Scalable Laws for Stable Network Congestion Control

Fernando Paganini* John Doyle† Steven Low‡

March 2, 2001

Abstract

This paper discusses flow control in networks, in which sources control their rates based on feedback signals received from the network links. This feature, present in current protocols such as TCP, has only recently been extensively studied from an analytical perspective. In particular, we are motivated by recent work that shows that TCP and its variants are prone to instabilities when combined with network delays and, more surprisingly, with increases in network capacity. In this paper we develop a congestion control system which is arbitrarily scalable, in the sense that its stability is maintained for arbitrary network topologies and arbitrary amounts of delay. Such a system can be implemented in a decentralized way with information currently available in networks plus a small amount of additional signaling.

1 Introduction

Flow control in a communication network such as the Internet concerns the adjustment of individual source transmission rates so that network resources are fully utilized, and link capacities are not exceeded. Existing protocols such as TCP regulate their transmission rates by a *congestion window* of outstanding packets; feedback signals from the network (packet acknowledgment or loss) are used to dynamically adjust this window to match available capacity. The additive increase multiplicative decrease (AIMD) algorithm of TCP Tahoe [7] and its enhancements (Reno, etc.), based on good engineering intuition supplemented with simulations, has performed remarkably well as the network scales up several orders of magnitude. However future growth combined with the upcoming diversity of traffic and communication substrates will stretch the limits of this simple algorithm, calling for a more extensive analytical investigation of the problem. Indeed, until recently it was impossible even to predict the equilibrium windows of TCP in a network that contains more than a single bottleneck link, let alone its stability, dynamics, and robustness to network parameters.

*UCLA Electrical Engineering, Los Angeles, CA 90095-1594. Corresponding author. Email: paganini@ee.ucla.edu.
†Control and Dynamical Systems, Caltech, Pasadena, CA 91125.
‡Computer Science & Electrical Engineering, Caltech, Pasadena, CA 91125.

Exciting progress has been made very recently towards a theoretical understanding of TCP, both of its equilibrium, using optimization theory, and its dynamics, using control theory. Key to these developments has been to explicitly model the feedback signal generated by links and communicated to sources; in practice these signals correspond to packet dropping, or packet marking when Explicit Congestion Notification (ECN) bits are available. Interpreting these signals as *prices* has allowed for the rate assignment problem to be cast in supply/demand terms [9, 10]; in a related interpretation [16], prices correspond to Lagrange multipliers of a certain optimization problem. When applied to the existing variants of TCP, this framework allows for the analysis of the resulting equilibria [15].

In terms of the *dynamics* of TCP, very recently an analytical model has been developed [19] that correlates well with standard simulators, and allows for stability studies [5, 6]. Note that window-based protocols clock their transmission rate to the packet acknowledgment rate, thus applying a conservation of packets principle that prevents unbounded queues. This does not, however, rule out oscillatory instabilities, which can seriously degrade performance; indeed the work of [5] for the single-bottleneck network case shows instabilities for high delays. Perhaps more surprisingly, the stability requirements are found to be *more* stringent as link capacity increases, to the point where instability is inevitable at the very high capacities one can expect from future networks. These considerations motivate the investigation of new protocols where one could provide guarantees of stability, robustness to delay and scalability to large capacities and arbitrary topologies.

Some work in recent years coming from the control community (e.g. [1, 20, 18]) has investigated the potential of both classical and modern tools in these problems, and in some cases addressed the issue of delay stability. However, results are mainly confined to single bottleneck networks (also the case typically explored by simulation). The most promising results applicable to networks of arbitrary topology are those of optimization-based flow control: in the absence of delay, global stability results can be given [10, 16, 21, 12] for price-based flow control in arbitrary networks. When considering delays, one is forced to slow down the control gains to retain stability (see, e.g. [10, 16, 22]). As noted in [11], this mechanism is already implicit in window-based protocols and therefore it is conceivable that one could obtain stability for arbitrary delays. In this vein, recent work in [8] has derived conditions for delay robustness for the protocols in [10]. While some restrictions are made, such as common round-trip time for all sources, and conditions are still route dependent, this is an important step in the desired direction.

In this paper we pose the objective of finding a protocol which can be implemented in a decentralized way by sources and routers, and which satisfies some basic objectives: efficient use of network resources in equilibrium, and local stability for arbitrary capacities, delays, and routing. These requirements, laid out in Section 3 lead us to adopt certain constraints on the linearized dynamics: integration at the links, and conditions on the gain at sources and links. In Section 4 we then prove a theorem showing that a system with these properties will satisfy our requirements; the proof combines ideas in [8] with quadratic separation arguments typical of multivariable robust control. In Section 5 we discuss the issue of implementation, and in particular characterize a family of nonlinear static functions that sources could use to adjust rates in a way that is consistent with the preceding linearization requirements. Conclusions are given in Section 6.

2 Problem Formulation

We are concerned with a system of L communication links shared by a set of S sources. The routing matrix R, of dimensions $L \times S$, is defined by

$$R_{li} = \begin{cases} 1 & \text{if source } i \text{ uses link } l \\ 0 & \text{otherwise} \end{cases}.$$

- **For each link l we have:**
  - the capacity c_l;
  - the aggregate rate y_l of all flows through the link;
  - the price signal p_l.
- **For each source i we have:**
  - The source rate x_i;
  - The aggregate price q_i of all links used by source i.

We will work with *flow* models, i.e. regard these variables as deterministic, non-negative real-valued. We use vector notation to collect the above variables across all links or sources; thus we define $c, y, p \in \mathbb{R}^L$, and $x, q \in \mathbb{R}^S$.

We now model the network, including the propagation delays, in terms of the relationships (expressed in the Laplace domain, ^{T} denotes transpose)

$$y(s) = R_f(s)x(s), \tag{1}$$
$$q(s) = R_b(s)^T p(s). \tag{2}$$

Here R_f and R_b are the delayed forward and backward routing matrices, defined by

$$[R_f(s)]_{l,i} = \begin{cases} e^{-\tau_{i,l}^f s} & \text{if source } i \text{ uses link } l \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and

$$[R_b(s)]_{l,i} = \begin{cases} e^{-\tau_{i,l}^b s} & \text{if source } i \text{ uses link } l \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Equations (1) and (3) express the fact that link flows are obtained by aggregation, with respective forward delays $\tau_{i,l}^f$, of source flows using the link. Similarly (2) and (4) say that the prices seen by sources are aggregations, with backward delay $\tau_{i,l}^b$, of link prices used by the source.

We define the total round trip delay by source,

$$\tau_i = \tau_{i,l}^b + \tau_{i,l}^f. \tag{5}$$

Note that this quantity is available to sources in real time. We will also assume that the quantities q_i and y_l are available to sources and links, respectively, in real time. More comments on implementation are given in Section 5.

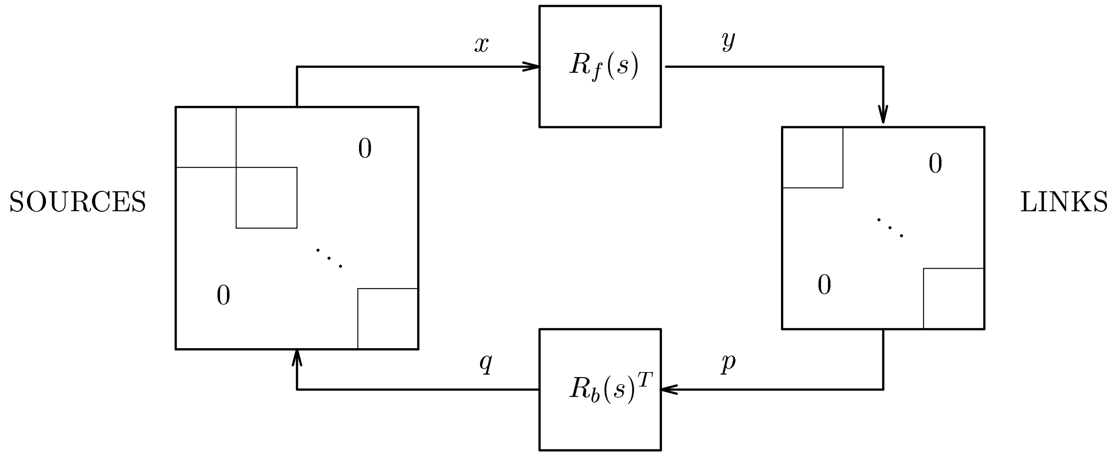


Figure 1: General flow control structure based on pricing signals.

The preceding definitions are illustrated by the diagram of Figure 1. What remains to be specified is:

- How the links fix their prices based on link utilization.
- How the sources fix their rates based on their aggregate price.

These operations are up to the designer, but have a main restriction: both must be *decentralized*, as indicated in the figure by a block diagonal structure. For instance the source rate x_i can only depend on the corresponding aggregate price q_i. The objective of this feedback is to allow for flows to dynamically adapt to changing conditions in traffic demand, link capacity, routing, etc. Key design considerations are thus to ensure dynamic stability and regulation of these systems around equilibria that satisfy some desirable static properties.

This paper is mainly concerned with the dynamic properties around a given equilibrium point x_0, y_0, p_0, q_0. For more discussion on equilibria see Section 5, but for the moment we only require that $y_{0l} \leq c_l$ (link capacities are not exceeded), and that non-bottleneck links (those where $y_{0l} < c_l$) have a price $p_{0l} = 0$.

Now consider a small perturbation around equilibrium; $x = x_0 + \delta x$, $y = y_0 + \delta y$, $p = p_0 + \delta p$, $q = q_0 + \delta q$. Assuming the set of bottlenecks is unchanged by this small perturbation, δp_l is only non-zero for bottleneck links. Therefore for the local analysis to follow, we can write the reduced model

$$\begin{aligned} \delta \bar{y}(s) &= \bar{R}_f(s)\delta x(s) \\ \delta q(s) &= \bar{R}_b(s)^T \delta \bar{p}(s), \end{aligned}$$

where the matrices $\bar{R}_f$, $\bar{R}_b$, and the vectors $\delta\bar{p}$, $\delta\bar{y}$ are obtained by eliminating the rows corresponding to non-bottleneck links. For simplicity in the sequel, we drop the "bar" notation for these reduced matrices, and also the δ from the perturbed quantities: equations (1-2) will be interpreted to refer to perturbed variables around equilibrium, and only involving the bottleneck links.

Remark 1. *There is a subtlety that arises when employing the linear equation (1) for incremental flows in the links, as discussed in [14]. If a bottleneck link is using its maximum capacity, it will saturate its outgoing flow rate, so that a positive increase δx in one of its source flows will not be propagated to bottlenecks downstream. This saturation nonlinearity would severely complicate the analysis. To avoid this issue we will make the following*

Assumption 1. *The target rate c_l for each link is slightly lower than its actual capacity.*

This assumption makes (1) valid around equilibrium, and also serves a separate, desirable purpose: since no link is really saturated, equilibrium queues will be empty, at least when one disregards stochastic queueing effects. In particular that there is no (deterministic) queueing delay, and the delays τ_i correspond exclusively to propagation. See [11, 8] for further discussion on this topic.

The next assumption helps eliminate degenerate cases from the network.

Assumption 2. *The matrix $R = R_f(0) = R_b(0)$ is of full row rank.*

This means that there are no algebraic constraints between bottleneck link flows. This rules out, for instance, the situation where all flows through one link also go through another. Note, however, that our R matrices refer now to bottlenecks only, and typically in the above scenario only one of the links would be a bottleneck, or both could be lumped into a common bottleneck. So Assumption 2 is quite generic. As a consequence, equilibrium prices are uniquely determined (see e.g. [16, 21]).

3 Control objectives and a proposed law

Having described the model, in this section we lay out a series of objectives for the feedback control laws in purely local (linearized) terms. These will lead us to conjecture a candidate local control law, which we argue is the simplest that can satisfy our requirements, and at the same time allows for the required decentralized implementation. In the next section we will prove that it actually achieves our objectives.

A first objective is that the target capacity c_l is matched at equilibrium; as in [16], this can be achieved when prices *integrate* the excess capacity:

$$\dot{p}_l = K_l y_l.$$

Is this the only choice? If one wishes exact tracking of capacity, there must be an integrator in the loop; and as explained in the next section, to have stability we must perform the integration at the lower-dimensional end of the problem, i.e. at the links. So the above is the simplest law consistent with this objective; the constant K_l will be chosen later.

The next, main objective is to have local dynamic stability for arbitrary network delays, link capacities and routing topologies.

Regarding delays, note that these are the only dynamics of the open loop, therefore the only thing that sets a time-scale to our closed loop behavior. Suppose, for a given controller, a certain response is obtained for certain values of the delays. If now one were to increase all delays by the same common factor, it is clear that, by appropriately re-scaling the control law, we could obtain the same response up to a change in time-scale. Furthermore, it is conceivable that such "delay-scheduling" could be performed automatically since round-trip times are known to sources.

To gain more insight into this issue, consider first a single-link, single-source problem. The network delay and the above link integrator will yield a term

$$\frac{e^{-\tau s}}{s}$$

in the loop transfer function. Thinking e.g. in terms of its Nyquist plot, instability will always occur at high values of τ unless the gain is made a function of τ. Indeed, introducing a gain K/τ in the loop (specifically at the source) leads to a loop gain

$$K\frac{e^{-\tau s}}{\tau s}$$

which is scale-invariant: namely, its frequency response is a function of $\tau\omega$, so Nyquist plots for all values of τ would fall on a single curve, and by choosing K appropriately one can ensure stability for all τ. Further, the time responses of such a loop will give the desired invariance up to scale. Inspired by this, for the multi-source case we are led to include a factor $1/\tau_i$ in the control gain at each source i.

Now we turn to invariance to capacity and routing. For this objective it is crucial that the overall loop gain introduced by the routing matrices R_f, R_b be kept under control; intuitively, as more links participate in the feedback the gain must be appropriately scaled down. The difficulty in implementing this in a decentralized fashion, without access to the global routing information.

One source of information that can be exploited is that at equilibrium, the aggregate source rates add up to capacity, so $Rx_0 = c$. Since sources know their rate, and links their capacity, one is led to introduce:

- A gain $\frac{1}{c_l}$ at each link.
- A gain x_{0i} at each source.

This makes the map from the vector of q's to each p_l have, in addition to the integrator, a gain bounded as

$$\left| \frac{1}{c_l} \sum_i x_{0i} e^{-j\tau_{i,l}^f \omega} \right| \leq \frac{1}{c_l} \sum_i x_{0i} = 1.$$

In the multiple link case, we are still left with the need of bounding the gain of the backward path from p's to q's. The analogous strategy would be in this case to introduce a gain $\frac{1}{M_i}$ at each source, M_i being the number of bottleneck links in the source's path. For a discussion on its implementation, see Section 5.

Summarizing the above requirements in their simplest possible form, we propose the following linear control laws:

- For the source, a gain (between q_i and x_i) of

$$-\kappa_i := -\frac{\alpha_i x_{0i}}{M_i \tau_i} \tag{6}$$

  where α_i is a parameter to be chosen, x_{0i} is the equilibrium rate, τ_i is round-trip delay as before, and M_i is the number of bottleneck links in source i's path. The sign is used to provide negative feedback from prices to rates.

- For the link, an integrator with gain normalized by capacity,

$$p_l = \frac{1}{c_l s} y_l. \tag{7}$$

  Note that the normalization gives this price units of *time*. In fact if c_l were the actual link capacity (it is not, see Assumption 1), p_l would represent the queueing delay at the link and is the price signal used in TCP Vegas [3, 17]; so here we can think of p_l as a "virtual" queueing delay (see [13] for related ideas).

It is shown below that these simple control laws we conjectured will indeed satisfy the desired objectives.

4 Linear stability results

With the source and link controllers described above, we proceed to study the linearized stability of the closed loop system from Figure 1. We express the overall, multivariable feedback loop in the classical configuration of Figure 2, with loop gain transfer function

$$L(s) = R_f(s)\mathcal{K}R_b^T(s)\mathcal{C}\frac{I_L}{s}, \tag{8}$$

where I_L is the identity matrix of size L, and we have introduced the notation

$$\mathcal{K} = \mathrm{diag}(\kappa_i), \quad \mathcal{C} = \mathrm{diag}(\frac{1}{c_l}).$$

Stability analysis can be reduced to Figure 2 as long as there are no unstable pole/zero cancellations within $L(s)$; note here that except for the integrators, the remaining blocks are stable. The theorem below provides stability conditions for such multivariable loops with integral control.

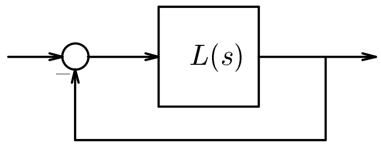


Figure 2: Overall feedback loop

Proposition 1. *Consider a standard unity feedback loop, with* $L(s) = \gamma F(s)\frac{I}{s}$. *Suppose:*

(i) $F(s)$ *is analytic in* $Re(s) > 0$*, and* $\|F(s)\| \leq \beta$ *in* $Re(s) \geq 0$.

(ii) $F(0)$ *has strictly positive eigenvalues.*

(iii) For all $\gamma \in (0, 1]$, -1 *is not an eigenvalue of* $L(j\omega)$, $\omega \neq 0$.

Then the feedback system is stable for all $\gamma \in (0, 1]$.

Proof: Note that (ii) implies $F(0)$ is invertible so there is no unstable cancellation at $s = 0$; so we can just impose the fact that

$$(I + L(s))^{-1} = s\big(sI + \gamma F(s)\big)^{-1}$$

is analytic and bounded in $Re(s) \geq 0$, or equivalently that there are no solutions in $Re(s) \geq 0$ to

$$\varphi_\gamma(s) := \det(sI + \gamma F(s)) = 0.$$

Using (ii) and continuity, select $\epsilon > 0$ such that for $|s| \leq \epsilon$, and $Re(s) \geq 0$, we have

$$Re(eig(F(s))) > 0. \tag{9}$$

Claim: For $0 < \gamma < \epsilon/\beta$, $\varphi_\gamma(s)$ has no roots in $Re(s) \geq 0$.

Indeed, $|s| \leq \epsilon$, $Re(s) \geq 0$ cannot be a root, otherwise $-s/\gamma \in eig(F(s))$, contradicting (9).

For $|s| > \epsilon$, note that

$$\left\| \frac{\gamma F(s)}{s} \right\| < \left\| \frac{\gamma F(s)}{\epsilon} \right\| < \left\| \frac{F(s)}{\beta} \right\| \leq 1,$$

so $I + \gamma F(s)/s$ is invertible and $\det(sI + \gamma F(s)) \neq 0$. This establishes the claim.

We now increase γ from this small value to 1. If at some point there is a root of $\varphi_\gamma(s)$ in $Re(s) \geq 0$, by continuity of the roots there must be a $\gamma \leq 1$ where the root is in $Re(s) = 0$. Since $s = 0$ is never a root, we conclude that

$$\det(j\omega I + \gamma F(j\omega)) = 0$$

for some $\omega \neq 0$. This means $-1 \in eig(L(j\omega))$, contradicting (iii).

■

In essence, the above conditions are a "nominal" stability requirement for small γ, that says that we have strictly negative feedback of enough rank to stabilize all the integrators, and a "robustness" argument that says we can perform a homotopy to $\gamma = 1$ without bifurcating into instability.

Returning now to the $L(s)$ in (8), we take

$$F(s) = R_f(s)\mathcal{K}R_b^T(s)\mathcal{C}$$

We will later add the scaling γ. Note that (i) is automatically satisfied. For (ii), note that

$$eig(F(0)) = eig(\mathcal{C}^{\frac{1}{2}}R\mathcal{K}R^T\mathcal{C}^{\frac{1}{2}})$$

where $R = R_f(0) = R_b(0)$ is the static routing matrix. Assuming R has full row rank, condition (ii) will hold. Here we see the importance of putting the integrators at the links (the lower dimensional portion). If, instead, we tried to integrate at the sources, the resulting feedback matrix at DC would never have enough rank to stabilize the (larger) number of integrators.

What remains is to establish (iii). For this purpose we bring in some more notation:

$$X_0 = \mathrm{diag}(x_{0i}), \quad \mathcal{M} = \mathrm{diag}(\frac{1}{M_i}), \quad \Lambda(s) = \mathrm{diag}(\lambda_i(s)), \quad \lambda_i(s) = \frac{\alpha_i e^{-\tau_i s}}{\tau_i s}.$$

We incorporate this notation to rewrite $L(s)$, for $s \neq 0$, in a more convenient form. Invoking (5), write

$$R_b(s) = R_f(-s)\mathrm{diag}(e^{-\tau_i s}),$$

and

$$R_b^T(s) = \mathrm{diag}(e^{-\tau_i s})R_f^{\sim}(s)$$

where $R_f^{\sim}(s) = R_f^T(-s)$ is the adjoint system. Then we have, for $s \neq 0$,

$$L(s) = R_f(s)X_0\mathcal{M}\Lambda(s)R_f^{\sim}(s)\mathcal{C} \tag{10}$$

We are now ready to tackle the robustness argument.

Theorem 2. *Consider an equilibrium point where rates match capacity, i.e.* $c = R_f(0)x_0$. *Let* $\alpha_i < 1$ *and the delays be arbitrary. Then with* $L(s)$ *as in (10),* $-1 \notin eig(L(j\omega))$, $\omega \neq 0$.

We first prove the following

Lemma 3. *If* $\alpha_i < 1$, *then*

$$Re[\lambda_i(j\omega)] > -1 \tag{11}$$

for all $\omega \neq 0$, τ_i.

Proof: Set $\theta = \tau_i\omega$, and write

$$\lambda_i(j\omega) = \frac{\alpha_i e^{-j\theta}}{j\theta} = \alpha_i\left(\frac{\cos(\theta)}{j\theta} - \frac{\sin(\theta)}{\theta}\right).$$

Therefore

$$Re(\lambda_i(j\omega)) = -\alpha_i\frac{\sin(\theta)}{\theta} \geq -\alpha_i > -1.$$

■

Proof of Theorem 2: Since nonzero eigenvalues are invariant under commutation, and also many of the factors in (10) are diagonal, we observe that

$$-1 \in eig\Big(L(j\omega)\Big) \Longleftrightarrow -1 \in eig\Big(P(j\omega)\Lambda(j\omega)\Big),$$

where

$$P(j\omega) = \mathcal{M}^{\frac{1}{2}}X_0^{\frac{1}{2}}R_f(j\omega)^*\ \mathcal{C}\ R_f(j\omega)X_0^{\frac{1}{2}}\mathcal{M}^{\frac{1}{2}} > 0.$$

Claim: The eigenvalues of $P(j\omega)$ are in the interval $[0,1]$.

Note first that they are non-negative, so we can just bound the spectral radius $\rho(P)$. Now

$$\rho(P) = \rho\Big(\mathcal{M}R_f(j\omega)^* \ \mathcal{C} \ R_f(j\omega)X_0\Big) \leq \|\mathcal{M}R_f(j\omega)^*\| \cdot \|\mathcal{C} \ R_f(j\omega)X_0\|.$$

Any induced norm will do, but if we use the l_∞-induced (max-row-sum) norm, we find that the rows of $\mathcal{C} \ R_f(j\omega)X_0$ are made of terms

$$\frac{1}{c_l} e^{-\tau_{l,i}^f j\omega} x_{0i}$$

for the i's that use link l. Since at equilibrium the sum

$$\sum_{i \text{uses } l} \frac{1}{c_l} x_{0i} = 1,$$

(we are dealing with bottlenecks) the second norm is 1. Also the norm $\|\mathcal{M}R_f^*\|$ is equal to 1, because each row contains exactly M_i elements of magnitude $1/M_i$. So we have $\rho(P) \leq 1$ as claimed. Indeed, it is not difficult to see that $\rho(P) = 1$ at $\omega = 0$, the eigenvector being the vector of all ones. So we conclude, as claimed, that

$$0 \leq P \leq I. \tag{12}$$

Now suppose $-1 \in eig(P(j\omega)\Lambda(j\omega))$ for some $\omega \neq 0$. We thus have a vector $u \neq 0$ such that

$$y = \Lambda u, \quad u = -Py.$$

Write

$$y^*u + u^*y + 2u^*u = u^*(\Lambda + \Lambda^* + 2I)u > 0$$

using (11). But also

$$y^*u + u^*y + 2u^*u = y^*(-2P + 2P^2)y = 2y^*P(P-I)y \leq 0,$$

using (12). This is a contradiction.

■

Remark 2. *Some elements of the preceding proof, in particular the use of l_∞ induced norms to prove a spectral radius bound, are inspired by the work of [8] for another set of control laws. We supplement this here with an argument of the style of robust control, based on quadratic graph separation.*

Theorem 2 establishes the requirement (iii) in Proposition 1, except for the factor γ; note however that scaling down by γ is equivalent to making the α_i smaller, and the result still applies. To summarize, we have:

Theorem 4. *Let $R_f(s)$, $R_b(s)$ denote the routing matrices of sources in relation to the bottleneck links (i.e., the equilibrium saturates all links, $c = R_f(0)x_0$). Suppose $R_f(0) = R_b(0)$ has full row rank. Then the system with source controller gain*

$$-\kappa_i = -\frac{\alpha_i x_{0i}}{M_i \tau_i},$$

and link controller $\frac{1}{c_l s}$ is linearly stable for arbitrary delays and link capacities.

We have modeled a law consisting only of the integrator and delay dynamics, but otherwise with only gains at link and source. We end this section with some comments about more general loops which would satisfy our requirements, possibly including additional dynamics. Key conditions for something like this to work are:

1. There can be no more pure integrators. Otherwise the loop gain $\Lambda(j\omega)$ can never be bounded to a right half-plane as in Lemma 3, and the preceding argument breaks down. For instance, the second order link laws used in REM [2] would not qualify.

2. To preserve the scale-invariance to delay, additional dynamic terms should be "clocked" at the rate of the round-trip time (and hence located at the sources). For instance a term

$$\frac{1}{\tau_i s + \beta_i}$$

could be added to the source controller; restrictions on β_i would be required.

5 Implementation and the equilibrium structure

We have presented a linearized control law with some desirable stability properties. When it comes to implementation, two types of issues arise:

- How can one embed such linear control laws in a global, nonlinear control scheme whose equilibrium would linearize as required? What freedom does this implementation leave in selecting the equilibrium structure?

- How does the necessary feedback information get communicated in a real, packet network?

These questions are now discussed.

5.1 Global implementation by nonlinear laws

In this section we seek feedback laws for links and sources which apply globally, and have suitable equilibrium points, around which the linearization performs as required in Section 4. For this section we restore the notation x, y, p, q to represent the relevant variables, not their perturbations around equilibrium.

For control at the links, since the control gains are constant it is straightforward to implement the price dynamics as

$$\dot{p}_l = \begin{cases} \frac{y_l - c_l}{c_l} & \text{if } p_l > 0; \\ \max\{0, \frac{y_l - c_l}{c_l}\} & \text{if } p_l(t) = 0. \end{cases} \tag{13}$$

Therefore prices integrate excess capacity in a normalized way, and are saturated to be always non-negative. At equilibrium, bottlenecks with nonzero price will have $y_{l0} = c_l$ as required. Non-bottlenecks with $y_{l0} < c_l$ will have zero price.

For the sources, the problem is somewhat more involved since the linearized gains depend on the equilibrium rate x_{0i}, and also depend on parameters τ_i and M_i. Let us assume those parameters are constant for the moment, and focus on the mapping between q_i and x_i. The simplest thing to try is a static, strictly decreasing function

$$x_i = f_i(q_i). \tag{14}$$

Indeed this choice of source and link laws makes this control a special case of the class of algorithms proposed in [16][1]. These algorithms are interpreted as carrying out a distributed computation to maximize the aggregate source *utility*

$$\sum_i U_i(x_i),$$

subject to capacity constraints, $y_l \leq c_l$ for all l. Here the utility function $U_i(\cdot)$ is related to f_i by $U_i' = (f_i)^{-1}$, and is thus strictly concave. One can also turn the question around and ask for the desirable utility functions, tailored to applications, and use them to determine the source algorithm, $x_i = U_i'^{-1}(q_i)$. It follows from the work in [16] that all these laws, for arbitrary strictly concave utility functions, have an unique equilibrium, and are stable in the absence of delay. The question we address here is what constraints do our conditions for linear stability in the *presence* of delay impose on f_i, and thus on the allowable utility functions.

Remark 3. *As shown in [15], all major TCP and active queue management algorithms have a similar interpretation as distributed optimization for respective utility functions. They thus have equilibria as in the basic algorithm from [16], but their dynamics are in general more complicated.*

[1] except that here we are using continuous time models.

Now given an equilibrium point x_0, q_0, let us impose the linearization requirement

$$\left.\frac{\partial f_i}{\partial q_i}\right|_{q_{0i}} = -\frac{\alpha_i x_{0i}}{M_i \tau_i},$$

for some $0 < \alpha_i < 1$, as needed in (6). In fact the parameter α_i could vary with the operating point, but let us assume initially that it is constant. Then the above gives a differential equation

$$\frac{\partial f_i}{\partial q_i} = -\frac{\alpha_i f_i(q_i)}{M_i \tau_i},$$

that can be solved analytically, and gives the control law

$$x_i = f_i(q_i) := x_{\max,i}\, e^{-\frac{\alpha_i q_i}{M_i \tau_i}}$$

as the static source law. Here $x_{\max,i}$ is a maximum rate parameter, which can vary for each source, and in fact can also be scheduled to depend on M_i, τ_i. All that is required is that it does not depend on q_i. So we find that an *exponential backoff* of source rates as a function of aggregate price, with slope limited by the round-trip times τ_i, number of bottlenecks M_i, can provide the desired control law, together with the link control in (13).

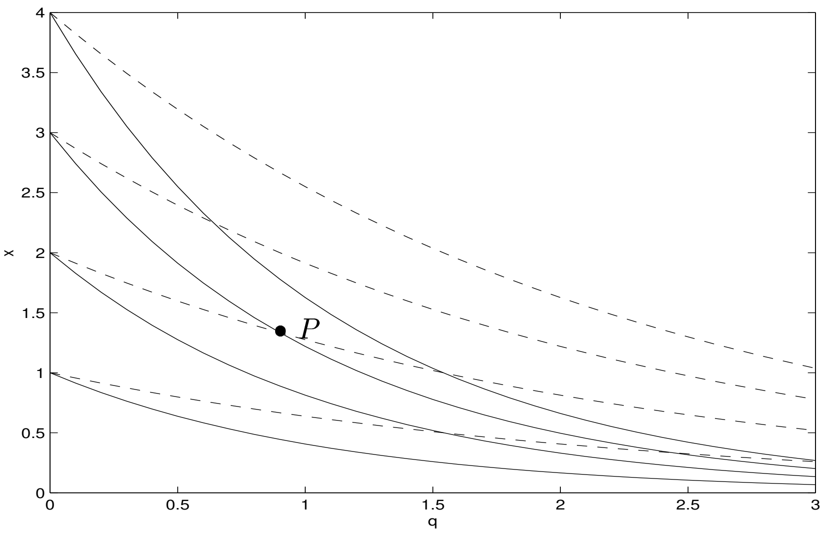


Figure 3: Source control laws for $\tau = 1$ (solid) and $\tau = 2$ (dashed)

To gain more insight into these exponential laws, consider the single link case $M_i = 1$, and two sources which use the same $\alpha_i = 0.9$ but have different round trip times τ_i. The static control laws for various choices of $x_{\max}$ are depicted in Figure 3. We see that the curves are more "elastic" for the smaller delays; in other words, they are more responsive in varying their rate as a function of price. The longer the delay, the more conservative the source control is, in order to avoid instability.

Note that if we used the same $x_{\max}$ for both sources, then the source with larger delay would get a larger rate, (this is similar to a version of TCP Vegas, and opposite of TCP Reno, see [17]), and the disparity could be significant. However one can inversely schedule $x_{\max,i}(\tau_i)$, aiming for sources to have the same rate around a nominal price value q_{0i}, as in the point P indicated in the figure. To achieve this we can, for instance choose

$$x_{\max,i}(\tau_i) = e^{\frac{\alpha_i q_{0i}}{M_i \tau_i}},$$

leading to

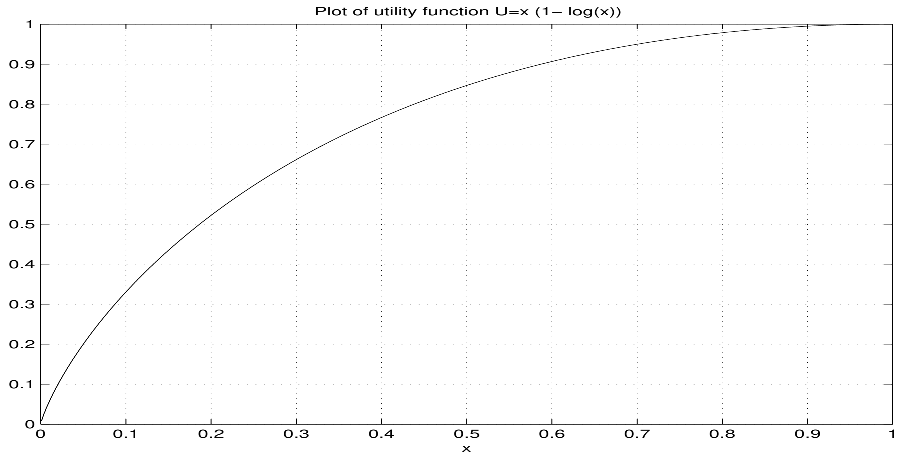
$$x_i(q_i, \tau_i) = e^{-\frac{\alpha_i (q_i - q_{0i})}{M_i \tau_i}}.$$

Then at price q_{0i} the achieved rate would be independent of τ_i.

The corresponding utility function (for which $f_i = (U_i')^{-1}$) is

$$U_i(x) = \frac{M_i \tau_i}{\alpha_i} x \left[1 - \log\left(\frac{x}{x_{\max,i}} \right) \right], \quad \text{for } x \leq x_{\max,i};$$

an illustrative sketch is given below.



Some more comments regarding the parameters M_i, τ_i.

- With our assumptions, recall equilibrium queues are empty, and so the equilibrium τ_i only correspond to propagation delays. However in a transient scenario the controller would be scheduled by the real-time round-trip delay $\tau_i(t)$.

- Regarding M_i, an issue arises: from a global perpective, sources cannot take this number of bottlenecks as given because it will depend itself on the equilibrium point, and therefore on the static laws employed. The simplest solution is to take M_i to be a constant *upper bound* on the number of bottleneck links in the path of source i, assumed known a priori. Then the proof of Theorem 2 will hold, and we will have the desired stability, albeit with some conservatism. In the next subsection we discuss a way for source i to estimate the true number of bottleneck links in real time, where a link would consider itself a bottleneck if currently $y_l(t) \geq c_l$. This would suggest the use of a less conservative, time varying law with $M_i(t)$ instead of M_i.

We remark that while the above implementations will behave as required at equilibrium, we have not provided results regarding convergence to equilibrium in the global nonlinear case.

Are these exponential laws the only possible choice? Among static laws, there is a further degree of freedom in letting the paramater α_i be a function of the operating point. In general, we would allow any mapping $x_i = f_i(q_i)$ that satisfies the differential inequality

$$0 \geq \frac{\partial f_i}{\partial q_i} \geq -\frac{f_i(q_i)}{M_i \tau_i}. \tag{15}$$

This essentially means that the static function is such that for fixed M_i, τ_i, it intersects the corresponding family of exponential curves as in Figure 3 with less absolute slope. This provides for a further degree of freedom in the static laws designs, and hence in the corresponding utility functions for which our stability theorem holds.

Finally, we ask the question: can more freedom be gained in the source control law by introducing additional dynamics? For example, by considering the general nonlinear controller (from q_i to x_i, subindex is dropped)

$$\dot{z} = g(z, q), \tag{16}$$
$$x = h(z, q). \tag{17}$$

Clearly, the linearization of such a law would give additional dynamics and not only the static gain (6); however as remarked at the end of the previous section it is possible that a generalized stability theorem can be obtained. What would be required in any event, is that the zero-frequency component of the linearized transfer function from q to z have the scalings in (6). This "DC" gain can be found by linearization to be

$$\left.\frac{\partial h}{\partial q}\right|_0 - \left.\frac{\partial h}{\partial z}\right|_0 \left(\left.\frac{\partial g}{\partial z}\right|_0\right)^{-1} \left.\frac{\partial g}{\partial q}\right|_0.$$

Note that $\frac{\partial g}{\partial z}$ must be non-singular since by our earlier remark, there cannot be any pure integrators at the source.

It is not difficult to show, however, that the above DC gain is the same as the derivative

$$\frac{\partial x_0}{\partial q_0}$$

along the equilibrium surface. Since this static information is what determines the utility functions and the equilibria, we see that constraining this gain as in (6) means that there is no added freedom in the choice of equilibria as compared to the static control law. There could still be reasons to use a dynamic controller (e.g., to provide smoothing to noisy estimates of q, see below), but it is important to recognize that stability does pose constraints on the allowable equilibrium structure.

5.2 Signaling requirements

We briefly discuss here the information needed at sources and links to implement the dynamic laws we defined, and resulting communication requirements.

Links must have access to their aggregate flow y_l; this quantity is not directly known, but can be estimated from arriving packets (see, e.g. [2]). Note also that since the rate is integrated there is smoothing already built into the process. The target capacity c_l is assumed known.

Sources must have access to the round-trip time τ_i, which can be obtained by timing packets and their acknowledgments, and two parameters from the network: the aggregate price q_i, and the number of bottlenecks M_i. To communicate q_i, the technique of random exponential marking [2] can be used. Here, an ECN bit would be marked at each link l with probability

$$1 - \phi^{-p_l}, \quad \phi > 1.$$

Assuming independence, the overall probability that a packet from source i gets marked is (see [2])

$$1 - \phi^{-q_i}$$

and therefore q_i can be estimated from marking statistics. Alternatively packet dropping can be used in lieu of marking as in other active queue management schemes [4]. To estimate q_i requires the knowledge of ϕ which is presumably a global constant set a priori.

Regarding M_i, in the simplest implementation one would simply employ an upper bound, which could be a globally known constant, or based on the total number of links in the route, found e.g. from traceroute information in IP.

For a more aggressive control one would need to communicate in real-time, how many links are bottlenecks. This can be done in an analogous way as what is done with prices, using a second ECN bit. Assume that links which are bottlenecks mark this bit with probability

$$1 - \phi^{-1},$$

and those who are not do not mark. Then the probability of marking during a route is

$$1 - \phi^{-M_i},$$

which allows for the estimation of M_i at the sources in real time. As discussed above, this introduces a time-varying source law whose closed loop stability requires further study.

6 Conclusion

We have obtained decentralized congestion control laws, which we have shown to provide local stability for arbitrary network routing, delay, and link capacities, improving certain limitations of current TCP. The required stability theorem employs tools from multivariable robustness theory. While these conditions are only sufficient, we have provided arguments that indicate that under decentralized information, these controllers are likely to be the simplest consistent with the above general requirements, and the exact matching of a target link capacity. We have also provided an implementation, based on integration at the links and a static control law at the source, that has desirable equilibria and linearizes as required. This law is a special case of those considered in [16], which are interpreted as maximizing aggregate source utility; from this point of view the result in this paper poses constraints on the allowable classes of utility functions that would yield the desired local stability. Finally we have described mechanisms based on marking ECN bits that can provide approximate implementations of these control laws.

A number of topics remain open for further research. A first issue is whether one can provide a global stability proof for these control laws of the type available in other cases, and for the case where $M_i(t)$ is the number of congested links, not just an upper bound. The difficulty here is that one must deal in this case with nonlinear, delay-differential equations. A second line of open questions concerns the effect of various random factors (marking, traffic fluctuations, etc.) in the performance of this system. In this regard, note that our system was designed to a target capacity c_l, somewhat below the true capacity. Under random effects, queueing delays inevitably occur and it is unclear how our current delay compensation is affected. Finally, we note that if one relaxes the requirement of exact capacity tracking (by our assumption, the target would be below actual capacity anyway), one can avoid the integration at links, opening the door for laws such as those in [10, 8]. A question we will address in the future is whether tools as in the present paper can be brought to bear to extend the argument in [8] to show stability in the case of arbitrary delays, as suggested by empirical evidence.

References

[1] E. Altman, T. Basar and R. Srikant. "Congestion control as a stochastic control problem with action delays", *Automatica*, December 1999.

[2] S. Athuraliya and S. H. Low, "Optimization flow control – II: Implementation", *preprint*, http://netlab.caltech.edu, May 2000

[3] Lawrence S. Brakmo and Larry L. Peterson. TCP Vegas: end to end congestion avoidance on a global Internet. *IEEE Journal on Selected Areas in Communications*, 13(8), http://cs.princeton.edu/nsg/papers/jsac-vegas.ps. October 1995.

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

[5] C. Hollot, V. Misra, D. Towsley and W.-B. Gong, "A control theoretic analysis of RED," to appear in IEEE INFOCOMM 2001.

[6] C. Hollot, V. Misra, D. Towsley and W.-B. Gong, "On Designing improved controllers for AQM Routers Supporting TCP Flows", to appear in IEEE INFOCOMM 2001.

[7] V. Jacobson, "Congestion avoidance and control", *Proc. ACM SIGCOMM '88.*

[8] R. Johari and D. Tan, "End-to-End Congestion Control for the Internet: Delays and Stability", Cambridge Univ. Statistical Laboratory Research Report 2000-2.

[9] F. P. Kelly, "Mathematical modeling of the Internet", Fourth International Congress on Industrial and Applied Mathematics, Edinburgh, Scotland, July 1999.

[10] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness, and stability". *Jour. Oper. Res. Soc.*, 49(3), pp. 237-252, 1998.

[11] F. P. Kelly, "Models for a Self-Managed Internet", *Philosophical Transactions of the Royal Society* A358 (2000) 2335-2348.

[12] S. Kunniyur and R. Srikant, "End–to–end congestion control schemes: utility functions, random losses and ECN marks", *Proc. IEEE INFOCOMM*, 2000.

[13] S. Kunniyur and R. Srikant, "A time–scale decomposition approach to adaptive ECN marking", to appear in 2001 INFOCOMM, Anchorage. http://comm.csl.uiuc.edu:80/ srikant/pub.html

[14] Steven H. Low. Optimization flow control with on-line measurement. In *Proceedings of the ITC*, volume 16, June 1999.

[15] S. H. Low. A duality model of TCP flow controls. In *Proceedings of ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management*, http://netlab.caltech.edu, September 18-20 2000.

[16] S. H. Low and D. E. Lapsley, "Optimization flow control – I: basic algorithm and convergence" *IEEE/ACM Trans. on Networking*, Vol 7, No 6. Dec 1999.

[17] Steven H. Low, Larry Peterson, and Limin Wang. Understanding Vegas: a duality model. In *Proceedings of ACM Sigmetrics*, http://netlab.caltech.edu/pub.html, June 2001.

[18] S. Mascolo, "Congestion control in high-speed communication networks using the Smith principle", *Automatica*, 1999.

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

[20] H. Ozbay, S. Kalyanaraman, A. Iftar, "On rate-based congestion control in high-speed networks: design of an H_∞ based flow controller for single bottleneck", *Proc. American Control Conference*, 1998.

[21] F. Paganini, "On the stability of optimization-based flow control", to appear in the 2001 American Control Conference. http://www.ee.ucla.edu/~paganini/

[22] F. Paganini, "Flow control via pricing: a feedback perspective", *Proceedings 2000 Allerton Conference*, Monticello, IL.