Design of Active Queue Management System for Scalable TCP in High Speed Networks

Harsha Sirisena, Aun Haider, and Victor Sreeram [§]

March 17, 2008

Abstract

Scalable TCP, based on a multiplicative-increase multiplicative-decrease congestion avoidance algorithm, has been proposed recently to overcome the inability of Standard TCP to utilize the full bandwidth in high-speed networks. This paper employs a novel approach to derive a transfer-function model of Scalable TCP that is then employed in a control-theoretic design of RED-based Active Queue Management for such a network. Robust stability of the proposed scheme is established under prescribed conditions and the design is validated by discrete-event simulations using the ns2 tool.

1 INTRODUCTION

A new generation of high-speed data networks is being deployed around the world for enabling collaboration between far-flung research groups via the sharing of computing and storage resources. Towards this end, paradigms such as Grid Computing, which pools distributed resources and services across a network, are rapidly evolving. Ultrafast data communication over the network is a key requirement for the success of Grid Computing, however the widely deployed standard transport protocol, TCP (Transmission Control Protocol), is inherently incapable of utilizing the full capacity of high-speed networks. In a steady-state environment, with a packet loss rate p, Standard TCP's average congestion window is roughly $1.2/\sqrt{p}$ segments. This places a serious constraint on the congestion windows that can be achieved by TCP in realistic environments. Firstly, a Standard TCP connection achieves full bandwidth utilization of a high-speed connection only at unrealistically low packet drop rates.

*This work was supported by a Gledden Visiting Senior Fellowship for the first author.

†H. Sirisena is with Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand harsha.sirisena@canterbury.ac.nz

‡A. Haider is with Institute of Information Science and Technology, Massey University, Palmerston North, New Zealand A.Haider@massey.ac.nz

§V. Sreeram (Corresponding Author) is with the School of Electrical, Electronic, and Computer Engineering, University of Western Australia, Crawley, Western Australia 6009 sreeram@ee.uwa.edu.au

Secondly, the window size would take an unreasonably long time to recover after a congestion event.

Two recent proposals, HighSpeed TCP [1] and Scalable TCP [2] address these fundamental limitations of Standard TCP by employing modified TCP response functions (the function mapping the steady-state packet drop rate to TCP's average sending rate in packets per round-trip time). Essentially, the Additive Increase Multiplicative Decrease (AIMD) congestion avoidance algorithm of standard TCP is replaced by Multiplicative Increase Multiplicative Decrease (MIMD) algorithms, exactly in the case of Scalable TCP and effectively in HighSpeed TCP. By letting the modified response functions take effect only with larger congestion windows, the TCP behavior in environments with heavy congestion is unmodified, obviating the danger of congestion collapse.

The performance of TCP with standard routers employing drop-tail queue management suffers from several drawbacks including global synchronization of multiple TCP flows. This occurs because senders go to slow start and increase their windows simultaneously, causing undesirable queue size fluctuations leading to high delay and jitter, less network utilization, and multiple consecutive packet drops that are bad for TCP fast recovery. These drawbacks are alleviated by active queue management (AQM) implemented by random early detection (RED) routers [3] that randomly drop packets with increasing probability as the averaged queue length increases. The randomization of congestion signals breaks the global synchronization, thereby improving network utilization. Moreover, by proactively dropping packets before the buffer is full, smaller average queue lengths and so less delay and delay jitter may be achieved. The performance of AQM can be further improved by the use of explicit congestion notification, which consists of marking packets during the times of low/moderate congestion before dropping them during high congestion.

It is worth to mention that efficacy of RED has been challenged in

The performance and design of RED algorithm with Standard TCP has been well studied. Early simulation studies demonstrated the benefits and some drawbacks of these algorithms when their parameter settings are chosen arbitrarily. More recently, control-theoretic analytical techniques [5] and [6] to better understand the performance, including stability, of TCP with RED as well as other forms of AQM such as PI /REM. The starting point has been the linearized fluid-flow model of Standard TCP Reno [7]. However, the design of AQM algorithms for the newer HighSpeed TCP and Scalable TCP algorithms has not been studied to date.

In this paper, we begin by deriving an analytical model for Scalable TCP by extending the deterministic analytical approach of Yeom and Reddy [8]. It turns out that, unlike in the Standard TCP model, the TCP pole is invariant with the window size, which simplifies the AQM controller design. We are then able to employ control theory to design a RED AQM for scalable TCP to achieve an acceptable stability margin. We verify the RED AQM design using Matlab to simulate a fluid approximation, and then perform proper discrete event simulations using ns2 to validate the control theoretic design.

This paper is organised as follows: The new dynamic model of scalable TCP is derived in section 2. The control theoretic design of AQM RED for Scalable TCP is described in section 3. A design example and ns2 simulation results are presented in section 4. Case studies are presented in section 5. Finally, conclusions are drawn in

section 6.

2 Dynamic Model of Scalable TCP

The generalized Scalable TCP congestion avoidance algorithm responds to each acknowledgment received with the update $cwnd < cwnd + a$ where a is a small positive constant. On the first detection of congestion in a given round trip time, the congestion window is decreased according to $cwnd = cwnd - b \cdot cwnd$ where $b,\ 0 < b < 1$, is a constant. It was suggested [2] that the parameter values be chosen as $a = 0.01$ and $b = 0.125$, and furthermore Scalable TCP should revert to legacy TCP when $cwnd < lwnd$, the typical maximum window size encountered with legacy TCP (about 32 packets) for the sake of fairness.

The steady-state window size of Scalable TCP can be approximated by $cwnd \approx (a/b)(1/p)$ for small packet drop rates p [2]. In contrast, the corresponding expression for legacy TCP is $cwnd \approx 1.2/\sqrt{p}$.

Trivial bounds on the rate of convergence of Scalable TCP were given in [2] and later [9] derived a stability result for Scalable TCP. However, an explicit transfer function model for Scalable TCP, similar to that derived for TCP Reno [5], does not appear to have been published and this issue is addressed next. Rather than the statistical approach of [6] we adopt a deterministic approach inspired by [8] but extended to the time-varying transient case.

Transfer function of Scalable TCP

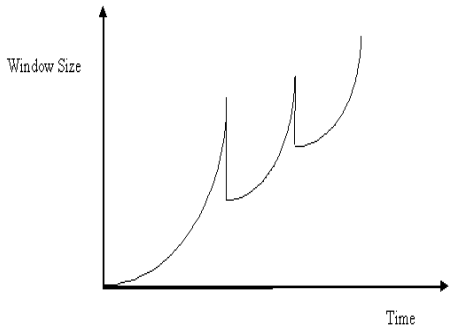

Figure 1: Window evolution for Scalable TCP

Figure 1, not drawn to scale, shows the evolution of the aggregate TCP window, W, which is the sum of the individual windows of the, say, N TCP connections. It follows a cyclic pattern where during cycle k, the aggregate window size increases from $\check{W}_k$ to $\hat{W}_k$. If the cycle comprises n_k round trips, then

$$\hat{W}_k \;=\; (1+a)^{n_k-1}\check{W}_k \tag{1}$$

The cycle ends when a congestion event occurs (triple dupacks) and the TCP connection in question reduces its window by a factor b. Because the mean window size

of a single TCP connection is W/N, the next cycle begins with an aggregate window of size $\check{W}_{k+1}$, where

$$\check{W}_{k+1} = (1 - \frac{b}{N})\hat{W}_k. \tag{2}$$

Now if the packet loss probability is p, then the expected number of packets sent during cycle k

$$E\left\{\sum_{i=0}^{n_k-1}(1+a)^i\check{W}_k\right\} = \frac{1}{p} \tag{3}$$

We make the "deterministic" approximation of omitting the expectation operator in (3). This is actually a very good approximation because the standard deviation, $1/\sqrt{p}$, of the number of packets is much smaller than its mean $1/p$ for small p, assuming that loss events form a Bernoulli process. Then by summing the geometric series we get

$$\left\{\frac{(1+a)^{n_k}-1}{(1+a)-1}\right\}\check{W}_k = \frac{1}{p}, \tag{4}$$

which gives

$$\check{W}_k = \frac{a}{p\cdot\{(1+a)^{n_k}-1\}}. \tag{5}$$

For determining n_k, we rewrite (5) as

$$(1+a)^{n_k} = 1 + \frac{a}{p\cdot\check{W}_k}, \tag{6}$$

which yields

$$n_k = \frac{ln\left(1+\frac{a}{p\check{W}_k}\right\}}{ln(1+a)} \tag{7}$$

From (1) and (2) we have that

$$\check{W}_{k+1} = (1-\frac{b}{N})(1+a)^{n_k-1}\check{W}_k \tag{8}$$

Substituting (6) in (8) we get the following linear difference equation governing the evolution of the sequence $\{\check{W}_k\}$:

$$\check{W}_{k+1} = \left(\frac{1-\frac{b}{N}}{1+a}\right)\check{W}_k + \frac{a\cdot(1-\frac{b}{N})}{p\cdot(1+a)} \tag{9}$$

which is a first-order discrete system with its pole at

$$z = \left(\frac{1-\frac{b}{N}}{1+a}\right), \tag{10}$$

Because $a > 0$ and $0 < b < 1$, the pole clearly lies inside the unit circle and so the system has a stable equilibrium. Moreover the equilibrium value of $\check{W}_k$ can be obtained by setting $\check{W}_{k+1} = \check{W}_k = \check{W}$ to get

$$\check{W} = \frac{a \cdot (1 - \frac{b}{N})}{p \cdot (a + \frac{b}{N})} \tag{11}$$

Then from (7) we obtain the equilibrium number n of round trips in a cycle as

$$n = \frac{ln\left(\frac{1+a}{1-\frac{b}{N}}\right)}{ln(1+a)} \tag{12}$$

Since the number of packets sent in a cycle is $1/p$, the mean aggregate window size W during a cycle is given by

$$W = \frac{1}{p \cdot n} = \frac{ln(1+a)}{p \cdot ln\left(\frac{1+a}{1-\frac{b}{N}}\right)} \tag{13}$$

For small a and b, (13) can be approximated by

$$W \approx \frac{a}{p \cdot (a + \frac{b}{N})}, \tag{14}$$

which is very close to the expression given in [2] for the case $N = 1$ that was derived by considering only a single cycle in equilibrium [1]. The new result we develop here is the following difference equation that describes the dynamic evolution of the aggregate window over multiple cycles. It is obtained from (9) using (11) and(13) to relate $\check{W}_k$ to W_k, assuming that the latter equations are approximately valid during transients, too:

$$W_{k+1} = \left(\frac{1 - \frac{b}{N}}{1+a}\right) W_k + \frac{ln(1+a)}{p \cdot ln\left(\frac{1+a}{1-\frac{b}{N}}\right)} \tag{15}$$

3 CONTINUOUS-TIME MODEL

A continuous-time model, the counterpart of the TCP model derived in [5] and [8], may be derived by observing that the sampling period, T_k, of the discrete system (9) is the duration of a cycle, i.e.,

$$T_k = n_k \cdot R, \tag{16}$$

where R is the round trip time, and n_k is the number of round trips in cycle k. At equilibrium, $n_k \to n$ and $T_k \to T$, a constant, that from (12) is given by

$$T = \left\{ \frac{ln\left(\frac{1+a}{1-\frac{b}{N}}\right)}{ln(1+a)} \right\} \cdot R \tag{17}$$

[1] Newton Mercator series: $ln(1+a) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} a^n = a - \frac{a^2}{2} + \frac{a^3}{3} - ...$, for $|a| \leq 1$. Thus, $ln(1+a) \approx a$ and $ln\left(\frac{1+a}{1-\frac{b}{N}}\right) = ln(1+a) - ln(1-\frac{b}{N}) \equiv a + \frac{b}{N}$, in (14), for smaller values of a.

The ODE which, when sampled with period $T = t_2 - t_1$, yields the difference equation (9) is obtained as:

$$\dot{W} = W_{t_2} - W_{t_1}, \tag{18}$$

which after substitution of (14), will yield:

$$\dot{W} = -AW + \frac{A \cdot ln(1+a)}{p \cdot ln\left(\frac{1+a}{1-\frac{b}{N}}\right)} \tag{19}$$

where

$$e^{-AT} = \left(\frac{1-\frac{b}{N}}{1+a}\right) \tag{20}$$

i.e.,

$$A = \frac{1}{T} \cdot ln\left(\frac{1+a}{1-\frac{b}{N}}\right) \tag{21}$$

Substituting for T using (17) gives

$$A = \left\{\frac{ln(1+a)}{R}\right\}. \tag{22}$$

Now equation (19) is linear in W but nonlinear in p. Linearization about the equilibrium point (11) yields the perturbation dynamics

$$\delta\dot{W} = -A\delta W - \frac{A \cdot ln(1+a)}{p^2 \cdot ln\left(\frac{1+a}{1-\frac{b}{N}}\right)}\delta p \tag{23}$$

Hence the transfer function for the window dynamics of a *single* Scalable TCP connection is

$$P_{stcp}(s) = \frac{1}{N}\frac{\delta W(s)}{\delta p(s)} = \frac{1}{N}\frac{-\frac{[ln(1+a)]^2}{Rp^2 ln\left(\frac{1+a}{1-\frac{b}{N}}\right)}}{s + \frac{ln(1+a)}{R}} \tag{24}$$

This can be expressed in terms of the bottleneck link capacity, C, by equating the aggregate window to the bandwidth-delay product, i.e.,

$$W = RC \tag{25}$$

and using (13) to relate p to W. After simplification we get the transfer function

$$P_{stcp}(s) = -\frac{\frac{RC^2 ln\left(\frac{1+a}{1-\frac{b}{N}}\right)}{N^2}}{s + \frac{ln(1+a)}{R}} \tag{26}$$

For $a << 1$ and $b << 1$, this can be approximated by

$$P_{stcp}(s) \;=\; -\frac{\frac{(aN+b)RC^2}{N^2}}{s+\frac{a}{R}} \tag{27}$$

In comparison with the transfer function for TCP Reno [5],

$$P_{Reno}(s) \;=\; -\frac{\frac{RC^2}{2N^2}}{s+\frac{2N}{R^2C}} \tag{28}$$

observe that for Scalable TCP the pole is independent of the number of connections and the bandwidth, whereas it varies with these quantities in the case of TCP Reno.

Remarks:

1. The transfer function (27) can also be used to understand, qualitatively, the transient behavior of HighSpeed TCP which differs from Scalable TCP in that its parameters a and b vary with the window size.

2. Our approach may be applied to other flavors of TCP as well. In the case of TCP Reno, it enables an alternative simpler derivation of the transfer function (28).

4 Design of RED AQM for Scalable TCP

A RED based active queue management (AQM) system for scalable TCP can be modeled as the feedback control system shown in figure 2 [5]. The marking/dropping probability profile of RED is shown in Figure 3, where the Exponentially Weighted Moving Average (EWMA) $\hat{q}$ of the queue length is computed by the filter $\hat{q} \leftarrow (1 - w_q)\hat{q} + w_q \cdot q$. The transfer function of the RED AQM is given by

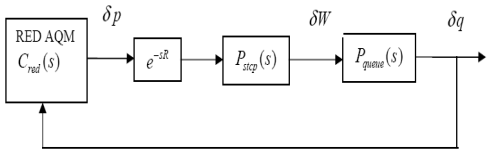

Figure 2: RED AQM Feedback Loop for Scalable TCP

$$C_{red}(s) \;=\; \frac{L_{red}}{s/K+1} \tag{29}$$

where L_{red} and K are the gain and bandwidth respectively of the RED filter. These are related to the RED parameters as follows:

$$L_{red} = \frac{max_p}{max_{th} - min_{th}};\; K = -C \cdot ln\,(1 - w_q) \tag{30}$$

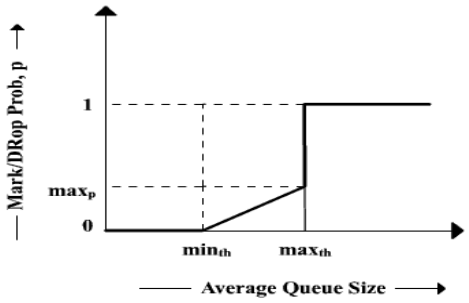

Figure 3: Marking/Dropping Profile of RED

The transfer function of the queue is given by

$$P_{queue}(s) \;=\; \frac{N/R}{s+1/R} \tag{31}$$

Thus this system has 3 poles:

$$p_{red} = -K, p_{stcp} = -a/R, \text{ and } p_{queue} = -1/R, \tag{32}$$

plus a pure delay of R. We first investigate the choice of a suitable RED filter bandwidth K to produce a robustly stable system with a gain margin greater than 4 (or 12 dB) and a phase margin greater than 40°. Simulations using Matlab show that this combination of margins can not be achieved for $K < \sqrt{a}/R$. At this critical value, the averaging time of the RED filter is 10 round trip times for the typical value $a = 0.01$, in accordance with the guideline determined empirically by Floyd et al. [10] in the case of TCP Reno. We initially fix the RED filter pole at

$$K \;=\; \frac{\sqrt{a}}{R} \tag{33}$$

which, not coincidently, is the geometric mean of p_{stcp} and p_{queue}. We then determine the gain L_{red} such that the gain crossover frequency is

$$\omega_g \;=\; K = \frac{\sqrt{a}}{R} \tag{34}$$

i.e.,

$$\left|\frac{L_{red}}{j1+1}\right| \cdot \frac{(aN+b)RC^2/N^2}{|j\sqrt{a}/R + a/R|} \cdot \frac{N/R}{|j\sqrt{a}/R + 1/R|} \;=\; 1 \tag{35}$$

which yields

$$L_{red} \;=\; \frac{N(1+a)\sqrt{2a}}{(aN+b)R^2C^2} \tag{36}$$

The justification for the choice (33) is that it yields a stable system with a phase margin of about 40° which is typical for a good process control system [11]:

$$\begin{aligned} \text{Phase Margin} \;&=\; \pi + \text{phase shift around loop at } \omega = \omega_g \\ &=\; \frac{\pi}{4} - \sqrt{a} \approx 39.5^\circ \;\text{ for }\; a = 0.01 \end{aligned} \tag{37}$$

We now generalize the design by relaxing the choice (33) and prove the following Proposition:

Proposition: Let L_{red} and K satisfy

$$L_{red} = \frac{N^-(1+a)\sqrt{2a}}{(aN+b)(R^+C)^2} \tag{38}$$

$$K \geq \frac{\sqrt{a}}{R^+} \tag{39}$$

Then the linear feedback control system in Figure 2, using $C_{red}(s)$ in (28) is stable for all $N \geq N^-$ and all $R \leq R^+$ with phase and gain margins bounded below as:

$$PM \geq \frac{\pi}{4} - \sqrt{a} \quad \text{and} \quad GM > 4. \tag{40}$$

Proof: Consider the frequency response of the loop transfer function

$$\begin{aligned} L(j\omega) &= -C_{red}(j\omega)P_{stcp}(j\omega)P_{queue}(j\omega)e^{-j\omega R} \\ &= \frac{L_{red}\frac{(aN+b)RC^2}{N^2} \cdot \frac{N}{R} \cdot \frac{R}{a} \cdot Re^{-j\omega R}}{\left(\frac{j\omega}{K}+1\right)\left(\frac{j\omega R}{a}+1\right)(j\omega R+1)} \end{aligned} \tag{41}$$

Its magnitude is

$$|L(j\omega)| = \frac{L_{red}\frac{(aN+b)R^2C^2}{aN}}{\sqrt{\frac{\omega^2}{K^2}+1} \cdot \sqrt{\frac{\omega^2R^2}{a^2}+1} \cdot \sqrt{\omega^2R^2+1}} \tag{42}$$

Now by virtue of (35), for $N = N^-$, $R = R^+$, $K = \sqrt{a}/R^+$, and L_{red} chosen according to (38),

$$|L(j\omega_g)| = 1 \tag{43}$$

Also, inspection of (42) shows that

$$|L(j\omega_g)| < 1 \tag{44}$$

if $N > N^-$ and/or $R < R^+$ with $K = \sqrt{a}/R = \omega_g$. Then because $|L(j\omega)|$ is a monotonically decreasing function of ω, the new gain cross-over frequency $\omega_g' < \omega_g$. Since, the phase angle of $L(j\omega)$ is also a monotonically decreasing function of ω, the new phase margin is

$$\pi + arg\{L(j\omega_g')\} > \pi + arg\{L(j\omega_g)\} \tag{45}$$

$$\text{i.e., } PM > \frac{\pi}{4} - \sqrt{a} \tag{46}$$

if K is kept constant at $\sqrt{a}/R$. To complete the proof, we need to consider the effect of increasing K, which is more complex because this increases the cross-over frequency but decreases the phase lag. We need to show that the latter effect dominates. Actually, from Bode plot theory, the change in gain magnitude at $\omega = \omega_g'$ is negligible

for $K > 10\omega_g'$, so we only need to consider this effect for $\omega_g' < K < 10\omega_g'$ and in particular at the lower limit $K = \omega_g'$. Furthermore we consider only the worst case scenario $N = N^-$, $R = R^+$. So $\omega_g' = \omega_g$.

Considering a perturbation $K \to K + \triangle K$ from $K = \omega_g = \sqrt{a}/R$, we find that

$$\triangle\left(ln|L(j\omega)|\right) \approx \frac{1}{2}\frac{\triangle K}{K} \tag{47}$$

Also, the change due to a frequency increment $\triangle\omega$ is found to be

$$\triangle\left(ln|L(j\omega)|\right) \approx -\frac{3}{2}\frac{\triangle\omega}{\omega} \tag{48}$$

in the vicinity of $\omega = \omega_g$. Then adding (47) and (48) and equating to zero gives the increase in the cross-over frequency as:

$$\omega_g' - \omega_g = \triangle\omega_g \approx \frac{1}{3}\triangle K \tag{49}$$

Remembering that we are considering $K = \omega_g$, the corresponding increment in the phase shift of the loop transfer function is then found to be

$$\triangle\left(\arg(L(j\omega))\right) \approx -\left(\frac{1}{2\sqrt{a}} + 3\right) R\triangle\omega_g + \frac{R}{2\sqrt{a}}\triangle K \tag{50}$$

Using (49)

$$\triangle(\arg(L(j\omega))) \approx \left(\frac{1}{3\sqrt{a}} - 1\right) R\triangle K \tag{51}$$

Clearly

$$\triangle(\arg(L(j\omega))) > 0 \quad \text{for} \quad a < \frac{1}{9} \tag{52}$$

Since, for scalable TCP, $a < \frac{1}{9}$, it is clear that any increase in K further strengthens the inequality (40), thus completing the proof with respect to the guaranteed phase margin.

The result on the guaranteed gain margin can be proved by a similar approach. First L_{red} clearly decreases montonically for increasing N, thus increasing the gain margin when R and K are kept constant. Next the phase angle of $L(j\omega)$ clearly increases monotonically as R decreases and/or K increases, thus increasing the phase crossover frequency. Hence the gain margin increases because $|L(j\omega)|$ is a monotonically decreasing function of ω. We have verified these guaranteed margins as well as the intermediate results (49) and (50) numerically using Matlab.

5 CASE STUDY

Consider the "dumbbell" network topology illustrated in Figure 4 with the link bandwidths and propagation delays indicated. Assuming a TCP packet size of 1000 bytes,

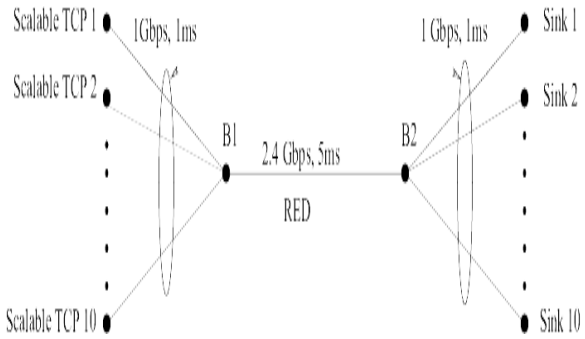

Figure 4: Network Topology

the bottleneck link bandwidth in packets/s is $C = 300,000$. Assuming also that the maximum queuing delay is 6 ms, the upper bound on the round-trip time is $R^+ = 20$ ms. Also, let the lower bound on the number of Scalable TCP connections be $N^- = 10$. Then adopting the typical Scalable TCP parameters $a = 0.01$ and $b = 0.125$, we are now able to design the RED AQM for this scenario.

First, from (38) and (39), we get $L_{red} = 1.763e - 007$ and $K > 5$. Using Matlab we obtain the Bode plots, shown in Figure 5, of the RED AQM loop's frequency response for $K = 5$, the lower limit and $K = 50$. Notice that increasing K increases the stability margins, as well as the control bandwidth as predicted by the analysis in the previous section.

We next present the results of ns2 simulations that validate our RED AQM design for Scalable TCP. First consider the mapping of L_{red} and K to the parameters of RED. In view of the assumed 6 ms upper bound on the mean queuing delay, we set RED router buffer size $= max_{th} = 6C = 2000$ packets, and then choose $min_{th} = 500$ packets. Then using (30), we get $max_p = 0.000265$. Also $w_q = 0.0000167$ for $K = 5$, and $w_q = 0.000167$ for $K = 50$.

Figures 6 and 9, respectively, show the window plots for two of the ten Scalable TCP connections and the instantaneous queue size at the router for $K = 5$, while Figures 7 and 10 show the corresponding plots for $K = 50$. The window plots

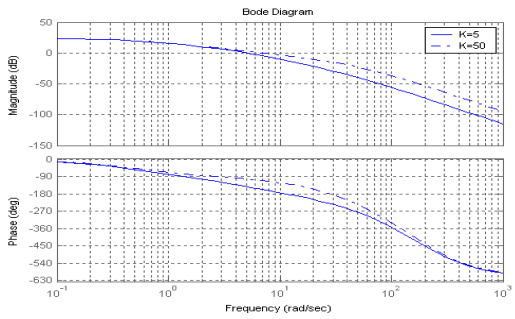

Figure 5: Bode Plots for RED AQM Feedback Loop

for the two values of K look quite similar. However, the EWMA of queue plot for $K = 50$ shows higher frequency oscillations than the plot for $K = 5$, which is to be expected in view of the higher control bandwidth with $K = 50$.

We next investigate the robustness of the design with respect to changes in max_p .

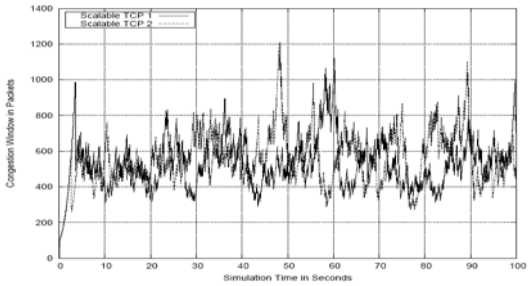

Figure 6: Window Plots for base case, $K = 5$

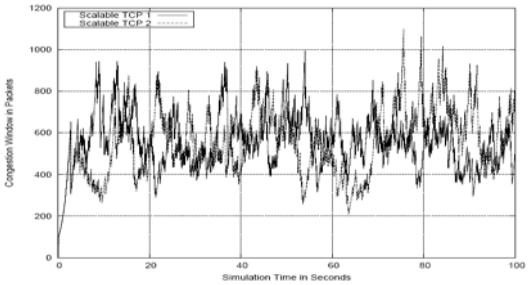

Figure 7: Window Plots, $K = 50$

Figures 8 and 11, respectively, show two window plots and the EWMA of queue plot for $K = 5$, the base value, but with max_p (and hence L_{red}) increased by a factor of 10. The window plots are not significantly different from those for the other two scenarios, confirming the robustness of the base design defined in the Proposition. The main difference is in the EWMA of queue plot. First, the oscillation frequency is the highest of the three cases, which agrees with the control theoretic prediction of increased control bandwidth with loop gain. Second, the mean queue level, approximately 570 packets, is much lower than the mean of approximately 1200 packets for the base value of max_p. This is precisely in accordance with theory: the difference between the mean queue length and min_{th} is reduced by a factor of 10 so that the mean packet dropping probability remains the same as can be understood by considering the RED dropping probability profile, Figure 3. Finally, it may seem surprising that the system is still stable after a tenfold increase in max_p, and hence the loop gain, despite the guaranteed gain margin being 4. However, this can be explained by noting that the decrease in the queuing delay reduces R and hence increases the phase, $\arg L(j\omega)$, of the RED AQM loop, which is stabilizing.

6 CONCLUSION

A nonlinear model of the window dynamics of Scalable TCP in the congestion avoidance phase was developed by a deterministic analytical approach. Linearization of this model at the equilibrium state produced a transfer function model, similar to that derived previously for Standard TCP by a different approach. However, a note-

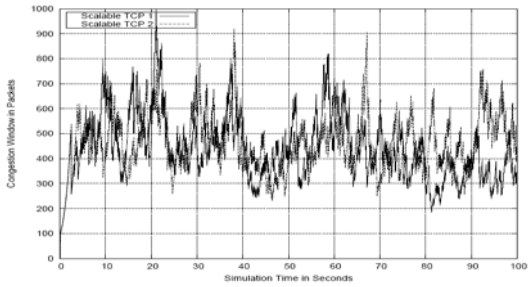

Figure 8: Window Plots, $K = 5$, max_p increased tenfold

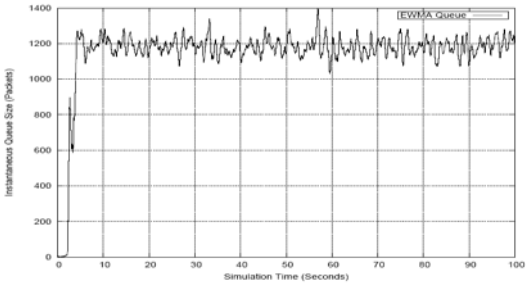

Figure 9: EWMA of Queue for base case, $K = 5$

worthy feature of this model is that, unlike in Standard TCP, the pole is independent of both the available bandwidth and the round-trip time. The new transfer function model was then employed in a control-theoretic design of RED-based Active Queue Management for high-speed networks. Conditions for the robust stability of the proposed scheme were established and the design was validated by simulation studies using ns2. The analytical model of Scalable TCP derived here may also be employed to design other types of AQM such as PI and PID controllers; this is left for future work.

References

[1] S. Floyd, HighSpeed TCP for Large Congestion Windows, *RFC 3649*, Category: Experimental, December 2003.

[2] T. Kelly, Scalable TCP: Improving Performance in HighSpeed Wide Area Networks, *ACM SIGCOMM Computer Communication Review*, Vol. 33, No. 2, pp. 83-91, April 2003, http://www-lce.eng.cam.uk/ ctk21/scalable/#publications.

[3] Floyd S. and Jacobson V, Random Early Detection Gateways for Congestion Avoidance, *IEEE/ACM Trans. on Networking*, vol. 1, no. 4, pp. 397-413, Aug 1993.

[4] Thomas Bonald, Martin May and Jean-Chrysostome Bolot, Analytic Evaluation of RED, *IEEE/ACM Trans. on Networking*, vol. 1, no. 4, pp. 397-413, Aug 1993.

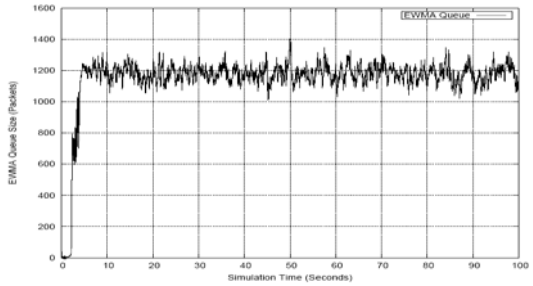

Figure 10: EWMA of Queue, $K = 50$

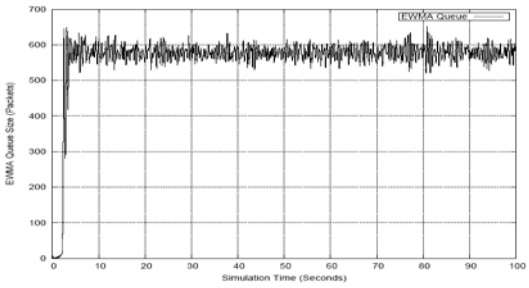

Figure 11: EWMA of Queue, $K = 5$, max_p increased tenfold

[5] C.V. Hollot, V. Misra, D. Towsley, and W.-B. Gong, A Control Theoretic Analysis of RED, *Proceedings of IEEE INFOCOM 2001*, Vol. 3, pp. 1510-1519, April 2001.

[6] C.V. Hollot, V. Misra, D. Towsley, and W.-B. Gong, On Designing Improved Controllers for AQM Routers Supporting TCP Flows, *Proceedings of IEEE INFOCOM 2001*, Vol. 3, pp. 1726-1734, April 2001.

[7] V. Misra, W.-B. Gong, and D. Towsley, Fluid-Based Analysis of a Network of AQM Routers Supporting TCP Flows with an Application to RED, *Proceedings of ACM/SIGCOMM 2000*.

[8] I. Yeom, and A.L. Narasimha Reddy, Modeling TCP Behaviour in a Differentiated Services Network, *IEEE/ACM Transactions on Networking*, Vol. 9, No. 1, pp. 31-46, February 2001.

[9] G. Vinnicombe, On the Stability of Networks Operating TCP-Like Congestion Control, 2002, http://www-control.eng.cam.ac.uk/gv/internet/index.html

[10] S. Floyd, R. Gummadi, and S. Shenker, Adaptive. RED: An Algorithm for Increasing the Robustness of RED, Technical Report, www.icir.org/floyd/papers/adaptiveRed.pdf, 2001.

[11] R.C. Dorf, and R.H. Bishop, *Modern Control Systems*, Tenth Edition, Prentice Hall, NJ, U.S.A, 2005.