Robust congestion control for the Internet

Glenn Vinnicombe
submitted to SIGCOMM 2002

Abstract

We derive decentralized and scalable robust stability conditions for a fluid approximation of a class of Internet-like communications networks operating a modified form of TCP-like congestion control. The network consists of an arbitrary interconnection of sources and links with heterogeneous propagation delays. Unlike previous results of this kind, the model here allows for dynamics at both the sources and the links.

Keywords: Robust control, Communication networks, Complex systems

Notation

$\sigma(Z)$ denotes the spectrum of a square matrix Z and $\rho(Z)$ its spectral radius. Whenever the meaning is unambiguous, $\{f_i\}$ is used as an abbreviation for $\{f_i : i = 1, 2, \ldots\}$. In particular, $\mathrm{Co}\{x_i\}$ denotes the convex hull of the set of points $\{x_1, x_2, \ldots\}$ and $\mathrm{diag}(x_i)$ denotes the matrix with the elements $x_1, x_2, \ldots$ on the leading diagonal and zeros elsewhere.

1 Introduction

We consider a communications network, such as the Internet, consisting of an interconnection of users/sources which generate data and resources/links which carry it. The key constraint of the network that we are interested in is its decentralized nature: control information can only be passed along the same routes as the data that is being transmitted, and with the same propagation delay as that data. We wish to investigate the limitations imposed by this structure. Such a structure might arise, for example, if the resources are allowed to communicate with the users by manipulating certain reserved portions of the data packet header (by marking for example). Kelly *et al* [4] have shown that a certain network utility optimization problem may be solved in a decentralized manner over this structure. In this scheme, each link sets a price per unit flow, based on the aggregate flow through that link, and the sources set their transmission rates based on the aggregate price they see. In the absence of delays, this scheme is globally stable. Moreover, this stability is maintained even if the sources are allowed to adapt their rates arbitrarily fast to achieve arbitrarily high utilization of the network. In the presence of delays this is no longer true. In this paper we derive a sufficient condition for the local stability of such a scheme, for arbitrary network topologies and heterogeneous round trip times. Moreover, we show that this condition captures a trade off between utilization of the network and its speed of response.

Associated with each source is a route, which is the collection of links through which information from that source is flowing. If x_r is the sending rate of source r then the flow through each link is given by

$$y_l(t) = \sum_{r:r \text{ uses } l} x_r(t - \overrightarrow{\tau}_{lr}). \tag{1}$$

where $\overrightarrow{\tau}_{lr}$ denotes the propagation delay from source r to link l. Similarly, if p_l is the price per unit flow at link l, then the aggregate price back at a source is given by

$$q_r(t) = \sum_{l:l \text{ used by } r} p_l(t - \overleftarrow{\tau}_{lr}) \tag{2}$$

where $\overleftarrow{\tau}_{lr}$ represents the return delay from link l to source r. We assume throughout that

$$\overrightarrow{\tau}_{lr} + \overleftarrow{\tau}_{lr} = T_r \quad \forall l, \tag{3}$$

where T_r is the round-trip delay of the rth route. This assumption is consistent with the price information being communicated back to the source via the recipient, along with acknowledgements.

We assume that the link prices are set according to the law

$$p_l = f_l(z_l), \quad f_l > 0, \; f_l' > 0 \tag{4}$$

where z_l is an exponentially smoothed flow rate, satisfying

$$\beta_l \dot{z}_l + z_l = y_l, \quad \beta_l \geq 0 \tag{5}$$

$\beta_l = 0$ corresponds to the price being set as an instantaneous function of the rate, $\beta_l > 0$ corresponds to the more realistic scenario where the rate is estimated either by arrivals over an exponentially weighted window or from queue lengths. We derive such models for link dynamics in Section 3. (Alternatively, it might be the price information itself that is smoothed, i.e. (4) and (5) might be replaced by $\hat{p}_l = f_l(y_l)$ and $\beta_l \dot{p}_l + p_l = \hat{p}_l$. This would give the same linearized transfer function from y_l to p_l and so all of our results will still hold.)

To complete the picture, the sources each have a utility function $U_r(\cdot)$ and attempt to maximize their net utility $U_r(x_r) - q_r x_r$ by setting their rates according to

$$\dot{x}_r(t) = k_r x_r(t - T_r)\left(1 - \frac{q_r(t)}{U_r'(x_r(t))}\right). \tag{6}$$

We shall assume that these utility functions satisfy

$$U_r'(x_r) > 0, \quad U_r''(x_r) < 0 \tag{7}$$

and so are concave. It is shown in [4] that if $T_r = 0$ for each route and $\beta_l = 0$ for each link then the interconnection is globally stable and converges to the unique equilibrium which maximizes the global utility $\sum_r U_r(x_r) - \sum_l \int_0^{y_l} p_l(y)\,dy$. For a closely related source law a simple condition on the gains k_r was derived in [2] which guarantees local stability of this equilibrium whenever all the round-trip trip times are equal. This same condition was conjectured to guarantee stability in the case of heterogeneous round trip delays. This conjecture is true ([5],[7]). In [8] similar stability conditions were given for the general source law (6) and allowing for lags at the links, i.e. $\beta_l > 0$. The condition is that if

$$\frac{\hat{y}_l f_l'(\hat{y}_l)}{f_l(\hat{y}_l)} \cdot k_r T_r < 1$$

whenever link l is on route r, then the network is locally stable about its equilibrium for any collection of lag time constants $\beta_l \geq 0$ and any admissible utility functions. Note, however, without any constraints on the β_l the robustness of this interconnection may be arbitrarily small. In fact there may be infinitesimal perturbations of the link dynamics which destabilize the network[1]. In this paper we shall show that, provided there exists a constant K such that $\beta_l \leq K\tau_l$ for each link, where τ_l is the propagation delay of link l, then this stability is quantifiably robust to perturbations in both gain and phase at both the links and the sources. The tools we develop are also suitable for the analysis of more complex dynamics at both the sources and the links.

We apply these results to a fluid flow approximation of a network with sources operating TCP-like algorithms, where the price is interpreted as the probability that a packet is marked and the dynamics at the links may be interpreted as a combination of queueing effects and the deliberate smoothing introduced by systems such as RED. For the usual TCP algorithm, the conclusion is that the network can be guaranteed to be stable only if the number of packets in flight is sufficiently large. We propose a modification of TCP which avoids this problem. (We should point out, though, that our results are for rate-based control and ignore the potentially stabilizing effects of the window-based control used in current TCP. It is argued elsewhere, though, (e.g. [3]) that this distinction disappears in the limiting regime where capacities increase and queueing delays and queue emptying times become small in relation to propagation delays)

[1] Note that although the focus of this paper is stability, as this is most easy to quantify, in practice we are more concerned about the high variability of rates that will occur as instability is approached, which is likely to drive the system out of the linear regime. Lack of robustness should be understood as being roughly equivalent to such high variance

2 Main result

We can write relationship (1) in terms of Laplace transforms as

$$\bar{y} = R(s)\bar{x}$$

where

$$R_{lr} = \begin{cases} e^{-s\overrightarrow{\tau}_{lr}} & \text{if route } r \text{ uses link } l \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Since

$$\overrightarrow{\tau}_{lr} + \overleftarrow{\tau}_{lr} = T_r \quad \forall l,$$

where T_r is the round-trip time on route r, we can write (2) as

$$\bar{q}(s) = \mathrm{diag}(e^{-sT_i})R^T(-s)\bar{p}(s).$$

These routing relations also hold for small perturbations of course, that is

$$\overline{\delta y} = R(s)\overline{\delta x}$$

and

$$\overline{\delta q}(s) = \mathrm{diag}(e^{-sT_i})R^T(-s)\overline{\delta p}(s)$$

where $y(t) = \hat{y} + \delta y(t)$ etc. Note that the incidence matrix $R(0)$ also determines the static relationships between equilibrium values, i.e.,

$$\hat{y} = R(0)\hat{x}, \quad \hat{q} = R(0)^T\hat{p}. \tag{9}$$

The source law (6) may be linearized around the equilibrium $q_r = U_r'(x_r)$ to give

$$s\overline{\delta x}_r(s) = k_r\frac{\hat{x}_r}{\hat{q}_r}\left(U_r''(x_r)\overline{\delta x}_r(s) - \overline{\delta q}_r(s)\right)$$

or

$$\overline{\delta x}_r(s) = -k_r T_r \frac{\hat{x}_r}{\hat{q}_r}\frac{1}{sT_r + \alpha_r}\overline{\delta q}_r(s) \tag{10}$$

where $\alpha_r = -T_r\frac{\hat{x}_r}{\hat{q}_r}U_r''(x_r) > 0$.

The link law (4–5) may be linearized to give

$$\overline{\delta p}_l(s) = f_l'(\hat{y}_l)\frac{1}{s\beta_l + 1}\overline{\delta y}_l(s) \tag{11}$$

First we state a stability result from [8] The proof of this Proposition is a development of that in [7] for the slightly different source law mentioned in the introduction.

First we recall some language typically used to describe feedback systems. For the linearized system described above the *return ratio* seen at the links (i.e. $-1\times$ the loop transfer function from $\delta y(s)$ back around to $\delta y(s)$) is given by

$$L(s) = R(s)\,\mathrm{diag}\left(k_i T_i \frac{\hat{x}_i}{\hat{q}_i}\frac{e^{-sT_i}}{sT_i+\alpha_i}\right) \cdot R^T(-s)\,\mathrm{diag}\left(f_i'(\hat{y}_i)\frac{1}{s\beta_i+1}\right)$$

The closed loop transfer function from disturbances at the link arrival rates (due to uncontrolled flows for example or stochastic effects) to the actual arrival rates is given by $(I + L(s))^{-1}$. We call the network stable if this transfer function is stable (i.e. analytic and bounded in $\Re(s) > 0$), corresponding to a finite gain from these disturbances to the arrival rates.

Proposition 1. *The interconnection described by (1–7) is locally asymptotically stable around the equilibrium* y_l, $z_l = \hat{y}_l$; $q_r = U'(x_r)$ *if*

$$\frac{\hat{y}_l f_l'(\hat{y}_l)}{f_l(\hat{y}_l)} \cdot k_r T_r < 1\ \forall l, r : r \textit{ uses } l$$

Sketch of proof: The proof in [8] proceeds by showing that the eigenloci of the feedback system's return ratio are contained within

$$\frac{\mathrm{Co}\{\frac{e^{jx}}{jx+\alpha} : x, \alpha > 0\}}{\mathrm{Co}\{jy+1 : y \geq 0\}}$$

which is the region below the curve in Figure 1, and so cannot encircle the point -1.

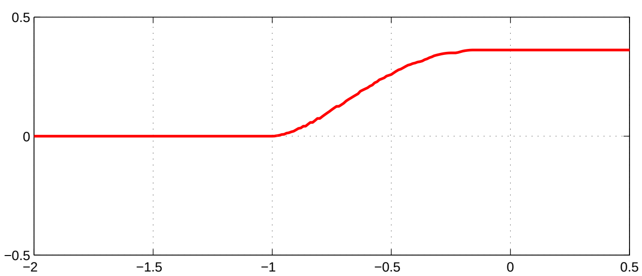

Figure 1: $K = \infty$ (Proposition 1)

Remark 1. *Note that the condition in this Theorem may be satisfied, for example, by ensuring that* $k_r T_r < 1/B$ *and* $f_l(y_l) = \left(\frac{y_l}{C_l}\right)^B$ *(giving* $\frac{y_l f_l'(y_l)}{f_l(y_l)} = B$*) for some global constant* B*. This* $f_l(\cdot)$ *is precisely the probability that an M/M/1 queue with an arrival rate* y_l *and capacity* C_l *is of length* B *or greater, and so is a fairly natural price in this context. This result shows that such a pricing function remains desirable in a more general context.*

The result also captures a tradeoff between utilization of the network and speed of convergence. In [4] the prices $f_l(\cdot)$ *are regarded as barrier functions in the global maximization of* $\sum_r U_r(x_r)$ *subject to the constraint that the flow at each link is no greater than the capacity, a larger* B *would thus correspond to the price remaining small until closer to capacity and then increasing more rapidly around capacity. However, a larger* B *also requires that the sources react more slowly to fluctuations in the network if stability is to be maintained.*

Notice also that it is not necessary that the constant B *be universally agreed upon. In principle, each link could choose its own* B*, perhaps in an attempt to optimize this trade off locally, and communicate it*[2] *to each source using that link.*

The main problem with the previous proposition is that is does not directly guarantee any level of robust stability. This is not a failing of the methods, as without placing bounds on the time constants β it is possible to construct examples which satisfy the theorem statement but which will be destabilized by arbitrarily small amounts of extra phase lag at the links. The following is our main theorem. As before, the source dynamics are written in terms of the roundtrip time, now though the link dynamics are written in terms of the link propagation delay. The key assumption is that the roundtrip time on any route is at least as great as the propagation delay associated with any link on that route, expressed below as $\tau_l \leq T_r$ whenever r uses l, i.e. $R_{lr} \neq 0$. For X a nonempty subset of $\mathbb{C}$ we write

$$S(X) := \left\{\left(\mathrm{Co}\sqrt{X}\right)^2\right\}$$

where

$$\sqrt{X} := \{y : y^2 \in X\}.$$

Note that, for any X, $S(X)$ always contains the origin (since if $y \in \sqrt{X}$ then so is $-y$). In fact $S(X)$ is typically just a little larger than $\mathrm{Co}(0 \cup X)$. See the appendix for more discussion of the operator S and the key Lemma which the following Theorem depends upon.

Theorem 1. *Consider a feedback system with return ratio*

$$L(s) = R(s)\,\mathrm{diag}\big(\tilde{k}_i h_i(sT_i)\big) R^*(s)\,\mathrm{diag}\big(\bar{k}_i g_i(s\tau_i)\big)$$

where R *is defined by* (8), h_r, g_l *are stable for all* r, l *and* $T_r > 0$, $\tau_l \geq 0$, $\tilde{k}_r > 0$, $\bar{k}_l > 0$ *for all* l *and* r*. Let* $\hat{y}$, $\hat{x}$, $\hat{q}$, $\hat{p}$ *be any real and positive vectors satisfying*

$$\hat{y} = R(0)\hat{x}, \quad \hat{q} = R(0)^T \hat{p}.$$

and further assume that

$$R_{lr} \neq 0 \implies \begin{cases} \tau_l \leq T_r \\ \tilde{k}_r \bar{k}_l \leq \dfrac{\hat{p}_l}{\hat{y}_l}\dfrac{\hat{x}_r}{\hat{q}_r} \end{cases}.$$

Finally, assume that there exist parameterized regions H_x, G_y *such that* $h_r(jx) \in H_x$ *for all* r *and* x *and* $g_l(jy) \in G_y$ *for all* l *and* y*.*

[2] e.g. by overwriting some information in a control packet if its B is larger than the B represented there

Under these conditions, the feedback system is stable (i.e. $(1+L(s))^{-1}$ is analytic and bounded in $\Re(s)>0$) if

$$-1 \notin \mathrm{Co}\left\{\cup_x H_x S(\cup_{y\leq x} G_y)\right\}$$

Example 1. *Before proving Theorem 1 we show the result of its application to the network described above, for which the return ratio is given by*

$$R(s)\,\mathrm{diag}\left(k_i T_i \frac{\hat{x}_i}{\hat{q}_i}\frac{e^{-sT_i}}{sT_i+\alpha_i}\right) R^T(-s)\,\mathrm{diag}\left(f_i'(\hat{y}_i)\frac{1}{s\beta_i+1}\right)$$

This corresponds to the return ratio of the Theorem, with $h_i = \frac{e^{-sT_i}}{sT_i+\alpha_i}$ and $g_i = \frac{1}{s\beta_i+1}$. The condition on the gains (with $\hat{y}$, $\hat{x}$, $\hat{q}$, $\hat{p}$ taking the same meanings as outside the Theorem statement) then becomes

$$k_r T_r f_l'(\hat{y}_l) \leq \frac{\hat{p}_l}{\hat{y}_l}$$

(as in Proposition 1). We assume that $\beta_l \leq K\tau_l$, where τ_l is the propagation delay of the lth link. So, $g_i(s\tau_i) = \dfrac{1}{s\tau_i\frac{\beta_i}{\tau_i}+1}$ i.e. $g_i(jy) = \dfrac{1}{jy\frac{\beta_i}{\tau_i}+1}$ and we can take $G_y = \left\{\dfrac{1}{jz+1} : z \leq Ky\right\}$ which is an arc of the circle centred at $+1/2$, from the point $+1$ to the point $1/(jKy+1)$. In this case $\cup_{y\leq x} G_y = G_x$. Figure 2 illustrates G_y and $S(G_y)$ for $Ky = 0.5$ and $Ky = 5$. Similarly, we take $H_x = \left\{\dfrac{e^{-jx}}{jx+\alpha} : \alpha > 0\right\}$. Figure 3 shows the final region, obtained by overlying the $S(G_y)$ on the H_x, for $K = 2$. In this case (and, in fact, for all K up to around 6) the regions $H_x S(G_x)$ are all contained within the union of the Nyquist contours of the set of single-input systems $\left\{\frac{e^{-sT}}{sT(s\tau+1)} : \tau \leq T\right\}$, which is shown as the thicker solid line. In particular, note that the region does not include -1 and so the feedback system is stable. This in itself is unremarkable, since we already know it from Proposition 1. What is important is that the eigenloci are strictly bounded away from 1. Since the operations involved in generating this picture are all continuous this means that the stability is robust to perturbations of the link and/or source dynamics (which would simply result in perturbations to the regions G_y and H_x respectively). In addition a rather large amount of guaranteed robustness can be gained by reducing the gains. For example, Figure 4 shows the corresponding picture when each of H_x and G_y are increased by a radius of 0.1, that is each H_x is replaced by $H_x + \Delta$ where $\Delta = \{z : |z| \leq 0.1\}$ and similarly for G_y. The intercept with the negative real axis now occurs at about -1.3. The conclusion is then that if

$$k_r T_r \frac{f_l'(\hat{y}_l)\hat{y}_l}{\hat{p}_l} \leq \frac{1}{1.3}$$

whenever r uses l then the network will be stable for all stable link dynamics satisfying $\|g_l(s) - \frac{1}{s\beta_l+1}\|_\infty \leq 0.1$ for some $\beta_l \leq 2\tau_l$ and all stable source+delay dynamics satisfying $\|h_r(s) - \frac{e^{-sT_r}}{sT_r+\alpha_r}\|_\infty$ for some $\alpha > 0$. ($\|\cdot\|_\infty$ denotes the $\mathcal{H}_\infty$ norm, $\sup_{\Re(s)>0}|\cdot|$)

In contrast, without using any bound on the βs, the proof of Proposition1 only guarantees that the eigenloci lie underneath the curve in Figure 1. Thus they may approach -1 arbitrarily closely and, even worse, the situation is not improved if the gains are reduced.

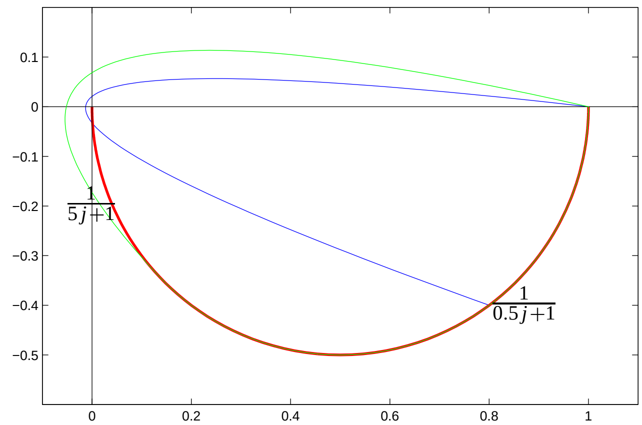

Figure 2: $S(\{1/(jx+1) : x < X\})$

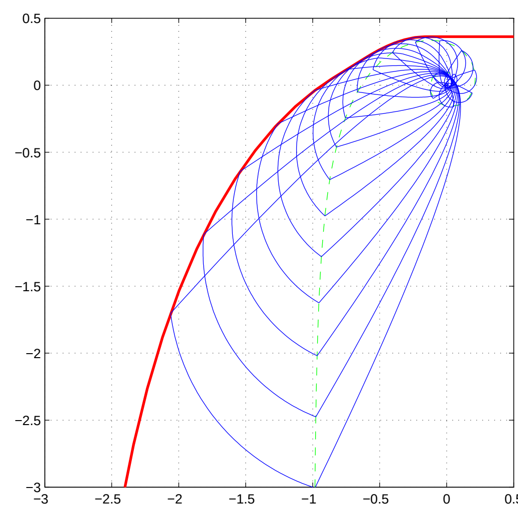

Figure 3: $K = 2$

Proof of Theorem 1. The conditions on the $\tilde{k}_r$s and $\bar{k}_l$s are sufficient to rescale R such that the return ratio is similar to (i.e. shares the same eigenvalues as)

$$\hat{R}(s)\,\mathrm{diag}\big(h_i(sT_i)\big)\hat{R}^*(s)\,\mathrm{diag}\big(g_i(s\tau_i)\big)$$

where now $\rho(|\hat{R}|^T|\hat{R}|) = \rho(\hat{R}(0)^T\hat{R}(0)) \leq 1$ (this can be shown by taking row sums as in [2]). Lemma 1 now applies

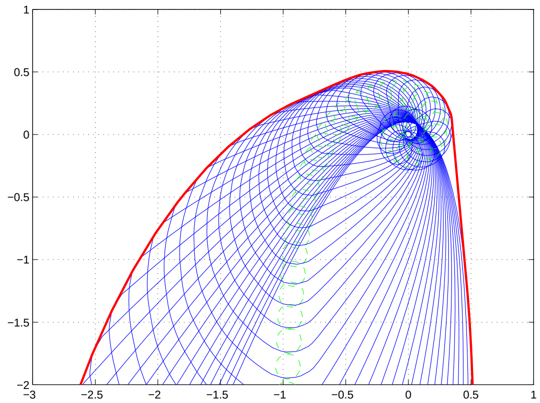

Figure 4: $K = 2$, $\delta = 0.1$

to give

$$\begin{aligned}\sigma\big(L(j\omega)\big) &\subset \mathrm{Co}\,\{h_i(j\omega T_i)S\{g_k(j\omega\tau_k) : R_{ki} \neq 0\}\} \\ &\subset \mathrm{Co}\,\big\{h_i(j\omega T_i)S\{G_y : y < \omega T_i\}\big\} \\ &\subset \mathrm{Co}\,\big\{h_i(jx)S\{G_y : y < x\} : x \in \mathbb{R}_+\big\} \\ &= \mathrm{Co}\,\big\{H_x S\{G_y : y < x\} : x \in \mathbb{R}_+\big\}.\end{aligned}$$

By assumption, the point -1 is not contained in this hull. Since the hull contains the origin this means that no point on the real axis to the left of the point -1 can be included either. It follows that the eigenloci cannot cross the real axis at or to the left of the point -1, and hence that the closed-loop system is stable by the generalized Nyquist stability criterion ([1]). $\square$

3 Link dynamics

In this section we consider how the link dynamics (4–5) may arise when the flow consists of discrete packets, and the link price is set either as a function of packet arrival rate or as a function of queue lengths.

If the arrival rate is to be measured then a natural way to achieve this is to average arrivals over an exponentially weighted window. If the arrival times are τ_i then we can define an average arrival rate at time t as

$$y(t) = \sum_{i:\tau_i < t} \frac{1}{\beta} e^{\frac{-(t-\tau_i)}{\beta}}$$

(for $t \neq \tau_i$). If the arrivals are a time varying Poisson process with rate μ then y is a stochastic process whose mean evolves as

$$\beta \frac{d}{dt} E(y(t)) + E(y(t)) = \mu(t).$$

That is, we can write $y(t) = E(y(t)) + w(t)$ where $E(y(t))$ is determined by μ and w is a zero mean noise signal. In addition, it is straightforward to verify that w is uncorrelated with μ. The choice of β is a tradeoff between noise attenuation at source and lack of robustness (leading to noise amplification). We would typically be interested in the value of y actually *at* the arrival instants, in which case it would be reasonable to define $y(\tau_k)$ as $\sum_{i:\tau_i<\tau_k} \frac{1}{\beta} e^{\frac{-(\tau_k-\tau_i)}{\beta}} + \frac{1}{2\beta}$. These values can be calculated recursively as

$$y(\tau_i) = e^{\frac{-(\tau_i-\tau_{i-1})}{\beta}} \left(y(\tau_{i-1}) + \frac{1}{2\beta} \right) + \frac{1}{2\beta}.$$

If C_l is the capacity of the link, then we can set the price p_l as

$$p_l = (y_l/C_l)^{B_l}.$$

Provided $k_r T_r < B_l$ for all routes using link l then the network is stable. This stability is robust if the time constant β_l is not large in comparison to the propagation delay.

We now consider the scheme where prices/marking probabilities are set in terms of queue lengths. Kelly *et al* [4] assume that queue dynamics may be averaged over round trip times, and consequently that marking probabilities at each link may be approximated by a static function of flow rate. In contrast, Paganini *et al* [6] in common with much of the literature, model each queue as a saturated (at zero) integrator. In this section, we derive a single dynamic model of a queue which reduces to each of these special cases under appropriate limiting conditions, with low load/short queues leading to the former model and high load/long queues leading to the latter. In particular, we show that, for an M/M/1 queue with threshold marking, the linearized transfer function from arrival rate to marking probability has a time constant $N(1 + q_0)/\lambda$ (where q_0 is the expected equilibrium queue length for the current load, N is the marking threshold and λ is the service rate). As noted above, any route using this buffer must have a round trip time of at least the propagation time of the link, and so the effect of queue dynamics become small on links for which the maximum number of packets in flight is greater than $N(1+q_0)$, at least for the purposes of the local stability analyses of the previous sections. Equivalently, the ratio of propagation delay to queueing delay should be greater than N. We recognize that an M/M/1 queue is not particularly representative of a buffer on the Internet, but should at least give results of the right form.

Consider an M/M/1 queue with service rate λ and arrival rate $\mu(t)$, which is assumed to be time varying (i.e. the expected number of arrivals in the interval $[t_1, t_2]$ is $\int_{t_1}^{t_2} \mu(t)\, dt$ and is independent of arrivals in any other interval). For any given function of the queue length, e.g. $F(q)$, we wish to write

$$F\big(q(t)\big) = f(\{\mu(\tau) : \tau \leq t\}) + w(t)$$

where $w(t)$ is a noise source uncorrelated with μ. It is straightforward to show that this is achieved by letting $f(\{\mu(\tau) : \tau \leq t\})$ be simply the expected value of $F(q)$. This choice also results in w being zero mean and having the smallest variance of all possible choices.

If we let

$$x = \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_k \\ \vdots \end{bmatrix}$$

where $p_k(t)$ is the probability that the queue is of length k at time t, then the evolution of x is governed by the differential equation

$$\dot{x} = Ax$$

where

$$A = \begin{bmatrix} -\mu & \lambda & 0 & 0 & \cdots \\ \mu & -\lambda-\mu & \lambda & 0 & \\ 0 & \mu & -\lambda-\mu & \ddots & \\ 0 & 0 & \ddots & \ddots & \\ \vdots & & & & \end{bmatrix}$$

For fixed $\mu_0 < \lambda$, this has an equilibrium

$$x_0 = (1-r)\begin{bmatrix} 1 \\ r \\ r^2 \\ \vdots \end{bmatrix}$$

where $r = \frac{\mu_0}{\lambda}$. Linearizing about this equilibrium, i.e. letting $\mu(t) = \mu_0 + \delta\mu(t)$ and $x(t) = x_0 + \delta x(t)$, we obtain

$$\dot{\delta x} = A_0 \delta x + B\delta\mu \qquad (12)$$

where A_0 is obtained from A by replacing μ with μ_0, and

$$B = \frac{\partial A}{\partial \mu} x_0 = (1-r)\begin{bmatrix} -1 & 0 & 0 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \\ 0 & 1 & -1 & \ddots & \\ 0 & 0 & \ddots & \ddots & \\ \vdots & & & & \end{bmatrix}\begin{bmatrix} 1 \\ r \\ r^2 \\ \vdots \end{bmatrix} = (1-r)^2 \begin{bmatrix} -1/(1-r) \\ 1 \\ r \\ r^2 \\ \vdots \end{bmatrix}$$

3.1 Transfer function to queue length

First we derive the transfer function from arrival rate to expected queue length. We shall not use these results directly, but they are illustrative.

If we let

$$C = \begin{bmatrix} 0 & 1 & 2 & \ldots \end{bmatrix}$$

then $q_0 = Cx_0 = \frac{r}{1-r}$ is the expected equilibrium queue length and $\delta q = C\delta x$ is the deviation of the expected queue length from this equilibrium.

The transfer function from $\delta\mu$ to δq is infinite dimensional, but may be approximated reasonably well by a first order lag. We shall approximate it by the lag which has the identical low and high frequency asymptotic behaviour. To achieve this, we need to calculate $\dot{\delta q}(0)$ and $\delta q(\infty)$ in response to

$$\delta\mu = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}.$$

From (12),

$$\dot{\delta q}(0) = CB = 1$$

as expected (the queue must integrate excess arrivals over short periods). Furthermore, setting $\dot{\delta x} = 0$, we can solve (12) to obtain

$$\delta x(\infty) = \frac{1}{\lambda}\begin{bmatrix} 1 \\ (1-r)-r \\ 2r(1-r)-r^2 \\ \vdots \\ kr^{k-1}(1-r)-r^k \\ \vdots \end{bmatrix}$$

(where we have used the fact sum $\sum \delta x(t) = 1$ for all t) and so

$$\delta q(\infty) = C\delta x(\infty) = \frac{1}{\lambda(1-r)^2}.$$

This results in the approximate transfer function

$$\overline{\delta q}(s) \approx \frac{1}{s+\lambda(1-r)^2}\overline{\delta\mu}(s),$$

which approaches a pure integrator as load $r \to 1$.

3.2 Transfer function to threshold marking probability

Alternatively, if we let

$$C = \begin{bmatrix} 0_{1\times N} & 1 & 1 & \cdots \end{bmatrix}$$

then $\delta p = C\delta x$ is the change in probability that the queue is of length N or greater. It may be similarly calculated that

$$\delta p(\infty) = \frac{1}{\lambda}\sum_{k=N}^{\infty} kr^{k-1}(1-r) - r^k = \frac{N}{\lambda}r^{N-1}$$

and

$$\dot{\delta p}(0) = 1\sum_{k=N}^{\infty}(1-r)^2 r^{k-1} = r^{N-1}(1-r)$$

and so the time constant is

$$T = \frac{N}{\lambda(1-r)} = \frac{N(q_0+1)}{\lambda}$$

and the approximate transfer function

$$\overline{\delta p}(s) \approx \frac{\frac{N}{\lambda}r^{N-1}}{\frac{Ns}{\lambda(1-r)}+1}\overline{\delta\mu}(s) = \frac{Np_0/\mu_0}{\frac{Ns}{\lambda(1-r)}+1}\overline{\delta\mu}(s).$$

Figure 5 shows the true Bode diagrams for this Transfer function normalized to the above time constant and DC gain. Notice the identical low and high frequency behaviour and that they all give approximately 45° of phase lag around $\omega = 1/T$, confirming that a first order lag is a reasonable approximation (although the regions required for Theorem1 could just as easily be determined for these exact frequency responses).

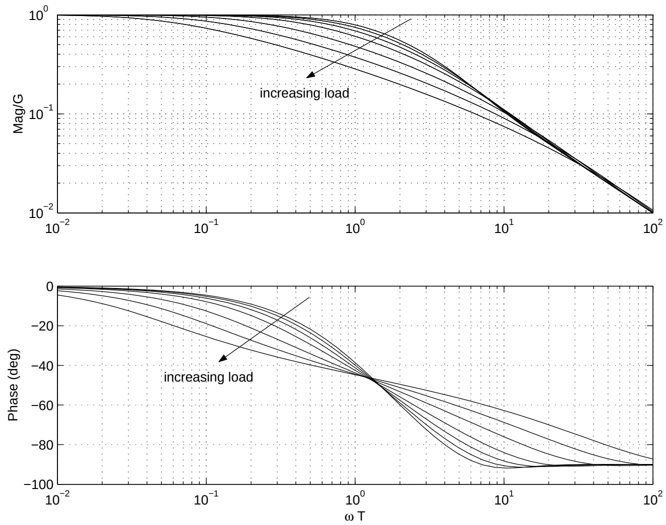

Figure 5: Bode diagrams for arrival rate → marking prob

Transfer functions can be derived, in a similar manner, for marking according to the RED scheme. In this case $C = [c_i]$ where c_i is zero for i less than the minimum threshold, ramps up to $\mathtt{max_p}$ at the maximum threshold, at which point it jumps to 1. In this case an extra lag should be added, since RED is based on an averaged queue length. This situation may also be analysed using Theorem 1, the conclusion being that this extra lag does not compromise stability as long as its time constant is not much larger than the propagation delay of the link.

4 TCP-like algorithms

We consider a class of algorithms where, as in TCP, the source maintains a window $\mathtt{cwnd}$ of sent but not yet acknowledged packets. This window is incremented by $a\mathtt{cwnd}^n$ for each unmarked acknowledgement, and decremented by $b\mathtt{cwnd}^m$ for each marked acknowledgement, where $m > n$. Conventional TCP in its congestion avoidance phase uses an increment of $1/\mathtt{cwnd}$ and a decrement $\mathtt{cwnd}/2$ (and with packets being dropped rather than marked). Properties of this algorithm may be analysed using the methods of the previous section provided it is assumed that queueing delays are small relative to propagation delays, which is equivalent to the assumption that queue dynamics may be averaged over periods significantly shorter than round trip times. See [3] for justification of this limiting regime, and the derivation of models of this kind. We let

$$x_r(t) = \mathtt{cwnd(t)}/T_r$$

be a continuous approximation of the sending rate to obtain

$$\frac{d}{dt}\mathtt{cwnd}(t) = \frac{a\mathtt{cwnd}^n\big(1 - q_r(t)\big) - b\mathtt{cwnd}^m q_r(t)}{T_r/\mathtt{cwnd}(t - T_r)}$$

where the "price" $q_r(t)$ should now be interpreted as the probability that an acknowledgement received at time t carries a mark. In terms of rates, this becomes

$$T_r\dot{x}_r(t) = x_r(t - T_r)\Big(a(x_r(t)T_r)^n\big(1 - q_r(t)\big) - b(x_r(t)T_r)^m q_r(t)\Big) \quad (13)$$

This equation does not quite fall into the framework of the previous section, because of the $(1-q_r)$ term. However, the presence of this term does not affect the linearizations. The equation may be linearized about its equilibrium

$$\hat{q}_r = \frac{a(\hat{x}_r T_r)^n}{a(\hat{x}_r T_r)^n + b(\hat{x}_r T_r)^m}$$

to give

$$\begin{aligned} T_r\frac{d}{dt}\delta x_r(t) &= \hat{x}_r\Big(-\delta q_r(t)\left(a(\hat{x}_r T_r)^n + b(\hat{x}_r T_r)^m\right) + \\ &\Big(aT_r n(\hat{x}_r T_r)^{n-1}\big(1 - \hat{q}_r\big) - bT_r m(\hat{x}_r T_r)^{m-1}\hat{q}_r\Big)\delta x_r(t)\Big) \\ &= \hat{x}_r\left(-\delta q_r(t)\frac{a(\hat{x}_r T_r)^n}{\hat{q}_r} - bT_r(m-n)(\hat{x}_r T_r)^{m-1}\hat{q}_r\delta x_r(t)\right) \\ &= -\left(\frac{a\hat{x}_r(\hat{x}_r T_r)^n}{\hat{q}_r}\right)\delta q_r(t) - \alpha_r\delta x_r(t). \quad (14) \end{aligned}$$

where

$$\alpha_r = b(m-n)(\hat{x}_r T_r)^m\hat{q}_r > 0$$

with the transfer function

$$\overline{\delta x_r} = -\frac{a\hat{x}_r(\hat{x}_r T_r)^n}{\hat{q}_r}\frac{1}{(sT_r + \alpha_r)}\overline{\delta q_r}$$

Comparison with (10) shows that Proposition 1 guarantees local stability whenever $a(\hat{x}_r T_r)^n < 1/B$. As mentioned above, conventional TCP has $n = -1$, $a = 1$ and so stability is only guaranteed if the equilibrium congestion window $\hat{x}_r T_r$ is *greater* than B. For routers operating threshold marking, B might correspond to something like the buffer size (in packets) at which packets are marked. (This would be exact it the buffers behave as M/M/1 queues). An appealing alternative is to take $n = 0$ and choose $a = 1/B$, that is choose a fixed increment of something like the reciprocal of the average buffer level at which packets are marked.

5 Conclusions

We have derived a simple decentralized stability condition for a network consisting of an interconnection of links and sources. The results suggest that TCP as usually implemented is likely to be prone to instabilities when the congestion window is small, and overly sluggish when it is large (at least in the limiting regime as capacities increase and queueing delays and queue emptying times become small in relation to propagation delays). We have suggested a simple scalable modification which avoids this fact, where the TCP window length is increased by a fixed amount of $1/B$ for each acknowledgement and links either perform threshold marking at the level B, or (preferably) calculate a marking probability in terms of measured rate which is equivalent to the marking probability of an M/M/1 queue at threshold B. This scheme is shown to be locally stable about all equilibria, and moreover is robust to dynamic uncertainty at both the links and the sources (including the delay). A further conclusion is that the exponential smoothing introduced by protocols such as RED, or associated with rate measurement, need not compromise stability (at least as long as the time constants are not large in relation to link propagation delays.)

Appendix

If R is a matrix, certain of whose elements are known to be zero, then intuitively the spectrum of $R^* \operatorname{diag}(f_1,\dots) R \operatorname{diag}(g_1,\dots)$ should not depend too much on the products $f_i g_k$ for any i, k for which $R_{ik} = 0$. The following theorem shows that this is indeed the case if a bound on the spectral norm of the absolute value of R is known, in which case then the spectrum of this matrix may be located in terms of the products $f_i g_k$ for which R_{ik} is nonzero.

Let $S\{x_i : i = 1,\dots\} = \left(\operatorname{Co}\{\pm\sqrt{x_i} : i = 1,\dots\}\right)^2$. This set clearly contains the origin as well as each of the points x_i, and is typically a little larger than $\operatorname{Co}(0 \cup \{x_i : i = 1,\dots\})$ (see Figure 6 for an example). We use $|R|$ to denote the elementwise absolute value of a matrix, i.e. $|[R_{ij}]| := [|R_{ij}|]$.

Lemma 1. *Let $R \in \mathbb{C}^{m\times n}$ satisfy $\rho(|R|^T|R|) \le 1$, and $G = \operatorname{diag}(g_1,\dots,g_n)$, $F = \operatorname{diag}(f_1,\dots,f_m)$, $g_i,\ f_i \in \mathbb{C}$ $\forall i$ then*

$$\begin{aligned}\sigma\left(R^*FRG\right) &\subset \operatorname{Co}\left\{ f_i S\{g_k : R_{ik} \neq 0\} : i = 1, m\right\}\\ &= \operatorname{Co}\left\{\left(\operatorname{Co}\{\pm\sqrt{f_i g_k} : R_{ik} \neq 0\}\right)^2 : i = 1, m\right\}\end{aligned}$$

Proof. First note that $\rho(|R|^T|R|) \le 1$ implies that $v^*|R|^T|R|v \le 1 \forall v \in \mathbb{C}^n : v^*v = 1$ or, equivalently,

$$\sum_i \left|v_1|R_{i1}| + v_2|R_{i2}| + \cdots\right|^2 \le 1\,\forall v \in \mathbb{C}^n : v^*v = 1.$$

Since this is true for all such v, it must also be the case that

$$\sum_i \left(|v_1 R_{i1}| + |v_2 R_{i2}| + \cdots\right)^2 \le 1\,\forall v \in \mathbb{C}^n : v^*v = 1. \tag{15}$$

Also,

$$\sigma\left(GR^*FR\right) \subset \left\{v^* G^{1/2} R^* F R G^{1/2} v : v \in \mathbb{C}^n,\ v^*v = 1\right\}$$

(where $G^{1/2} = \operatorname{diag}(\sqrt{g_1},\dots,\sqrt{g_n})$, where either value of each square root may be used)

$$= \left\{\sum_k f_k v^* G^{1/2} R^*_{k\bullet} R_{k\bullet} G^{1/2} v : v \in \mathbb{C}^n,\ v^*v = 1\right\}$$

$$\begin{aligned}= \Big\{\sum_k f_k (v_1^* R_{k1}^* \sqrt{g_1} + v_2^* R_{k2}^* \sqrt{g_2} + \cdots)\\ (v_1 R_{k1}\sqrt{g_1} + v_2 R_{k2}\sqrt{g_2} + \cdots) : v \in \mathbb{C}^n,\ v^*v = 1\Big\}\end{aligned} \tag{16}$$

Next, note that for any $\alpha \in \mathbb{C}^n$,

$$(\alpha_1^*\sqrt{g_1} + \alpha_2^*\sqrt{g_2} + \cdots)(\alpha_1\sqrt{g_1} + \alpha_2\sqrt{g_2} + \cdots)$$

$$= |\alpha_1|^2 g_1 + |\alpha_2|^2 g_2 + \cdots + 2\Re\left(\alpha_1^*\alpha_2\right)\sqrt{g_1 g_2} + \cdots$$

$$\begin{aligned}\in \operatorname{Co}\Big\{&(|\alpha_1|\sqrt{g_1} + |\alpha_2|\sqrt{g_2} + \cdots)^2,\\ &(|\alpha_1|\sqrt{g_1} - |\alpha_2|\sqrt{g_2} + \cdots)^2, \dots\Big\}\end{aligned}$$

(since $\Re\left(\alpha_1^*\alpha_2\right) \in [-|\alpha_1\alpha_2|, |\alpha_1\alpha_2|]$ etc)

$$\begin{aligned}\subset (|\alpha_1| + |\alpha_2| + \cdots)^2 \operatorname{Co}\Big\{&(\operatorname{Co}\{\sqrt{g_1}, \sqrt{g_2}, \dots\})^2,\\ &(\operatorname{Co}\{\sqrt{g_1}, -\sqrt{g_2}, \dots\})^2, \dots\Big\}\end{aligned}$$

(since $(|\alpha_1|\sqrt{g_1} + |\alpha_2|\sqrt{g_2} + \cdots) \in (|\alpha_1| + |\alpha_2| + \cdots)\operatorname{Co}\{\sqrt{g_1}, \sqrt{g_2}, \dots\}$)

$$\subset (|\alpha_1| + |\alpha_2| + \cdots)^2 \operatorname{Co}\left\{(\operatorname{Co}\{\pm\sqrt{g_1}, \pm\sqrt{g_2}, \dots\})^2\right\}$$

(since each term, e.g. $(\operatorname{Co}\{\sqrt{g_1}, -\sqrt{g_2}, \dots)^2, \subset (\operatorname{Co}\{\pm\sqrt{g_1}, \pm\sqrt{g_2}, \dots\})^2$)

In particular, following on from (16),

$$\begin{aligned}&\sigma\left(R^*FRG\right)\\ &\subset \Big\{\sum_k f_k (|v_1 R_{k1}| + |v_2 R_{k2}| + \cdots)^2\\ &\cdot \operatorname{Co}\left\{(\operatorname{Co}\{\pm\sqrt{g_1}, \pm\sqrt{g_2}, \dots\})^2\right\} : v \in \mathbb{C}^n,\ v^*v = 1\Big\}\end{aligned}$$

$$= \left(\sum_k (|v_1 R_{k1}| + |v_2 R_{k2}| + \cdots)^2 \right)$$
$$\cdot \text{ Co } f_k \{ (\text{Co}\{\pm\sqrt{g_1}, \pm\sqrt{g_2}, \ldots\})^2 \}$$
$$\subset \text{ Co } f_k \{ (\text{Co}\{\pm\sqrt{g_1}, \pm\sqrt{g_2}, \ldots\})^2 \}$$

using (15). □

An (obvious) inclusion we have used a several times in this proof is: For $x_i \in \mathbb{R}^+$, $X_i \subset \mathbb{C}$, $i = 1, \ldots$ then $\sum_i x_i X_i \subset \left(\sum_i x_i\right) \text{Co}\{X_i : i = 1, \ldots\}$.

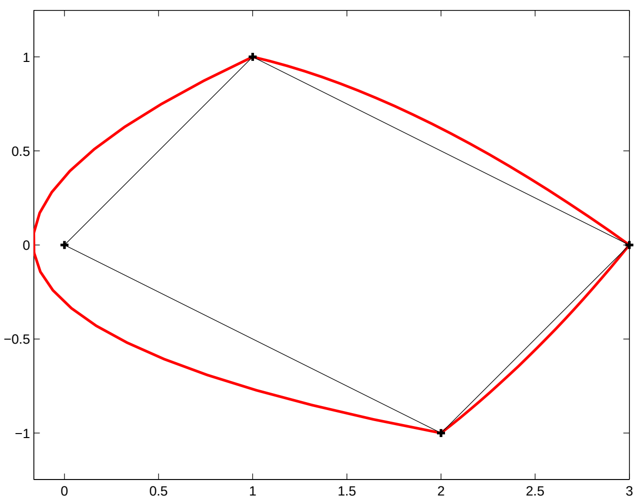

Figure 6: $\text{Co}\{\pm\sqrt{g_i}\}$ vs $\text{Co}(0 \cup \{g_i\})$

The result of Lemma 1 is demonstrated in Figure 7, which shows the eigenvalues of $R^* FRG$ for 1000 random values of R (all with $R_{22} = 0$, $\rho(|R|^T|R|) = 1$) together with the bounding region from Lemma 1. The products $f_i g_k$ are also marked. Note how the region is bounded away from the product $f_2 g_2$ (at $6 + 2j$). Note also that the eigenvalues do not always lie inside the convex hull of $f_1 g_1$, $f_1 g_2$ and $f_2 g_1$; showing that the larger region defined in the Theorem is, in some sense, necessary.

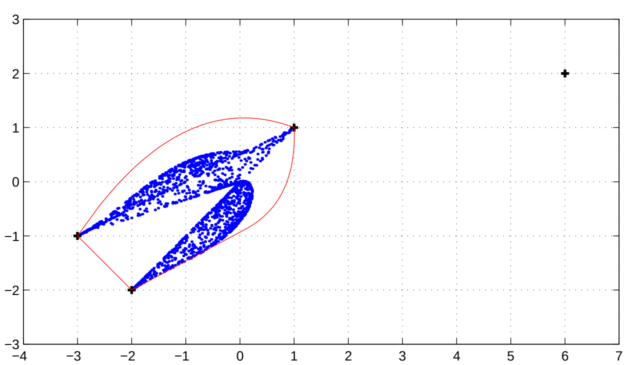

Figure 7: $f = (1 + j, -2 - 2j)$, $g = (1, -2 + j)$, $R_{22} = 0$

References

[1] C. A. Desoer and Y. T. Yang. On the generalized Nyquist stability criterion. *IEEE Transactions on Automatic Control*, 25:187–196, 1980.

[2] R. Johari and D. Tan. End-to-end congestion control for the Internet: delays and stability. Technical report, Statistical Laboratory, University of Cambridge, 2000.

[3] F. P. Kelly. *Mathematics Unlimited – 2001 and Beyond*, chapter Mathematical modelling of the Internet, pages 685–702. Springer-Verlag, Berlin, 2001.

[4] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control in communication networks: shadow prices, proportional fairness, and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[5] L. Massoulie. Stability of distributed congestion control with heterogeneous feedback delays. Technical Report MSR-TR-2000-111, Microsoft Research, Nov 2000.

[6] F. Paganini, J. C. Doyle, and S. H. Low. Scalable laws for stable network congestion control. CDC01, 2001.

[7] G. Vinnicombe. On the stability of end-to-end congestion control for the Internet. Technical Report CUED/F-INFENG/TR.398, Engineering Department, University of Cambridge, 2000. `http://www-control.eng.cam.ac.uk/gv/internet`.

[8] G. Vinnicombe. On the stability of networks operating TCP-like congestion control. In *IFAC*, 2002. to appear.