ACK-Clocking Dynamics: Modelling the Interaction between Windows and the Network

Krister Jacobsson*, Lachlan L. H. Andrew†, Ao Tang‡, Karl H. Johansson*, Håkan Hjalmarsson*, Steven H. Low†
*ACCESS Linnaeus Centre, Electrical Engineering, KTH, Stockholm, SE-100 44, Sweden
†California Institute of Technology, Pasadena, CA 91125, USA
‡Cornell University, Ithaca, NY 14853, USA

***Abstract*—A novel continuous time fluid flow model of the dynamics of the interaction between ACK-clocking and the link buffer is presented. A fundamental integral equation relating the instantaneous flow rate and the window dynamics is derived. Properties of the model, such as well-posedness and stability, are investigated. Packet level experiments verify that this new model is more accurate than existing models, correctly predicting qualitatively different behaviors, for example when round trip delays are heterogeneous.**

I. Introduction

The Transmission Control Protocol (TCP) is the predominant transport protocol of the Internet today, carrying about 83% of the total traffic volume [1]. Since Jacobson's work on the Tahoe release of BSD Unix in 1988 [2], many modifications and replacements have been proposed [2–9] to meet the demands of a modern Internet scaled up in size and capacity.

Most proposed algorithms are *window based*, meaning that a source explicitly controls a window size, that is the number of packets that are sent before the sender must wait for an acknowledgment packet. Research has focused on how to determine that window size.

There exist many experimental TCP proposals ranging between purely loss-based versions like CUBIC [5] and H-TCP [6], and purely delay based schemes like TCP Vegas [7] and FAST TCP [8], with many algorithms that use both delay and loss as congestion measures, such as TCP Africa [9] and TCP Illinois [10].

All of these rely on detailed dynamics of instantaneous rates and network queue sizes, either to determine which flow's packet is being received at the exact time a packet is dropped, or to determine the precise queuing delays. In window based schemes, ACK-clocking governs these sub-RTT phenomena. Despite its importance, the dynamics of the window mechanism is still not well understood.

A. Window-based transmission control

A schematic picture of the control structure for window-based transmission control is displayed in Fig. 1. The dynamics of the endpoint protocol are represented by the three blocks: transmission control, window control, and congestion estimator. The system consists of an inner loop and an outer loop. In the outer loop, the window control adjusts the transmission window size based on the estimated congestion level of the network. This congestion level is estimated based on the ACKs, which carry implicit (often corrupted) information in the form of duplicate, missing and delayed ACKs.

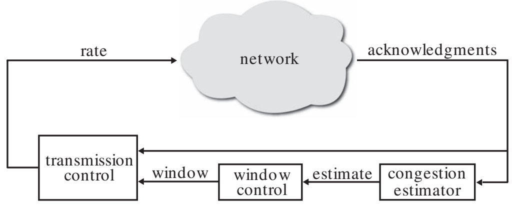

Fig. 1. System view in window-based congestion control.

B. ACK-clocking

The dynamics of the inner loop are given by so called ACK-clocking. The transmission of new packets is controlled or "clocked" by the stream of received ACKs by the transmission control. A new packet is transmitted for each received ACK, thereby keeping the number of outstanding packets, i.e. the window, constant. More sophisticated traffic shaping could also be considered, but we do not consider such dynamics in this paper.

The design of the outer loop, i.e., the window adjustment mechanism, has received ample attention in the literature [2–9] while the properties of ACK-clocking are often ignored. ACK-clocking operates at a per-packet time-scale. This makes it better suited to handle short-term queue fluctuations than the outer-loop. Furthermore, ACK-clocking has stabilizing properties in itself.

C. Network fluid flow modeling

To ensure that the network will reach and maintain a favorable equilibrium, it is important to assess its dynamical properties such as stability and convergence. Instability means that small fluctuations due to varying cross traffic are amplified, and manifests itself as severe oscillations in aggregate traffic quantities, such as queue lengths. Following the seminal work by Kelly [11] there have been numerous studies on network stability. Network fluid flow models, where packet level information is discarded and traffic flows are assumed to be smooth in space and time, have shown to be useful

978-1-4244-2026-1/08/$25.00 © 2008 IEEE

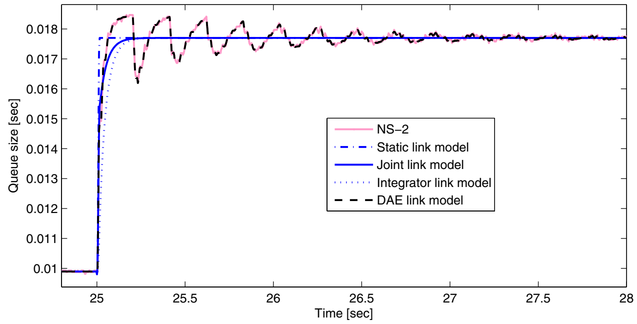

Fig. 2. The new model presented in this paper is able to capture a dynamic step response much better than traditional models in the literature. The graph shows the queue size. Two window based flows with propagation delays $d_1 = 10\,\text{ms}$ and $d_2 = 190\,\text{ms}$ are sharing the bottleneck link. The window of the first source is subject to a step after 25 s.

in such analysis, for example in [8,12–16]. The validity of results concerning dynamical properties, however, rely heavily on the accuracy of the models. Models with fundamentally different dynamical properties have been used to model ACK-clocking in window-based schemes (often referred to as "the link"), i.e., the inner loop in Fig. 1. In [12–15] an "integrator" link model is used, integrating the approximate excess rate on the link. On the other hand, in [17] transients are ignored and a "static" link model is proposed. Furthermore, in [18] a "joint" link model combining the immediate and long term integrating effect is proposed and used for stability analysis in [16]. The motivating example in Section II illustrates the limited accuracy of these models. This is further elaborated on in a companion paper [19], which highlights the need for incorporating important microscopic sub-RTT effects in macroscopic fluid flow models.

A new link model which captures these sub-RTT effects is derived in Section III. Properties of this model are found in Section IV, and it is rigorously validated in Section V. Conclusions are drawn in Section VI.

II. A MOTIVATING EXAMPLE

Consider a system of two window based flows sending over a single bottleneck link with capacity 150 Mbit/s, with 1040 byte packets, where the sources' window sizes are kept constant, i.e., the outer loop in Fig. 1 is disabled. The round trip delays excluding the queuing delay are $d_1 = 10\,\text{ms}$ and $d_2 = 190\,\text{ms}$. The window sizes are initially $w_1 = 210$ and $w_2 = 1500$ packets respectively. After convergence, at 25 seconds, w_1 is increased step-wise from 210 to 300 packets. The solid pink line in Fig. 2 shows the bottleneck queue size (in seconds) when this scenario is simulated in NS-2, exhibiting significant oscillation in the queue. This is in contrast to the dash-dotted, solid and dotted blue lines in Fig. 2, showing predictions made by existing models of the inner loop dynamics (see [16] for a discussion). They all predict smooth convergence similar to first order filter step responses (with varying time constants) and the reasons will be discussed in section IV-D. The dashed black line shows the continuous time fluid model derived in this paper, it shows almost perfect agreement with the packet level simulation, even at sub-RTT time scales.

Analyses based on the cruder models may give results that are qualitatively different than those for the more accurate model proposed here [19].

III. MODELLING

A. Preliminaries

Consider a single bottleneck link with capacity c and time varying queuing delay $p(t)$. Traffic consists of N flows, with $w_n(t)$ the time varying number of packets "in flight" (sent but not acknowledged). The instantaneous rate at which traffic from flow n enters the link is $x_n(t)$. The round trip time between the time a packet of flow n enters the link and the time that the "resulting" packet transmitted in response to the acknowledgment of that packet enters the link is denoted $\tau_n(t)$. It consists of a fixed component d_n and a time varying component due to queuing delay. In the single-link case, $\tau_n(t) = d_n + p(t)$.

The link carries cross traffic $x_c(t)$ which is not window controlled.

Packets are assumed to be transmitted greedily and in FIFO order at links, which reflects the reality of the current Internet.

Without loss of generality, forward propagation delay is assumed to be zero.

B. ACK-clocking model

In terms of rates, a link buffer is simply an integrator, integrating the excess rate at the link (modulo static non-linearities present in the system, such as non-negativity constraints or drop-tail queues). Thus, the buffer dynamics are naturally given by

$$\dot{p}(t) = \frac{1}{c}\left(\sum_{n=1}^{N} x_n(t) + x_c(t) - c\right). \tag{1a}$$

It remains to define the instantaneous rates $x_n(t)$.

1) Instantaneous rate: To discover what can be known about a source's instantaneous transmission rate x_n based on knowledge of the window size w_n, consider an arbitrary time t. Packets transmitted up to time t will be acknowledged by time $t + \tau_n(t)$, and thus the number of packets "in flight" at time $t + \tau_n(t)$, namely $w_n(t + \tau_n(t))$, will exactly equal the number of packets transmitted in the interval $(t, t + \tau_n(t)]$. That is, for $n = 1, \ldots, N$, we have the constraints

$$\int_t^{t+\tau_n(t)} x(T)dT = w_n(t + \tau_n(t)). \tag{1b}$$

(This equation was introduced in passing in [20], but not pursued.) The whole system is described by the delay *Differential Algebraic Equation* (DAE) defined by (1a) and (1b).

It may be convenient to reformulate the model (1) by, for example, differentiating the constraints (1b). Applying a variable transformation $\dot{B}_n(t) = x_n(t)$, working with accumulated packets B_n instead of rates, can also be useful. The model then reduces from a delay DAE to a recursive update law.

The model is supported by numerical results in Section V.

IV. Analysis

The model is shown in [21] to have a unique equilibrium, by showing that the equilibrium rate x^* maximizes the sum of the concave utility functions $U_n(x_n^*) = w_n \log(x_n^*) - d_n x_n^*$.

This section uses a linearization to prove that the queuing delays are asymptotically stable but the rates may have sustained oscillations.

A. Linearization around equilibrium

In order to study the stability, let us linearize (1) around its equilibrium (p, w, x, x_c). Following the convention that time delays in variables' arguments are modeled by their equilibrium values yields, for $n = 1, \ldots, N$,

$$\begin{aligned} \dot{p}(t) - \sum_{n=1}^{N} x_n(t)/c - x_c(t)/c &= 0, && \text{(2a)} \\ x_n \dot{p}(t) - \dot{w}(t + \tau_n) + x_n(t + \tau_n) - x_n(t) &= 0. && \text{(2b)} \end{aligned}$$

Here variables now denote small perturbations. Taking the Laplace transform gives an explicit expression for the sources' queue input rates

$$x_n(s) = \frac{s}{e^{-s\tau_n} - 1} \left(x_n e^{-s\tau_n} p(s) - w_n(s) \right). \tag{3}$$

Thus the linear ACK-clocking dynamics are described by

$$\left(c + \sum_{n=1}^{N} x_n \frac{e^{-s\tau_n}}{1 - e^{-s\tau_n}} \right) p(s) = \sum_{n=1}^{N} \frac{w_n(s)}{1 - e^{-s\tau_n}} + \frac{1}{s} x_c(s). \tag{4}$$

Modeling non-zero forward propagation delay, τ_n^f, is achieved simply by multiplying $w_n(s)$ by $e^{-s\tau_n^f}$ in (4). The linear model is validated in Section V-B and used for analysis below.

B. Stability

As pointed out in [2], window flow control is stable in the sense that signals remain bounded. The following theorem shows the stronger result that the linearized single bottleneck dynamics (4) relating the windows w to the queue p are asymptotically stable, ruling out persistent oscillations in these quantities, at least locally. Let $\mathbb{C}^+$ be the open right half plane, $\{z : \text{Re}(z) > 0\}$, and $\bar{\mathbb{C}}^+$ be its closure, $\{z : \text{Re}(z) \geq 0\}$.

Theorem 1: For all $0 < x_n \leq c$, $\tau_n > 0$, $n = 1, \ldots, N$, $\sum_n x_n \leq c$, the function $G_{pw} : \bar{\mathbb{C}}^+ \to \mathbb{C}^{1 \times N}$ whose ith element is given by

$$G_{pw_i}(s) = \frac{1}{(1 - e^{-s\tau_i}) \left(c + \sum_{n=1}^{N} x_n \frac{\exp(-s\tau_n)}{1 - \exp(-s\tau_n)} \right)}, \tag{5}$$

is stable.

Proof: It is sufficient to confirm that [22]:

(a) $G_{pw}(s)$ is analytic in $\mathbb{C}^+$;

(b) for almost every real number ω,

$$\lim_{\sigma \to 0^+} G_{pw}(\sigma + j\omega) = G_{pw}(j\omega);$$

(c) $\sup_{s \in \bar{\mathbb{C}}^+} \bar{\sigma}(G_{pw}(s)) < \infty$

where $\bar{\sigma}$ denotes the largest singular value.

Conditions (a) and (b) are satisfied if they hold elementwise. Furthermore

$$\sup_{s \in \bar{\mathbb{C}}^+} \bar{\sigma}(G_{pw}(s)) \leq \sum_{i=1}^{N} \sup_{s \in \bar{\mathbb{C}}^+} |G_{pw_i}(s)|. \tag{6}$$

Thus, condition (c) holds if

$$\inf_{s \in \bar{\mathbb{C}}^+} |1/G_{pw_i}(s)| > 0. \tag{7}$$

It is therefore sufficient to establish (a), (b) and (c) for the ith transfer function element $G_{pw_i}(s)$.

Start with the boundedness condition (c). It is sufficient to show that there is no sequence $s_l = \sigma_l + j\omega_l \in \bar{\mathbb{C}}^+$ with $\lim_{l \to \infty} |1/G_{pw_i}(s_l)| = 0$. This will be established by showing that the limit evaluated on any convergent subsequence is greater than 0. Consider a subsequence with $\sigma_l \to \sigma$, $\omega_l \to \omega$.
Case 1, $\sigma = \infty$: $1/G_{pw_i}(s_l) \to c > 0$.
Case 2, $\sigma \in (0, \infty)$: By the triangle inequality,

$$|1 - e^{-s_l\tau_i}| \geq \left|1 - |e^{-s_l\tau_i}|\right| \to 1 - e^{-\sigma\tau_i} > 0. \tag{8}$$

Furthermore, $1/(e^{s_l\tau_n} - 1)$ lies on the circle with center $1/(A_l^2 - 1) + j0$ and radius $A_l/(A_l^2 - 1)$, where $A_l = |e^{s_l\tau_n}|$. Thus $\lim_{l \to \infty} \text{Re}(1/(e^{s_l\tau_n} - 1)) \geq -1/(e^{\sigma\tau_n} + 1)$, hence

$$\lim_{l \to \infty} \text{Re} \left(c + \sum_{n=1}^{N} \frac{x_n}{e^{s_l\tau_n} - 1} \right) \geq c - \sum_{n=1}^{N} \frac{x_n}{e^{\sigma\tau_n} + 1} = \\ c - \sum_{n=1}^{N} x_n + \sum_{n=1}^{N} \frac{x_n e^{\tau_n\sigma}}{e^{\tau_n\sigma} + 1} \geq \sum_{n=1}^{N} \frac{x_n}{1 + e^{-\tau_n\sigma}} \geq \sum_{n=1}^{N} \frac{x_n}{2} > 0. \tag{9}$$

Multiplying (8) and (9) gives $\lim_{l \to \infty} |1/G_{pw_i}(s_l)| > 0$.
Case 3, $\sigma = 0$: Note that $\text{Re}(1/(e^{j\omega_l\tau_n} - 1)) = -1/2$, so

$$\lim_{l \to \infty} \text{Re} \left(c + \sum_{n=1}^{N} \frac{x_n}{e^{(\sigma_l + j\omega_l)\tau_n} - 1} \right) = c - \sum_{n=1}^{N} \frac{x_n}{2} > 0. \tag{10}$$

Thus $\lim_{l \to \infty} |1/G_{pw_i}(s_l)| \neq 0$ except possibly when the first factor of (5) $1 - e^{-s_l\tau_i} \to 0$, which occurs when $\omega\tau_i = 2\pi m$, $m \in \mathbb{Z}$. Let $\mathbf{I}_n = 1$ if $m\tau_n/\tau_i \in \mathbb{Z}$, and 0 otherwise. Now

$$\begin{aligned} &\lim_{s \to j2\pi m/\tau_i} |1/G_{pw_i}(s)| \\ &= \lim_{s \to j2\pi m/\tau_i} \left| c(1 - e^{-s\tau_i}) + x_i + \sum_{\substack{n=1 \\ n \neq i}}^{N} x_n e^{-s\tau_n} \frac{1 - e^{-s\tau_i}}{1 - e^{-s\tau_n}} \right| \\ &= x_i + \sum_{\substack{n=1 \\ n \neq i}}^{N} x_n \frac{\tau_i}{\tau_n} \mathbf{I}_n > 0, \end{aligned} \tag{11}$$

using L'Hôpital's rule in the second step when $\mathbf{I}_n = 1$. Thus $\lim_{l \to \infty} |1/G_{pw_i}(s_l)| > 0$ for all sequences s_l in $\bar{\mathbb{C}}^+$ for which the limit exists, whence (7) holds, and thus (c).

Furthermore, since $1/G_{pw_i}(s) \neq 0$, $G_{pw_i}(s)$ is also nonsingular in $\bar{\mathbb{C}}^+$, and therefore analytic as its components are analytic. This establishes (a). Condition (b) holds since $G_{pw_i}(s)$ is analytic in $\bar{\mathbb{C}}^+$. ■

C. Uniqueness of rates

The results presented until now hold for any $x(t)$ satisfying (2). It is possible for the windows not to define unique rates, due to sub-RTT burstiness. Consider a network in which two flows with equal RTTs τ share a bottleneck link of capacity C, and each has window $C\tau/2$. If the flows alternate between sending at rate C for time $\tau/2$ and sending at rate 0 for $\tau/2$, and if the "on" periods of flow 1 coincide exactly with the "off" periods of flow 2, then the total rate flowing into the bottleneck link is constant, and (2) is satisfied. It is also satisfied if both sources send constantly at rate $C/2$.

For a single bottleneck, the rates will be unique unless one flow has a RTT which is a rational multiple of another flow's RTT.

To see this, note that sustained oscillations in the rate for a constant window correspond to marginally stable (pure imaginary) poles of (2). Taking the Laplace transform of (2) and eliminating p, gives

$$\mathrm{diag}(se^{s\tau_k})w(s) = \left(\frac{1}{c}\mathrm{diag}(x_k)E + \mathrm{diag}(e^{s\tau_k} - 1)\right)x(s), \tag{12}$$

where $E_{k,l} = 1$ for all $k, l = 1, \ldots, N$. Since $\mathrm{diag}(se^{s\tau_k})$ is never singular for $s \neq 0$, the poles of (12) are the non-zero values of s for which the coefficient of $x(s)$ is singular. The only imaginary values for which this occurs are when $s\tau_i = j2\pi b$ and $s\tau_k = j2\pi a$ for some $i, k = 1, \ldots, N$ and integers a and b.

For two flows, oscillation occurs at $\min(a, b)$ times per RTT for the smaller RTT flow. Even if τ_1/τ_2 is not exactly rational, or is a ratio of large integers, slowly decaying oscillations may exist corresponding to an approximation $\tau_1/\tau_2 \approx b/a$ for smaller a and b. For $\epsilon = b\tau_2 - a\tau_1$, there is a pole with $\sigma + j\omega \approx -(2\pi\epsilon)^2/(\tau_1 + \tau_2)^3 + j2\pi(a + b)(\tau_1 + \tau_2)$. When b/a is a poor approximation to τ_1/τ_2, ϵ will be large making σ very negative, and the oscillations will diminish rapidly.

D. Relation to existing models

The model may be simplified by approximating the integral equation (1b) defining the instantaneous rate, and the N integral constraints in (1). Let $H_t(z) = \int_t^z x(T)dT - w(z)$. By (1b), $H_t(t + \tau(t)) = 0$. Standard approximations to $H_t(z)$ yield several popular models. Rigorous analysis of the accuracy in the queuing delay p would demand considering the coupling between the constraints (1b) and the integration (1a). However, intuitively, better approximations of the constraint should lead to greater model accuracy.

1) Ratio models: Most common models take $x_n(t) \approx w(t-\Delta_a)/\tau(t-\Delta_b)$, for some choice of Δ_a and Δ_b [12–16]. Applying the right-side rectangle rule to $H_t(t + \tau(t))$ gives $x_n(t + \tau_n(t)) \approx w_n(t + \tau_n(t))/\tau_n(t) + \mathcal{O}(\tau_n)$ whence

$$x_n(t) \approx w_n(t)/\tau_n(t - \tau_n(\tilde{t})) \tag{13}$$

where $\tilde{t}$ satisfies $\tilde{t} + \tau_n(\tilde{t}) = t$. This is similar to the integrator model shown in [17] to be overly pessimistic for large RTTs. More accurate numerical quadrature rules can also be applied. However such approximate models are of the same (or higher) complexity as the more accurate model and they furthermore, surprisingly, seem to tend to be unstable.

By further assuming in (13) that the deviation from the equilibrium rates are negligible, $x_n(t) = x_n + \delta x_n(t) \approx x_n$, we get a static update of the queue in terms of window updates as suggested in [17].

2) "Joint" models: Taylor expansion of H_t around t yields

$$\begin{aligned} 0 &= H_t(t + \tau(t)) = H_t(t) + H_t'(t)\tau(t) + \mathcal{O}(\tau^2) \\ &= -w(t) + (x(t) - \dot{w}(t))\tau(t) + \mathcal{O}(\tau^2). \end{aligned} \tag{14}$$

Dividing by $\tau(t)$ gives the rate used by the "joint link model" [18] as an $\mathcal{O}(\tau)$ approximation

$$x_n(t) \approx w_n(t)/\tau_n(t) + \dot{w}_n(t). \tag{15}$$

Ignoring the $\dot{w}(t)$ in (14) gives $x_n(t) \approx w_n(t)/\tau_n(t)$. If $\dot{w}_n = \mathcal{O}(\tau_n)$ then this is again an $\mathcal{O}(\tau_n)$ approximation, albeit less accurate than (15); otherwise it is $\mathcal{O}(\dot{w}_n)$.

Taking higher order terms in the Taylor expansion of $H(t+\tau(t))$ gives more accurate models. However, this leads to high order ODE models.

3) Models by Padé approximations: An alternative is to study the linearized model in the Laplace domain (4), and use, for example, different orders of Padé approximations to $e^{-s\tau_n}$. In this context a (0,0) Padé approximation (i.e. $e^{-s\tau_n} \approx 1$) gives the "static link model" introduced in [17], while the "joint link model" [18] corresponds to a (0,1) approximation. By a (1,0) approximation, a time-scaled ratio model is achieved, c.f. [12–16]. A suitable order of approximation can be chosen, and a nonlinear ODE may then be "reverse engineered" to approximate the DAE model. This approach is used with good accuracy in the linear validation example in Section V-B.

All of the above models are based on small τ approximations. However, $\tau(t)$ need not be small; in particular $\tau(t)$ does not approach zero in the fluid limit of many packets. Thus, (1b) should be used whenever it results in a tractable problem formulation, such as the analysis of loss synchronization and stability of delay based protocols in [19].

V. Model Validation

In this section the model derived in Section III is validated. The model is simulated in Simulink, and the simulation output is compared with packet level data achieved using NS-2. Note that in the experiments we only execute positive changes of the window $w(t)$ (remember it represents the packets "in flight" here). This is to decouple the dynamics of the studied mechanism from the dynamics of the inherited traffic shaping. Recall that a negative change is dependent on the rate of received ACKs.

A. Nonlinear model

We refer to the motivating example in Section II due to limited space. The solid pink line in Fig. 2 shows the queue size when the system is simulated in NS-2, the dashed black line the DAE model (1). The model fits almost perfectly.

Further validation is provided in [21].

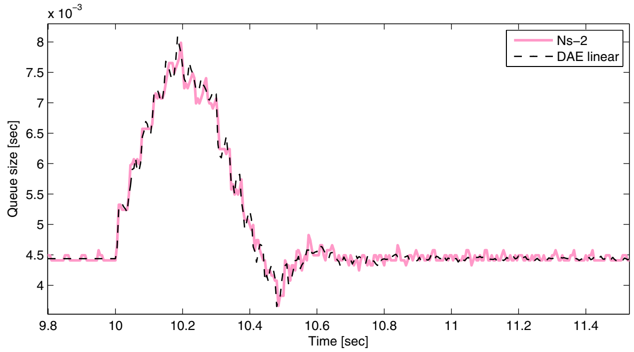

Fig. 3. Validation example. Solid line: NS-2 simulation. Dashed line: Continuous time DAE model (1).

B. Linearized model

Two window based flows are sending over a bottleneck link with capacity $c = 100$ Mbit/s. There is no non-window based cross traffic, so $x_c = 0$. Initially, $w_1 = 60$ packets and $w_2 = 2000$ packets, with packet size $\rho = 1040$ byte. Furthermore, $d_1 = 10$ ms and $d_2 = 190$ ms, with no forward delay. The system is started in equilibrium, and w_1 is increased by 10 at $t = 10$ s, and 300 ms later it is decreased back to 60. The solid line in Fig. 3 shows the queue size when the system is simulated in NS-2, the dashed line the linear approximation (4). The model is good, so the linear approximation seems valid. (In the simulation of (4), a Padé approximation of order (17,17) of the exponential functions has been used.)

VI. Conclusion

We have provided a rigorous analysis of the dynamics of the ACK-clocking mechanism in window-based congestion control for a single bottleneck link. The main result is a fluid flow model of the system. The model is shown in packet level experiments using NS-2 to be very accurate and qualitatively different from its predecessors. The model can be generalized to a multi-link network [21].

We define the instantaneous rate of each window based source (as seen by the link) by a fundamental integral equation. This is in contrast to the customary approach for approximating the window based sources' sending rates, and is the key in the modeling. The system has unique equilibrium rates. Furthermore we show that a linear approximation of the model around the equilibrium is asymptotically stable from the window sizes to the queue size. Many existing models in the literature are shown to be certain approximations to this new accurate model. This procedure also provides insight into how to derive other simplified models.

A natural application of the model is stability analysis of window based congestion control algorithms. Since the model captures sub-RTT burstiness it can be used to analyze, e.g., loss synchronization. Analyzing how such microscopic effects influence macroscopic properties is future work, although exciting initial steps are given in a companion paper [19]. It also remains to explore the implications of the model for general networks.

Acknowledgments

The authors thank Niels Möller for useful discussions. This work was supported by the KTH ACCESS Linnaeus Centre, the Swedish Research Council, the Swedish Foundation for Strategic Research, and the European Commission through the Network of Excellence HYCON and the Integrated Project RUNES; by NSF under grants 0435520 and EIA-0303620, the WAN-in-Lab project.

References

[1] M. Fomenkov, K. Keys, D. Moore, and K. Claffy. Longitudinal study of Internet traffic in 1998-2003. In *WISICT '04: Proc. Winter Int. Symp. Info. Commun. Technol.*, 2004.

[2] V. Jacobson. Congestion avoidance and control. *ACM Comput. Commun. Rev.*, vol. 18, no. 4, pp. 314–329, 1988.

[3] S. Floyd and T. Henderson. The NewReno modification to TCP's fast recovery algorithm. RFC 2582, April 1999.

[4] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow. TCP selective acknowledgements options RFC 2018, October 1996.

[5] L. Xu, K. Harfoush and I. Rhee. Binary Increase Congestion Control for Fast Long-Distance Networks. In *Proc. IEEE INFOCOM*, 2004.

[6] D. J. Leith, R. Shorten. H-TCP Protocol for High-Speed Long Distance Networks. In *Proc. PFLDnet*, 2004.

[7] L. S. Brakmo, S. W. O'Malley, and L. L. Peterson. TCP Vegas: new techniques for congestion detection and avoidance. In *Proc. ACM SIGCOMM*. 1994, pp. 24–35.

[8] D. Wei, C. Jin, S. H. Low, and S. Hegde. FAST TCP: motivation, architecture, algorithms, performance. *IEEE/ACM Trans. Networking*, December 2006.

[9] R. King, R. Baraniuk and R. Riedi. TCP-Africa: An Adaptive and Fair Rapid Increase Rule for Scalable TCP. In *Proc. IEEE INFOCOM*, 2005.

[10] 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.

[11] F. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *J. Op. Res. Soc.*, 49:237–252, 1998.

[12] E. Altman, C. Barakat, and V. Ramos. Analysis of AIMD protocols over paths with variable delay. In *Proc. IEEE INFOCOM*, 2004.

[13] F. Baccelli and D. Hong. AIMD, fairness and fractal scaling of TCP traffic. In *Proc. IEEE INFOCOM*, 2002.

[14] C. Hollot, V. Misra, D. Towsley, and W. B. Gong. A control theoretic analysis of RED. In *Proc. IEEE INFOCOM*, Anchorage, AK, April 2001, pp. 1510–1519.

[15] S. H. Low, F. Paganini, and J. C. Doyle. Internet congestion control. *IEEE Control Systems Magazine*, 22(1):28–43, Feb. 2002.

[16] 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.

[17] J. Wang, D. X. Wei, and S. H. Low. Modeling and stability of FAST TCP. In *IMA Volumes in Mathematics and its Applications*, Volume 143: Wireless Communications. Springer Science, 2006.

[18] K. Jacobsson, H. Hjalmarsson, and N. Möller. ACK-clock dynamics in network congestion control – an inner feedback loop with implications on inelastic flow impact. In *Proc. IEEE Conf. Decision Control*, 2006.

[19] A. Tang, L. L. H. Andrew, K. Jacobsson, K. Johansson, S. H. Low and H. Hjalmarsson. Window Flow Control: Macroscopic Properties from Microscopic Factors In *Proc. IEEE INFOCOM*, 2008.

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

[21] K. Jacobsson, L. L. H. Andrew, A. Tang, K. Johansson, S. H. Low and H. Hjalmarsson. ACK-Clocking Dynamics: Modelling the interaction between windows and the network. [online] Available ⟨http://netlab.caltech.edu/publications/AckClockTR07.pdf⟩.

[22] G. E. Dullerud and F. Paganini *A Course in Robust Control Theory*. Springer, 2000.