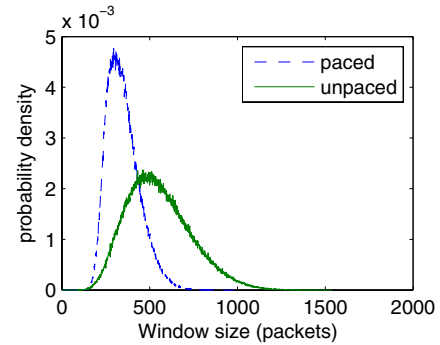Fig. 2.   Ratio of window sizes for paced and unpaced flows, and ratios of simulated rates.

Fig. 3.   Histogram of window sizes for paced and unpaced TCP Reno flows.

IV. APPLICATION TO DELAY BASED PROTOCOLS: RTT RATIOS DETERMINE STABILITY

Among the research work on congestion control, besides steady state analysis, another line of work of fundamental interest is the dynamics of congestion control algorithms. In particular, stability despite feedback delay is required in order to ensure that the system operating point is indeed the intended equilibrium, with the desired equilibrium properties, such as efficiency and fairness. However, due to the lack of an accurate model whose prediction can be verified by packet level simulations, there has been certain confusion in this area. Take FAST TCP as an example; existing experiments always showed it to be stable regardless of feedback delay [30] while analysis has made diverse predictions (see [24]).

In this section, by applying the new model to this problem, we show that for *any* step size γ, there is a (possibly pathological) network in which FAST is unstable, as verified by both analysis and packet level simulations. We finally also provide practical conditions which guarantee FAST to be stable.

A. Model of FAST and the network

FAST TCP is an algorithm which aims at improving TCP Reno's performance especially for networks with large bandwidth delay products [30]. FAST sets the congestion window based on the queuing delay, $p_i(t - \tau)$, seen by the packets. Its continuous time form is

$$\dot{w}_i(t) = -\gamma\frac{p_i(t-\tau)}{(d_i + p_i(t-\tau))^2}w_i(t) + \gamma\frac{\alpha_i}{d_i + p_i(t-\tau)} \qquad (19)$$

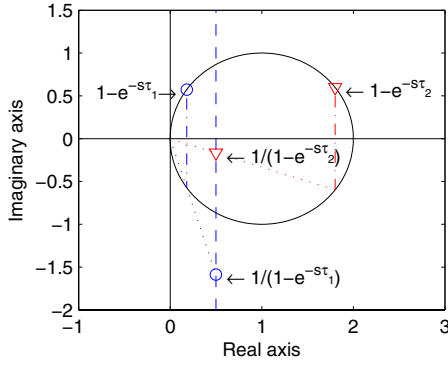Fig. 4. Construction of $T(j\phi)$. The circle is $1 - e^{-j\phi}$; $T(j\phi)$ is where the line joining the conjugate of $1 - e^{-j\phi}$ to the origin cuts $\mathrm{Re}(z) = 1/2$. The circles show $j\phi = s\tau_1$, and the triangles show $j\phi = s\tau_2$. The dash-dot line joins $1 - e^{-j\phi}$ to its conjugate, and the dotted line joins $T(j\phi)$ to the origin.

where $\gamma \in (0, 1]$ is a step size and α_i is a constant measured in packets. Since we are interested in local stability, we linearize and take the Laplace transform of (19). That yields

$$\left(s + \gamma \frac{p}{\tau_i^2}\right) w_i(s) = -\gamma \frac{\alpha_i d_i}{q \tau_i^2} p_i(s) e^{-s\tau_i}, \qquad (20)$$

while the linearization of the network model (1), (3) gives

$$\left(c + \sum_{i=1}^{N} x_i \frac{e^{-s\tau_i}}{1 - e^{-s\tau_i}}\right) p(s) = \sum_{i=1}^{N} \frac{w_i(s)}{1 - e^{-s\tau_i}}. \qquad (21)$$

Combining (20) and (21) gives a negative feedback system with an open loop transfer function

$$L(s) = \sum_{i=1}^{N} \mu_i L_i(s) \qquad (22a)$$

where

$$\mu_i = \frac{\alpha_i}{cq} = \frac{\alpha_i}{c \sum_{n=1}^{N} \alpha_n} = \frac{x_i}{c} \qquad (22b)$$

$$L_i(s) = \frac{d_i \gamma e^{-s\tau_i}}{\tau_i^2 s + \gamma q} \frac{T(s\tau_i)}{\sum_{n=1}^{N} \mu_n T(s\tau_n)} \qquad (22c)$$

$$T(s\tau) = \frac{1}{1 - e^{-s\tau}}. \qquad (22d)$$

Note that $\mathrm{Re}(T(j\phi)) = 1/2$ for all ϕ. Define y such that

$$\frac{1}{2} + jy(\phi) := \frac{1}{1 - e^{-j\phi}} = \frac{1}{2} - j \frac{\sin(\phi)}{2(1 - \cos(\phi))}. \qquad (23)$$

Geometrically, $T(j\phi)$ is the intersection between the line $\mathrm{Re}(z) = 1/2$ and the line passing through the origin and the point $1 - e^{+j\phi}$, as shown in Figure 4. As a consequence, $\hat{T} = \sum_{i=1}^{N} \mu_i T(s\tau_i)$ also lies on the line $\mathrm{Re}(z) = 1/2$. In particular, as ω increases from 0 to $2\pi/\tau_{\max}$, $\hat{T}$ monotonically traces the line from $1/2 - j\infty$ to $1/2 + j\infty$.

We now focus on the case of $q \to 0$, in which

$$L(s) = \gamma \sum_{i=1}^{N} \mu_i \frac{T(s\tau_i) e^{-s\tau_i}/s\tau_i}{\sum_{n=1}^{N} \mu_n T(s\tau_n)}. \qquad (24)$$

This model will now be used to predict two previously unknown modes of instability of FAST.

B. Instability due to RTT heterogeneity

Since each individual flow does not have complete knowledge of the network, we would like to be able to set FAST's parameters, such as γ, so that it will be stable in *all* networks. In the following, we show that that is impossible. For a given γ, we construct a network carrying two flows with very different RTTs such that FAST is unstable.

For two flows with equal α, $\mu_1 = \mu_2 = 1/2$, so the loop gain (24) reduces to

$$L(s) = \gamma \frac{\dfrac{e^{-s\tau_1}}{s\tau_1}(1 - e^{-s\tau_2}) + \dfrac{e^{-s\tau_2}}{s\tau_2}(1 - e^{-s\tau_1})}{2 - e^{-s\tau_2} - e^{-s\tau_1}}. \qquad (25)$$

The crux of the proof is that the loop gain becomes very large near $\omega = 2\pi/(\tau_1 + \tau_2)$ as the heterogeneity increases. Let $\lambda = \tau_1/(\tau_1 + \tau_2)$. Let $\omega(\tau_1, \lambda, \beta)$ be the solution to

$$\omega\tau_2 = 2\pi - \omega\tau_1 - \beta, \qquad (26)$$

with $\omega\tau_1, \omega\tau_2 \in (0, 2\pi)$ and $\beta \in (0, \omega\tau_1)$. Let

$$L_\lambda(\beta) = L(j\omega(\tau_1, \lambda, \beta)) \qquad (27)$$

$$= \gamma \frac{\dfrac{e^{-j\omega\tau_1}}{j\omega\tau_1}(1 - e^{j(\omega\tau_1 + \beta)}) + \dfrac{e^{j\omega\tau_1} e^{j\beta}(1 - e^{-j\omega\tau_1})}{j(2\pi - \omega\tau_1 - \beta)}}{2 - e^{j\omega\tau_1} e^{j\beta} - e^{-j\omega\tau_1}} \qquad (28)$$

where each ω in (28) refers to $\omega(\tau_1, \lambda, \beta)$.

The following lemma is proved in the appendix.

Lemma 3. *Let $\omega^* = 2\pi/(\tau_1 + \tau_2)$. Then (25) satisfies*

$$\mathrm{Im}(L(j\omega^*)) > 0. \qquad (29)$$

Let $\beta = (2\pi\lambda)^3$. For $0 < \lambda < 1/(2\pi)^2$, (27) satisfies

$$\mathrm{Im}(L_\lambda(\beta)) < 0 \qquad (30)$$

and for all $\omega \in [\omega^* - \beta/\tau_1, \omega^*]$, (25) satisfies

$$|L(j\omega)| \geq \frac{\gamma}{411\lambda^2}. \qquad (31)$$

For all $\omega > 0$, the general case (21), and hence (25), satisfies

$$\frac{d}{d\omega} \arg(L(j\omega)) \leq 0. \qquad (32)$$

The primary result of Lemma 3 is (31) which shows that there is an arbitrarily large positive feedback near ω^*. To see where this arises, consider the limit as $\omega\tau_1 \to 0$. As $\tau_2 \gg \tau_1$, the second term in the numerator of (28) is small, and so to first order,

$$L_\lambda(\beta) \approx \gamma \frac{(-1 + \cos(\omega\tau_1))/(j\omega\tau_1) - (\sin(\omega\tau_1))/(\omega\tau_1)}{2 - 2\cos(\omega\tau_1) - j\beta}$$

$$\approx \gamma \frac{1}{\omega\tau_1} \frac{-\omega\tau_1 + j(\omega\tau_1)^2/2}{(\omega\tau_1)^2 - j\beta} \qquad (33)$$

using $\sin(a + \beta) - \sin(a) \approx \beta$ for small a, β. The large gain results from the cancellation of the imaginary part of the denominator, by the τ_2, leaving only $j\beta$. Physically, this is because there is feedback on the timescale of $\tau_2 \gg \tau_1$; the feedback gain should normally be reduced in proportion to the feedback delay [21], but flow 1 scales its gain in proportion to its own, much smaller, RTT.
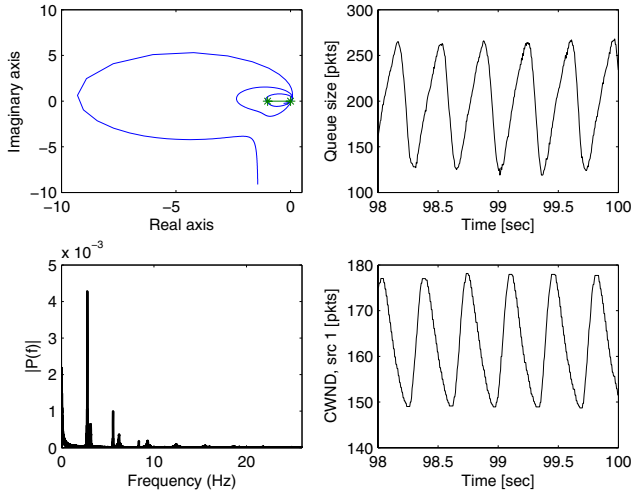
542

Fig. 5. Instability of FAST due to heterogenous RTTs: $d_1 = 10$ ms, $d_2 = 303$ ms. Top left: Nyquist plot of loop gain. Top right: Bottleneck queue size. Lower left: Magnitude spectrum (FFT) of queue size, without DC component. Lower right: Window size, source 1.

Theorem 4. *For all $\gamma > 0$ and τ_1 there exists a $\tilde{\tau}$ such that for all $\tau_2 > \tilde{\tau}$ the model (25) is unstable.*

Proof: By (32) and the Nyquist criterion, (25) is unstable if and only if $L(j\omega) \in (-\infty, -1)$ for some ω, since low frequency behavior is as in [21].

Let $\tilde{\tau} = \max(\sqrt{411/\gamma}, (2\pi)^2)$. For any $\tau_2 > \tilde{\tau}$, the right hand side of (31) exceeds 1, and (29) and (30) of Lemma 3 hold. By (29) and (30), $L(j\omega)$ crosses the real axis for some $\omega \in [\omega^* - \beta/\tau_1, \omega^*]$, and by (32), this crossing must be clockwise and must be a crossing of the negative real axis. By (31) of Lemma 3, this crossing must be to the left of $-1 + j0$, proving the instability. ∎

The following numerical results support Theorem 4. This is the first example to show instability of FAST TCP; previous work failed to show this as it did not explore cases with sufficient heterogeneity in feedback delays [30].

Example 1: *Instability due to Heterogeneous RTTs*
Consider two FAST flows with 1040 byte packets sharing a 200 Mbit/s bottleneck, with $d_1 = 10$ ms and $d_2 = 303$ ms and FAST parameters $\gamma = 0.5$ and $\alpha = 100$ [30]. The Nyquist plot of (25) in Fig. 5 encircles -1, indicating instability. NS-2 simulations, reported in the three remaining plots, show that there is indeed sustained oscillation at around $1/d_2 \approx 3$ Hz. The variation in *window* size shows that this is not simply packet-level sub-RTT burstiness. For this and Example 2, FAST's multiplicative increase mode was disabled.

Similarly, consider two FAST flows with 1500 byte packets sharing a 1 Gbit/s bottleneck, with $d_1 = 6$ ms and $d_2 = 130$ ms and FAST parameters $\gamma = 0.5$ and $\alpha = 30$ packets. This system was implemented in WAN-in-Lab [15]. Fig. 6 shows that there is again sustained oscillation of over 50 packets at around $1/d_2 \approx 8$ Hz, indicating instability.

C. Instability due to RTT synchronization

The precise timing of the model (24) allows another mode of instability, now to be described. This instability, which is
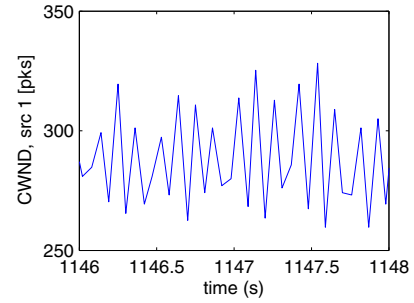

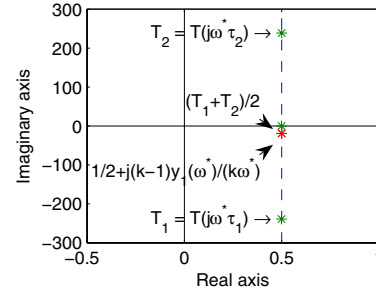Fig. 6. Window size, source 1: $d_1 = 6$ ms, $d_2 = 130$ ms.

Fig. 7. Values of $T(j\omega\tau) = 1/(1 - e^{-j\omega\tau})$ for $\mu_1 = \mu_2 = 1/2$, and $\tau_1 = 100.1$ ms, $\tau_2 = 200$ ms, $k = 2$. The numerator of $L(s)$ is bounded below in magnitude by $|y_1(\omega^*)/(2\omega^*)| \approx 19$, which is 4π times smaller than $|y_1(\omega^*)|$, but much larger than the denominator, $(T_1 + T_2)/2 = 1/2$.

confirmed by NS simulation, is expected not to appear in real networks with jitter as explained in Section IV-D (like the phase effect artifact [5]), but affects formal stability proofs.

Consider a network (24) with $N = 2$ flows with equal rate, $\mu_1 = \mu_2 = 1/2$ and RTTs $\tau_1 = 1 + \delta$ and $\tau_2 = k$, $k = 2, 3, \ldots$. If δ is sufficiently small, this system oscillates with a frequency near $\omega^* \approx 2\pi(1 - \delta/(k+1))$, as will now be shown.

Recall that the individual terms $T(j\omega\tau_i) = 1/2 + jy_i(\omega)$ lie on the line $\{z \in \mathbb{C}|\text{Re}(z) = 1/2\}$, as does their convex combination in the denominator of (24). Each of these points moves vertically upwards as ω increases, at a rate depending on τ_i, with all starting from $1/2 - j\infty$ at $\omega = 0$. For $\delta < \pi/k$, $y_1(\omega)$ is negative for $\omega \in (2\pi/(1+\delta), 2\pi)$ and takes all values less than $t_1(2\pi)$. Similarly, $y_2(\omega)$ takes all values greater than $y_2(2\pi/(1+\delta))$ in that interval. Thus, for some ω^*, $y_1(\omega^*) + y_2(\omega^*) = 0$ making the denominator of $L(s)$ equal 1/2, as illustrated in Figure 7. However, since $\omega^* \approx 2\pi$, for $\delta \ll 1$ the numerator of (24) is approximately

$$\frac{(k+1)/2 + j(k-1)y_1(\omega^*)}{jk\omega^*}.$$

Since $y_1(\omega^*) < y_1(2\pi) \to -\infty$ as $\delta \to 0$, the magnitude of $L(s)$ can be made arbitrarily large by taking δ small. Moreover, this large value occurs when $L(s)$ lies approximately on the negative real axis. This causes the Nyquist curve to encircle $-1 + j0$, and the system to be unstable.

In contrast to the previous mode of instability, which was caused by the short-RTT flow over-reacting to congestion due to very heterogeneous RTTs, this new mode is caused by precise cancellations when the ratio of the RTTs is approximately rational [9]. In the previous example, the Nyquist curve
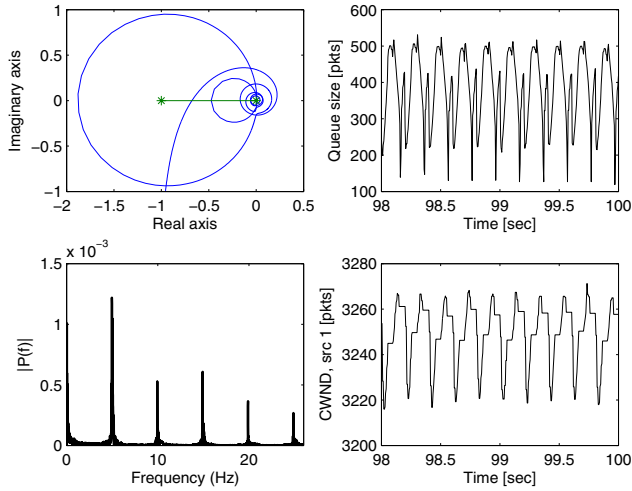
Fig. 8. Instability of FAST due to synchronized RTTs: $\tau_1 = 101\,\text{ms}$, $\tau_2 = 200\,\text{ms}$. Subfigures as for Figure 5.

encircles $-1 + j0$ in its first loop around 0. This time the Nyquist curve encircles $-1 + j0$ in the kth loop around 0, when the cancellation occurs. This corresponds to a frequency the reciprocal of the smaller RTT, rather than the larger. The following NS-2 simulation shows instability due to this mechanism.

Example 2: *Instability due to precisely matched RTTs*
Consider two FAST flows with $\gamma = 1$ and $\alpha = 200$ packets, sharing a bottleneck with $c = 500\,\text{Mbit/s}$, with $d_1 = 94.4\,\text{ms}$ and $d_2 = 193.4\,\text{ms}$ giving $\tau_1 = 101\,\text{ms}$ and $\tau_2 = 200\,\text{ms}$. The Nyquist plot of Fig. 8 again predicts instability, and NS-2 also exhibits oscillation. As well as the predicted peak at 10 Hz, there is one at 5 Hz, because the implementation of FAST freezes the window every alternate RTT, as seen in the bottom right hand figure, introducing severe nonlinearity.

D. Sufficient conditions for stability

With these two mechanisms which can cause instability for arbitrarily small gain γ, one might despair of finding any conditions under which (24) can be shown to be stable. However, it turns out that it is stable when there are many flows, or in the realistic case of networks with slight jitter.

Let $\mu : (0, \tau_{\max}] \mapsto \mathbb{R}^+ \cup \{\infty\}$ be the distribution of RTTs, weighted by the rate of flows with each RTT; for finitely many flows, this is a sum of impulses weighted by the discrete μ_i. Then sums weighted by μ_i are replaced by integrals, denoted

$$\mathbb{M}[f(\tau)] := \int_0^{\tau_{\max}} f(\tau)\mu(\tau)\,d\tau. \qquad (34)$$

Further, let $\text{sinc}(\theta) = \sin(\theta)/\theta$.

Define $H(\omega)$ as the half plane under the line through $-1 + j0$ with complex argument $\frac{\pi}{2} - \arg\left(\sum_{n=1}^N \mu_n T(s\tau_n)\right)$. Formally,

$$H(\omega) = \left\{ x \,\middle|\, \arg(x+1) - \frac{\pi}{2} + \arg\left(\sum_{n=1}^N \mu_n T(s\tau_n)\right) \in (-\pi, 0) \right\}. \qquad (35)$$
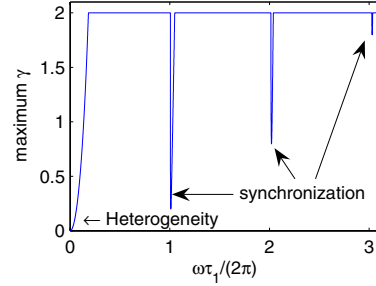

Fig. 9. Maximum γ for which (36) holds for $\tau_1 = 101\,\text{ms}$, $\tau_2 = 200\,\text{ms}$.

Lemma 5. *If*

$$\gamma\mathbb{M}[\text{sinc}(\omega\tau)] < \mathbb{M}[1 - \cos(\omega\tau)], \qquad (36)$$

then for $L(s)$ defined by (24),

$$L(j\omega) \in H(\omega). \qquad (37)$$

Proof: By definition, $L(j\omega) \in H(\omega)$ is equivalent to

$$\arg(L(j\omega) + 1) + \arg\left(\sum_{i=1}^N \mu_i T(s\tau_i)\right) \in (-\pi/2, \pi/2), \qquad (38)$$

or in other words,

$$\arg\left(\sum_i \mu_i \frac{\gamma e^{-j\omega\tau_i}}{j\omega\tau_i} T(j\omega\tau_i) + \sum_{i=1}^N \mu_i T(j\omega\tau_i)\right) \in (-\pi/2, \pi/2)$$

which is equivalent to

$$0 < \text{Re}\left(\mathbb{M}\left[T(j\omega\tau)(1 + \frac{\gamma e^{-j\omega\tau}}{j\omega\tau})\right]\right)$$

$$= \mathbb{M}\left[(1 - \cos(\omega\tau))(1 - \frac{\gamma\sin(\omega\tau)}{\omega\tau}) - \frac{\gamma\sin(\omega\tau)\cos(\omega\tau)}{\omega\tau}\right]$$

Thus, (38) is satisfied if (36) holds. ∎
Remark: The techniques used here, analogous to those discussed in [24], yield significantly tighter bounds than those in existing linear stability analysis of TCP, which is necessary for the analysis of this problem with the accurate model. □

Figures 9 shows the maximum γ for which (36) holds (truncated to 2). The frequencies at which instability occurred due to heterogeneity (example 1) and synchronization (example 2) are cases where (36) only holds for very small γ. Small perturbations of τ cause the "synchronization" dips vanish, but for any distribution of τ, (36) is violated at low frequency, ω. The next lemma shows that $L(j\omega)$ cannot cross $(-\infty, -1]$ in a low frequency region.

Lemma 6. *Let* $1/\hat{\tau} = \int (1/\tau)\mu(\tau)\,d\tau$,

$$\varphi(\omega) = \int \frac{\sin(\omega\tau)}{1 - \cos(\omega\tau)}\mu(\tau)\,d\tau \int \frac{\sin(\omega\tau)}{1 - \cos(\omega\tau)}\frac{\hat{\tau}}{\tau}\mu(\tau)\,d\tau \qquad (39)$$

and

$$\omega^* = \max\{\omega : \forall \omega \in (0, \omega^*), \varphi(\omega) > 1\}. \qquad (40)$$

Then $\omega^* \geq \pi/\tau_{\max}$ and $\text{Im}(L(j\omega)) < 0$ for all $\omega \in (0, \omega^*)$.

544

Proof: By (23), $y(\omega\tau) < y(\omega\tau_{\max}) \leq -1/2$ for all $\omega \in (0, \pi/\tau_{\max}]$, and $\tau \in (0, \tau_{\max})$. Then $\omega^* \geq \pi/\tau_{\max}$ since

$$\begin{aligned} \varphi(\omega) &= \int -2y(\omega\tau)\mu(\tau)\,d\tau \int -2y(\omega\tau)\frac{\hat{\tau}}{\tau}\mu(\tau)\,d\tau \\ &> \int \mu(\tau)\,d\tau \int \frac{\hat{\tau}}{\tau}\mu(\tau)\,d\tau \\ &= 1. \end{aligned}$$

If $K_1K_2 > K_r^2$ then $\mathrm{Re}([-K_r + jK_1]/[K_r + jK_2]) > 0$. Thus (40) implies that for all $\omega \in (0, \omega^*)$,

$$\begin{aligned} 0 &< \mathrm{Re}\left(\frac{-1 + j\int -2y(\omega\tau)(\hat{\tau}/\tau)\mu(\tau)\,d\tau}{1 + j\int -2y(\omega\tau)\mu(\tau)\,d\tau}\right) \\ &= -\omega\hat{\tau}\mathrm{Im}\left(\frac{\int(-(1/2) + j\mathrm{Im}(j\omega\tau))\mu(\tau)\,d\tau}{j\omega\tau\int((1/2) + j\mathrm{Im}(j\omega\tau))\mu(\tau)\,d\tau}\right) \\ &= -\omega\hat{\tau}\mathrm{Im}(L(j\omega)) \end{aligned}$$

whence $\mathrm{Im}(L(j\omega)) < 0$ for all $\omega \in (0, \omega^*)$. ∎

We can now show a sufficient condition on τ for FAST to be stable. It says that if there are many flows, with sufficiently uniformly spread RTTs, then FAST will be stable.

Theorem 7. *Let $0 < a < \tau_0$, $0 < h \leq 1/(2a)$ and let $\delta < 1$ be such that*

$$\delta > \mathrm{sinc}(a\pi/\tau_{\max}). \tag{41}$$

Then for any distribution $\mu(\tau) : (0, \tau_{\max}] \mapsto \mathbb{R}^+ \cup \{\infty\}$ with $\mu(\tau) > h$ for all $\tau \in (\tau_0 - a, \tau_0 + a)$, and for any

$$\gamma < 2ah(1 - \delta), \tag{42}$$

the transfer function $L(s)$ given by (24) is stable.

Proof: It is sufficient to show that the Nyquist curve does not intersect $(-\infty, -1]$. By Lemma 6, this does not happen for $\omega < \pi/\tau_{\max}$.

Since $(-\infty, -1] \cup H(\omega) = \emptyset$, it remains by Lemma 5 to show (36) holds for all $\omega \geq \pi/\tau_{\max}$. Since $1 - \cos(\omega\tau) \geq 0$,

$$\begin{aligned} \mathbb{M}[1 - \cos(\omega\tau)] &\geq h\int_{\tau_0-a}^{\tau_0+a} 1 - \cos(\omega\tau)\,d\tau \\ &= 2ah(1 - \cos(\omega\tau_0)\mathrm{sinc}(\omega a)). \end{aligned} \tag{43}$$

Similarly $\mathrm{sinc}(\theta) \leq 1$ and hence $\mathbb{M}[\mathrm{sinc}(\omega\tau)] \leq 1$.

Since $a\pi/\tau_{\max} < \pi/2$, and $\mathrm{sinc}(\theta) < \mathrm{sinc}(\pi/2)$ for $\theta > \pi/2$, it follows that $\mathrm{sinc}(\omega a) \leq \delta$ for all $\omega \geq \omega^* \geq \pi/\tau_{\max}$. Combining this with (36), (43) and $\cos(\omega\tau_0) \leq 1$ shows that (42) implies $L(s)$ is stable. ∎

Note that condition (42) is only a sufficient condition, and is loose for very peaked distributions μ. Typically $\mathbb{M}[\mathrm{sinc}(\omega\tau)] \ll 1$ for ω with $\cos(\omega\tau_0) \approx 1$ and vice versa.

V. CONCLUSION

We have studied window based flow control, the dominant congestion control mechanism in the current Internet, with a new model that naturally includes sub-RTT burstiness. This microscopic burstiness was analytically shown to have notable effects on macroscopic properties including both steady state bandwidth allocation and flow level dynamics. For loss based protocols, we analytically showed that paced and unpaced TCP flows, following the same window update rule, have different steady state throughputs. For delay based protocols, new modes of instability were identified, and it was demonstrated that the model is tractable enough to find regions of stability of practical importance. These results, ranging from equilibrium to dynamics, help bridge the gap between existing large bodies of work on fluid models and packet level studies.

There are several exciting directions in which to extend this work. For example, the number of flows is assumed to be fixed and it would be very interesting to see the effect of the new model on networks with dynamically arriving and departing flows [3]. Also, the analysis remains to be extended to general networks which can potentially have multiple congested links, on which we are optimistic as the model itself extends to general networks naturally as shown in a companion paper [9].

VI. APPENDIX

Lemma 3 follows from the following four lemmas.

Lemma 8. *Let $\omega^* = 2\pi/(\tau_1 + \tau_2)$. Then (25) satisfies*

$$\mathrm{Im}(L(j\omega^*)) > 0. \tag{44}$$

Proof: Let $\epsilon = \omega^*\tau_1$. Since $\omega^*\tau_2 = 2\pi - \omega^*\tau_1$, $\cos(\omega^*\tau_2) = \cos(\epsilon)$ and $\sin(\omega^*\tau_2) = -\sin(\epsilon)$, giving

$$L(j\omega^*) = \frac{j}{2}\left(\frac{1}{\epsilon} + \frac{1}{2\pi - \epsilon}\right) - \left(\frac{1}{\epsilon} - \frac{1}{2\pi - \epsilon}\right)\frac{\sin(\epsilon)}{2 - 2\cos(\epsilon)}.$$

Since $0 < \epsilon < 2\pi$, the imaginary part is positive. ∎

Lemma 9. *Let $\beta = (2\pi\lambda)^3$. For $\lambda < 1/(2\pi)^2$,*

$$\mathrm{Im}(L_\lambda(\beta)) < 0. \tag{45}$$

Proof: Let ω be the solution of (26). Note that $\omega\tau_1 < 2\pi\lambda$ and $\beta < \omega\tau_1$. If $L_\lambda(\beta)$ can be expressed as

$$\gamma\frac{N_r + jN_i}{D_r + jD_i} \tag{46}$$

with $N_r < 0$, $N_i > 0$, $D_r > 0$ and $D_i < 0$, then the lemma follows provided that $|N_i/N_r| < |D_i/D_r|$, or equivalently

$$\left|\frac{N_rD_i}{N_iD_r}\right| > 1. \tag{47}$$

Let the numerator of (28) be $N_r + jN_i$ with $N_r, N_i \in \mathbb{R}$. Then, using $\theta - \theta^3/6 \leq \sin(\theta) \leq \theta$ and $\cos(\theta) \leq 1 - \theta^2/2$, we have Moreover, using $1 - \cos(x) \leq x^2/2$ and $\omega\tau_1 < 2\pi\lambda < 1/(2\pi)$, we have $0 < N_i \leq 10\lambda/3$.

Let the denominator of (28) be $D_r + jD_i$ with $D_r, D_i \in \mathbb{R}$. Then, using $1 - \cos(x) \leq x^2/2$, $\beta = (2\pi\lambda)^3 \leq \lambda$ and $\omega\tau \leq 2\pi\lambda \leq 1/(2\pi)$, we have $0 < D_r \leq \lambda^2((2\pi)^2 + 3/2)$. Moreover, using $1 - \cos(x) < x^2/2$, $\sin(x) < x$, $\beta < 2\pi\lambda$

and $\omega\tau_1 < 2\pi\lambda \le 1/(2\pi)$, we have $D_i \le (2\pi\lambda)^3(-1 + 1/(2\pi)^2) < 0$.

Given the signs of the foregoing,

$$\left|\frac{N_r D_i}{N_i D_r}\right| \ge \frac{10/11}{10\lambda/3}\frac{(2\pi\lambda)^3(1 - 1/(2\pi)^2)}{\lambda^2((2\pi)^2 + 3/2)} > 1$$

as required. ∎

Lemma 10. *Let* $\beta = (2\pi\lambda)^3$ *and* $\omega^* = 2\pi/(\tau_1 + \tau_2)$. *For all* $\omega \in [\omega^* - \beta/\tau_1, \omega^*]$, *(25) satisfies*

$$|L(j\omega)| \ge \frac{\gamma}{411\lambda^2}. \tag{48}$$

Proof: Let $D(s) = (2 - e^{-s\tau_2} - e^{-s\tau_1})$,

$$N_1(s) = \frac{e^{-s\tau_1}}{s\tau_1}(1 - e^{-s\tau_2}), \qquad N_2(s) = \frac{e^{-s\tau_2}}{s\tau_2}(1 - e^{-s\tau_1})$$

and $\eta = \omega^* - \omega \in [0, \beta/\tau_1]$. By (25),

$$|L(j\omega)| \ge \gamma(|N_1(j\omega)| - |N_2(j\omega)|)/|D(s)|.$$

Using $\cos(\theta) = \cos(2\pi - \theta)$, $2\pi - (\omega^*\tau_2 - \eta\tau_2) = \omega^*\tau_1 + \eta\tau_2$, $\cos(\theta) \ge 1 - \theta^2/2$ and $\sin(\theta_1) - \sin(\theta_2) \le |\theta_1 - \theta_2|$ gives

$$\begin{aligned}&|D(j(\omega^* - \eta))|\\ &\le \left|\frac{(\omega^*\tau_1 - \eta\tau_1)^2 + (\omega^*\tau_1 + \eta\tau_2)^2}{2} + j\eta(\tau_1 + \tau_2)\right|.\end{aligned}$$

Further using $|x + jy| \le |x| + |y|$, $\lambda < \tau_1/\tau_2 < 1$, $\omega^*\tau_1 = 2\pi\lambda$ and $\eta \le (2\pi\lambda)^3/\tau_1$ gives

$$|D(j(\omega^* - \eta))| \le 406\lambda^2. \tag{49}$$

Similarly, since $\sin(\theta) = -\sin(2\pi - \theta)$ and $\sin(\omega^*\tau_1 + \eta\tau_2) > \sin(\omega^*\tau_1 - \eta\tau_1)$ for $0 < \omega^*\tau_1 - \eta\tau_1 < \omega^*\tau_1 + \eta\tau_2 < \pi/2$,

$$|N_1(j(\omega^* - \eta))| \ge 0.995. \tag{50}$$

Moreover, using $1 - \cos(x) \le x^2/2$, $\sin(x) \le x$, $\lambda = \omega^*\tau_1/(2\pi) < 1/(2\pi)^2$ and $\eta \le (2\pi\lambda)^3/\tau_1$,

$$|N_2(j(\omega^* - \eta))| \le 0.04. \tag{51}$$

Combining (49), (50) and (51) gives

$$|L(j(\omega^* - \eta))| \ge \frac{0.99\gamma}{406\lambda^2} \ge \frac{\gamma}{411\lambda^2}. \tag{52}$$

∎

Lemma 11. *For all* $\omega > 0$, *the* $L(s)$ *of (24) satisfies*

$$\frac{d}{d\omega}\arg(L(j\omega)) \le 0. \tag{53}$$

Proof: As ω increases, the numerator and denominator of (24) are piecewise continuous, with simultaneous jump discontinuities and singularities at $\omega = 2k\pi/\tau_i$ for $k = 1, 2, \ldots$. Except for $\omega = 0$, these singularities are removable, and $L(j\omega)$ is continuous. Thus it suffices to show that $\arg(L(j\omega))$ is decreasing where the denominator is continuous.

Substituting (23) into (24), gives

$$L(s) = \frac{\gamma}{j\omega}\frac{\sum_{n=1}^{N}(\mu_n/\tau_n)(jy(\omega\tau_n) - 1/2)}{\sum_{n=1}^{N}\mu_n(jy(\omega\tau_n) + 1/2)}.$$

As ω increases, $y(\omega\tau_n)$ increases, except at the singularities. Since the numerator is in the left half plane, its argument decreases, while the denominator is in the left half plane, and its argument increases. ∎

REFERENCES

[1] A. Aggarwal, S. Savage and T. Anderson. Understanding the performance of TCP pacing. In *Proc. IEEE Infocom*, 2000.

[2] F. Baccelli and D. Hong. AIMD, fairness, and fractal scaling of TCP Traffic. In *Proc. IEEE Infocom*, 2002.

[3] T. Bonald and L. Massoulie. Impact of fairness on Internet performance. In *Proc. ACM SIGMETRICS*, Jun. 2001.

[4] L. S. Brakmo and L. L. Peterson. TCP Vegas: End-to-end congestion avoidance on a global Internet. *IEEE J. Select. Areas Commun.*, 13(8):1465–1480, 1995.

[5] S. Floyd and V. Jacobson. On traffic phase effects in packet-switched gateways. *Internetworking: Research and Experience*, 3(3):115-156, Sep. 1992.

[6] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Networking*, 1(4):397–413, Aug. 1993.

[7] M. Gerla and L. Kleinrock. Flow control: A Comparative survey. *IEEE Trans. Comm.*, 28(4):553–574, Apr. 1980.

[8] V. Jacobson. Congestion avoidance and control. In *Proc. ACM SIGCOMM*, 1988.

[9] K. Jacobsson, L. L. H. Andrew, A. Tang, K. Johansson, H. Hjalmarsson and S. H. Low. ACK-clocking dynamics: Modeling the interaction between windows and the network. In *Proc. IEEE Infocom*, 2008.

[10] R. Jain. Congestion control in computer networks: Issues and trends. *IEEE Network Magazine*, 4(3):24–30, May 1990.

[11] H. Jiang and C. Dovrolis Why is the Internet traffic bursty in short time scales. In *Proc. ACM SIGMETRICS*, 2005.

[12] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *J. Op. Res. Soc.*, 49(3):237–252, Mar. 1998.

[13] J. Kulik, R. Coulter, D. Rockwell and C. Partridge. Paced TCP for High Delay-Bandwidth Networks. In *Proc. IEEE GLOBECOM*, 1999.

[14] S. Kunniyur and R. Srikant. End-to-end congestion control: Utility functions, random losses and ECN marks. *IEEE/ACM Trans. Networking*, 11(5):689 – 702, Oct. 2003.

[15] G. S. Lee, L. L. H. Andrew, A. Tang and S. H. Low. WAN-in-Lab: Motivation, deployment and experiments. In *Proc. PFLDnet*, pp 85-90, Marina Del Rey, CA, 2007.

[16] D.J. Leith and R. Shorten. Impact of drop synchronisation on TCP fairness in high bandwidth-delay product networks. In *Proc. PFLDnet*, Nara, Japan, 2006.

[17] S. Liu, T. Basar and R. Srikant. TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks. In *Proc. First Int. Conf. on Perform. Eval. Methodol. Tools (VALUETOOLS)*, 2006.

[18] S. H. Low and D. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Trans. Networking*, 7(6):861–874, 1999.

[19] J. Mo, R. La, V. Anantharam, and J. Walrand. Analysis and comparison of TCP Reno and TCP Vegas. in *Proc. IEEE INFOCOM*, 1999.

[20] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Networking*, 8(5):556–567, Oct. 2000.

[21] F. Paganini, Z. Wang, J. C. Doyle and S. H. Low. Congestion control for high performance, stability, and fairness in general networks *IEEE/ACM Trans. Networking*, 13(1):43-56, Feb. 2005.

[22] I. Rhee and L. Xu Limitations of equation-based congestion control. in *Proc. ACM SIGCOMM*, 2005.

[23] R. N. Shorten, F. Wirth, D. J. Leith. A positive systems model of TCP-like congestion control: Asymptotic results. *IEEE/ACM Trans. Networking*, 14(3):616-629, June 2006.

[24] A. Tang, K. Jacobsson, L. L. H. Andrew and S. H. Low An accurate link model and its application to stability analysis of FAST TCP. in *Proc. IEEE INFOCOM*, 2007.

[25] A. Tang, J. Wang, and S. H. Low. Counter-intuitive throughput behaviors in networks under end-to-end control. *IEEE/ACM Trans. Networking*, 14(2):355–368, Apr. 2006.

[26] A. Tang, J. Wang, S. H. Low, and M. Chiang. Equilibrium of heterogeneous congestion control: Existence and uniqueness. *IEEE/ACM Trans. Networking*, 15(4):824–837, Aug. 2007.

[27] M. Vojnovic and J. Boudec. On the long-run behavior of equation-based rate control *IEEE/ACM Trans. Networking*, 13(3):568–581, Jun. 2005.

[28] Z. Wang and J. Crowcroft. Eliminating periodic packet losses in the 4.3-Tahoe BSD TCP congestion control algorithm. *ACM SIGCOMM Comp. Commun. Rev.*, 22(2):9–16, Apr. 1992.

[29] D. Wei. Microscopic Behavior of Internet Congestion Control. Ph.D. Thesis, California Institute of Technology, 2007.

[30] D. Wei, C. Jin, S. H. Low, and S. Hegde. FAST TCP: motivation, architecture, algorithms, performance. *IEEE/ACM Trans. Networking*, 14(6): 1246–1259, Dec. 2006.