On the stability of optimization-based flow control

Fernando Paganini[1]
UCLA Electrical Engineering
Los Angeles, CA 90095-1594.

Abstract

This paper concerns optimization-based network flow control; these recently proposed algorithms select transmission rates by maximizing a utility function for the set of sources, subject to link capacity constraints. A decentralized way to carry out this optimization has been proposed recently, based on the propagation of link prices, themselves updated dynamically. In particular we consider here the second-order update law of [8], which includes a backlog term in the price dynamics. We adopt a deterministic, continuous-time model which enforces non-negativity constraints in prices and backlogs. For this model, a Lyapunov-function based proof is given of global asymptotic stability, i.e. convergence to the optimal rates and prices. The paper concludes with simulation examples.

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. The main issues are the stability, efficiency, and steady-state fairness for these large-scale coupled dynamical systems. This is a particularly challenging problem since rates must be selected by sources in a decentralized way, with little information about the rest of the network. Existing protocols such as TCP and its variants [2, 3] employ ad-hoc probing schemes in which sources increase their rates until they detect congestion, then back off to avoid it.

Recently there has been substantial interest in a more mathematical theory of flow control (see [1, 4, 5, 7, 8] and references therein), with the objective of both providing an interpretation for the main aspects of current protocols [4, 9], and also suggesting directions for improvement. The common theme of these methods is that they can be viewed as decentralized algorithms to solve a convex optimization problem: namely, the maximization of an aggregate *utility* function across all sources, subject to link capacity constraints. Decentralization is achieved by means of pricing signals that are communicated from links to sources, which then use them to update their rates. In particular, Kelly and co-workers [4, 5] have employed continuous time models and proposed two alternate, first-order update schemes, which can be shown to be globally convergent via Lyapunov analysis; the equilibrium solves an approximation to the abovementioned optimization problem.

A related approach has been developed by Low and co-workers [7, 8], based on discrete-time models. In [7] it is shown that a gradient projection algorithm applied to the dual of our optimization problem leads directly to a decentralized algorithm, convergent to the global optimum. This first-order method has a drawback, however, observed in [8]: the algorithm is such that prices are proportional to link backlogs, and thus the equilibrium can have large backlogs. This has motivated the proposal in [8] to drive the price dynamics with an additional term involving the backlog; this "extra integrator" guarantees that any equilibrium will have empty buffers. The resulting second-order dynamics, however, has no simple gradient interpretation: consequently, despite positive empirical evidence, no stability proof has been given to date.

The main result of this paper is to provide such a proof. We adopt a continuous-time model, similar in flavor to those in [5], but applying to the second order dynamics of [8], and enforcing the non-negativity constraints for prices and backlogs implicit in this algorithm. We do not model, however, the stochastic "marking" used in [8] for price propagation; our model is deterministic. By constructing a suitable Lyapunov function, we prove global asymptotic convergence to the optimal equilibrium. Simulation examples are given to illustrate the dynamics and explore the relationship between continuous and discrete models.

2 Problem Formulation and Notation

We begin by setting up the problem in a suitable form for stability studies. We will follow in general the notation from [7, 8], with a few changes that are convenient for our development.

[1]Email: paganini@ee.ucla.edu. Research supported by NSF CAREER Award ECS-9875056, and EPRI-DoD Grant W0-8333-06.

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

$$R_{ls} = \begin{cases} 1 & \text{if source } s \text{ uses link } l \\ 0 & \text{otherwise} \end{cases}$$

For each link l we have:

- A link capacity c_l.
- A price p_l.
- A backlog b_l.
- The aggregate rate of all sources which use link l, which we denote by y_l. This notation differs from [8] where x^l is used.

The vectors $c, p, b, y \in \mathbb{R}^L$ are defined by the above components across the set of links.

For each source s we have:

- The source rate x_s.
- The aggregate price of all links used by source s, which we denote by q_s. Again this differs from [8] where p^s is used.

The vectors $x, q \in \mathbb{R}^S$ are defined by the above components across the set of sources.

The following relationships are immediate (R^T is the matrix transpose of R):

$$y = Rx, \quad (1)$$
$$q = R^T p. \quad (2)$$

Source rate computation. As explained in [7, 8], for a given total price q_s, the sources must pick the rate that maximizes

$$U_s(x_s) - x_s q_s$$

over x_s, where $U_s(x_s)$ is the source utility function, assumed to be strictly concave. [7] allows for the inclusion of maximum and minimum constraints for x_s, for simplicity we will not impose those here (as, for instance, in logarithmic utility functions $U_s(x_s) = w_s \log(x_s)$).

Assuming $U_s(\cdot)$ is differentiable, the maximum is achieved at

$$x_s = {U'_s}^{-1}(q_s),$$

where ${U'_s}^{-1}$ is the inverse function of the derivative of U_s. We denote henceforth

$$f_s(q_s) := {U'_s}^{-1}(q_s).$$

Notice that U'_s is strictly decreasing in $x_s > 0$, hence f_s is a strictly monotone decreasing function of q_s. In vector notation, we summarize the above equations for source rates as

$$x = f(q). \quad (3)$$

We assume that sources have instantaneous access to the price q_s (i.e. we do not model the marking process), and that they compute their value instantaneously. Therefore the only dynamics of the system is given by the update of prices at the links.

Price dynamics: We adopt a continuous time version of the dynamics from [8]; for each l,

$$\frac{db_l}{dt} = \begin{cases} (y_l - c_l) & \text{if } b_l(t) > 0; \\ [y_l - c_l]^+ & \text{if } b_l(t) = 0. \end{cases} \quad (4)$$

$$\frac{dp_l}{dt} = \begin{cases} \gamma(\alpha_l b_l + y_l - c_l) & \text{if } p_l(t) > 0; \\ \gamma\,[\alpha_l b_l + y_l - c_l]^+ & \text{if } p_l(t) = 0. \end{cases} \quad (5)$$

Here we have used the notation $[z]^+ := \max\{0, z\}$.

The above system of differential equations models the price update as well as the backlog dynamics, enforcing the non-negativity constraints. Here $\gamma > 0$ and $\alpha_l > 0$ are constants. We assume the links know exactly their total rate y_l. Further discussion on the comparison between this version and the discrete time algorithm in [8] is given in Section 4.

Let (b^*, p^*) be an equilibrium of the above system. We also use the notation $q^* = R^T p^*$ for the equilibrium source prices, $x^* = f(q^*)$ for the equilibrium source rates, and $y^* = Rx^*$ for the equilibrium link rates.

It is not difficult to see that we must have $b^* = 0$. Indeed, if $b^*_l > 0$ then we would have $y^*_l = c_l$ so $\dot{p}_l > 0$, which contradicts equilibrium. Now p^* need not be zero, indeed its nonzero components correspond to links where $y_l \equiv c_l$, i.e. where the capacity constraint is active (bottleneck links). This fact is the main motivation for the introduction of this second order update law in [8], as compared to the first order law in [7], which would correspond to setting $\alpha_l = 0$ in (5); in this case prices become proportional to backlogs and there is nothing to curtail the size of this backlog at equilibrium.

The relation of this algorithm with optimization is explained in detail in [7], but we briefly outline it here. The key observation is that an equilibrium point p^*, x^* satisfying the equations (1-5) will be a saddle point of the optimization

$$\min_{p \geq 0} \max_{x} \left(\sum_{s=1}^{S} U_s(x_s) + p^T(c - Rx) \right).$$

This is the Lagrangian dual of the convex program

$$\max_x \sum_{s=1}^{S} U_s(x_s),$$
$$\text{subject to} \quad Rx \le c.$$

It follows from duality theory that x^* must be the unique global optimum of the latter problem; therefore y^*, q^* are also unique. p^* need not be unique, because in general the capacity constraints might not be independent. To simplify the further development and obtain a unique equilibrium price, we make the following

Assumption: The matrix R is of full row rank.

This means that there are no algebraic constraints between link flows. Equivalently, given a vector q of aggregate source prices, there is a unique p satisfying $q = R^T p$. Removing this assumption does not affect the subsequent stability theory in a significant way, but makes the statements more complicated. Namely, in that case one has a *set* of equilibria in the system, and stability results must be formulated in terms of convergence to this set.

3 Stability

We are now in a position to state the main result of this paper:

Theorem 1. *Given the system (1-5), assume $f_s(q_s)$ is strictly decreasing in $q_s > 0$, and that R is of full row rank. Then the unique equilibrium point $b^* = 0$, p^* is globally asymptotically stable.*

Proof: The proof is based on Lasalle's invariance principle applied to a suitable Lyapunov function. We begin by defining, for each s, the function

$$\phi_s(q_s) = \int_{q_s^*}^{q_s} (x_s^* - f_s(\sigma))d\sigma.$$

Note that since $f_s(\cdot)$ is deceasing, we have $\phi_s(q_s) \ge 0$ for every q_s. Furthermore, since we have assumed the decreasing is strict (U_s is strictly concave, and there are no interval limits for x_s), we find that

$$\phi_s(q_s) > 0 \quad \text{for all } q_s \ne q_s^*.$$

Moreover, $\phi_s(q_s)$ goes to infinity with q_s.

Now introduce the candidate Lyapunov function

$$V(b,p) = \sum_{l=1}^{L} [\alpha_l \gamma \frac{b_l^2}{2} + (c_l - y_l^*)p_l] + \sum_{s=1}^{S} \phi_s(q_s).$$

Note that V is non-negative, since each of the terms is non-negative. In particular, the equilibrium link rates y_l^* are no larger than the link capacity. Also $V(b,p) = 0$ implies $b = 0$, $q = q^*$, and $p_l = 0$ for non-saturated links. Using the Assumption above, these conditions only hold for $b = 0$, $p = p^*$, i.e. the function only vanishes at equilibrium. Furthermore, this function is radially unbounded, i.e. the sets $\{(b,p) : V(b,p) \le K\}$ are bounded for each K, also using our Assumption.

We now take the derivative of $V(b,p)$ along trajectories of our system:

$$\dot{V} = \sum_{l=1}^{L} [\alpha_l \gamma b_l \dot{b}_l + (c_l - y_l^*)\dot{p}_l] + \sum_{s=1}^{S} (x_s^* - f_s(q_s))\dot{q}_s.$$

We focus on the last term above, and write it as

$$\sum_{s=1}^{S} (x_s^* - x_s)\dot{q}_s = (x^* - x)^T \dot{q} = (x^* - x)^T R^T \dot{p}$$
$$= (y^* - y)^T \dot{p} = \sum_{l=1}^{L} (y_l^* - y_l)\dot{p}_l.$$

Substituting back, we find that

$$\dot{V} = \sum_{l=1}^{L} [\alpha_l \gamma b_l \dot{b}_l + (c_l - y_l^*)\dot{p}_l + (y_l^* - y_l)\dot{p}_l] = \sum_{l=1}^{L} \nu_l,$$

where we have denoted

$$\nu_l := \alpha_l \gamma b_l \dot{b}_l + (c_l - y_l)\dot{p}_l.$$

We will now show that $\nu_l \le 0$ for each l. For this we must apply the dynamic equations (4-5), and distinguish between the four cases:

(a) $b_l > 0$, $p_l > 0$. Here

$$\nu_l = \alpha_l \gamma b_l (y_l - c_l) + (c_l - y_l)\gamma(\alpha_l b_l + y_l - c_l)$$
$$= -\gamma (y_l - c_l)^2$$

(b) $b_l = 0$, $p_l > 0$. Here

$$\nu_l = (c_l - y_l)\gamma(y_l - c_l) = -\gamma(y_l - c_l)^2$$

(c) $b_l > 0$, $p_l = 0$. Here

$$\nu_l = \alpha_l \gamma b_l (y_l - c_l)$$
$$+ (c_l - y_l)\gamma \max\{0, \alpha_l b_l + y_l - c_l\}.$$

Now we distinguish between the two possibilities for the maximum. If the maximum is 0, then $\alpha_l b_l + y_l - c_l \le 0$ so

$$\nu_l = \alpha_l \gamma b_l (y_l - c_l) \le -\gamma \alpha_l^2 b_l^2 < 0$$

If the maximum is $\alpha_l b_l + y_l - c_l$, then as in case (a) we obtain

$$\nu_l = -\gamma(y_l - c_l)^2.$$

(d) $b_l = 0$, $p_l = 0$. Here

$$\nu_l = (c_l - y_l)\gamma \max\{0, y_l - c_l\}$$

Once again there are two cases:

$$\begin{aligned} \nu_l &= 0 \quad \text{for } y_l < c_l; \\ \nu_l &= -\gamma(y_l - c_l)^2 \quad \text{for } y_l \geq c_l. \end{aligned}$$

We thus confirm that $\nu_l \leq 0$ for every l, and thus $\dot{V} \leq 0$. Invoking Lyapunov's stability theorem, we conclude that the trajectory $(b(t), p(t))$ must remain bounded over time, and that the equilibrium point (b^*, p^*) is stable in the sense of Lyapunov: trajectories starting close to it will remain inside a neighborhood.

To establish the stronger claim of *asymptotic* stability, we must show that trajectories will converge to equilibrium as time goes to infinity. We do this by means of Lasalle's invariance principle (see, e.g. [6]). To apply it, we must study the set of states (b, p) where the Lyapunov derivative is zero, or equivalently $\nu_l = 0$ for each l. From the cases above, we see this can only happen when either

(i) $y_l = c_l$, or

(ii) $y_l < c_l$ and $p_l = b_l = 0$.

The Lasalle principle is based on identifying an *invariant* set inside this set $\{(b, p) : \dot{V} = 0\}$. For this purpose, suppose a trajectory $(b(t), p(t))$ moves inside this set. Then for each l we must have one of the alternatives (i) or (ii).

Claim: Under this assumption, we must have

$$\begin{aligned} b_l(t) &\equiv b_{0l} \\ p_l(t) &= p_{0l} + \alpha_l b_{0l} t \end{aligned}$$

where (b_0, p_0) is the initial state. To see this, first note that if $b_{0l} = 0$ for a certain l, then it must remain this way because $y_l - c_l \leq 0$ under both alternatives (i) and (ii). Using this fact again, now (5) implies that $\dot{p}_l = 0$ under both alternatives, so $p_l(t)$ is also constant.

If instead $b_{0l} > 0$, we are initially in alternative (i) and thus b_l stays constant due to (4), and p_l grows linearly with rate $\alpha_l b_{0l}$. Then we stay in this alternative indefinitely.

Thus the claim is established. For compactness, now denote by $\dot{p}_0$ the vector of price rates $\dot{p}_{0l} = \alpha_l b_{0l}$. We have

$$\begin{aligned} q(t) &= R^T(p_0 + \dot{p}_0 t) \\ x(t) &= f\Big(R^T(p_0 + \dot{p}_0 t)\Big) \\ y(t) &= Rf\Big(R^T(p_0 + \dot{p}_0 t)\Big) \qquad (6) \end{aligned}$$

Now we observe that for a trajectory satisfying the alternatives (i) or (ii), we have

$$\dot{p}_0^T \dot{y}(t) = \sum_{l=1}^{L} \dot{p}_{0l} \dot{y}_l(t) \equiv 0.$$

The reason is that for those entries where $\dot{p}_{0l} = \alpha_l b_{0l} \neq 0$, we are always in alternative (i) and thus y_l is constant. Taking a derivative in (6) we obtain

$$\dot{p}_0^T R f'\Big(R^T(p_0 + \dot{p}_0 t)\Big) R^T \dot{p}_0 = 0.$$

Now $f'(\cdot)$ is the diagonal matrix of derivatives $f'_s(q_s(t))$, which are all *negative.* This means that the vector $R^T \dot{p}_0$ must be zero. Using now our rank assumption on R, we conclude that $\dot{p}_0 = 0$ and therefore our candidate trajectory is in effect an equilibrium point $(b_0 = 0, p_0)$. Given our assumptions, our equilibrium is unique and therefore we have established that the only invariant set inside the set $\{(b, p) : \dot{V} = 0\}$ is the equilibrium $(0, p^*)$.

Invoking Lasalle's principle, all trajectories of our system will converge to the equilibrium, as was to be proved. ■

As an additional remark, we note that the preceding argument contains, as a special case, a proof of stability for the simpler first order rule in [7], obtained by setting $\alpha_l = 0$ in (5). Indeed if we do the same with our Lyapunov function (effectively, we eliminate the buffer terms), the argument follows through. Thus we have a continuous alternative to the discrete time argument in [7].

4 Examples

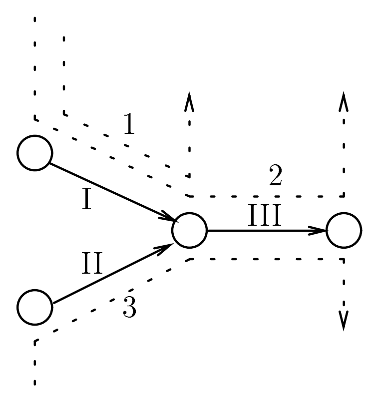

Figure 1: Example: network of 3 links, 3 sources

Our first example illustrates the dynamics of the system for a simple network of 3 sources, 3 links, of which only 2 are bottlenecks. Figure 1 depicts the network, where we have indicated the links with Roman numbers for easy distinction, all are assumed to have unit capacity, $c_l = 1$. The sources all use the utility function $U_s(x) = \log(x)$, and their link usage is depicted by dashed lines in the figure; the corresponding routing matrix is

$$R = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The continuous model (1-5) was simulated using the stiff ODE solver 'ode15s' in Matlab, using $\gamma = 0.02$, $\alpha_l = 0.1$, and initial conditions

$$p(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad b(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

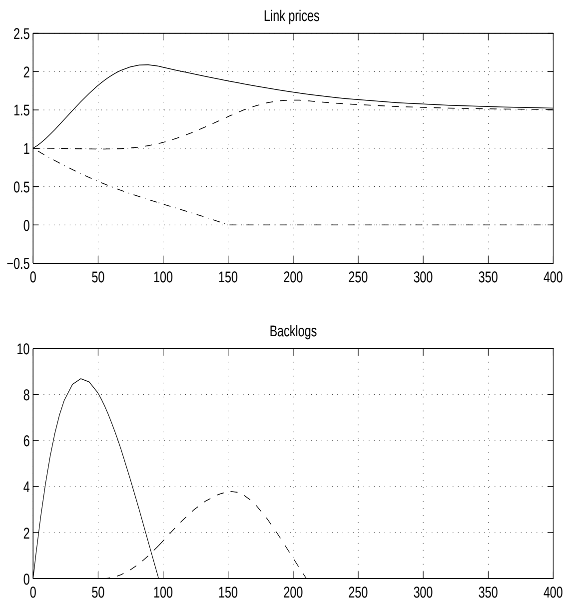

Figure 2: Link prices and backlogs: I, solid, II, dash-dot; III, dashed.

Results are depicted in Figures 2 and 3. We see that after a transient where prices are updated, and temporary backlogs occur in the first and third links, the system converges to an equilibrium of

$$p^* = \begin{bmatrix} 3/2 \\ 0 \\ 3/2 \end{bmatrix}, \; x^* = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \; y^* = \begin{bmatrix} 1 \\ 2/3 \\ 1 \end{bmatrix},$$

and zero backlogs. In particular, link II is not a bottleneck, which reflects itself on a zero equilibrium price. As expected, the Lyapunov function is monotonically decreasing during the simulation.

Our second example concerns the relationship between the continuous time model (4-5), and the discrete-time dynamics of the algorithm from [8], namely

$$b_l(t+1) = [b_l(t) + y_l(t) - c_l]^+, \tag{7}$$

$$p_l(t+1) = [p_l(t) + \gamma(\alpha_l b_l(t) + y_l(t) - c_l)]^+. \tag{8}$$

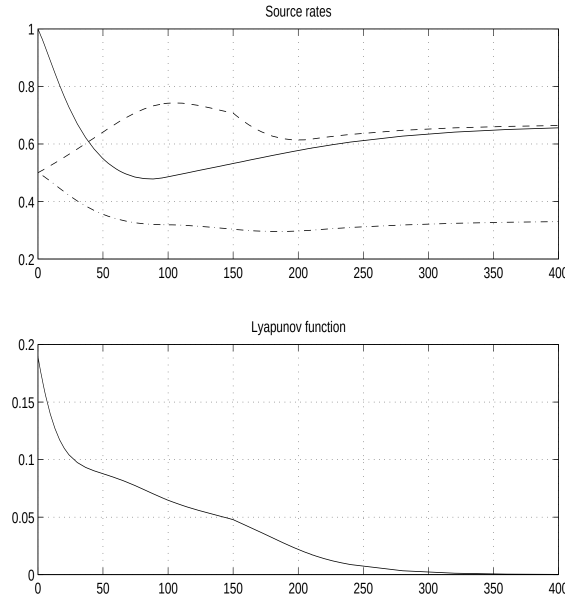

Figure 3: Source rates: 1, solid; 2, dash-dot; 3, dashed. Lyapunov function evolution.

Loosely speaking, these could be thought of as Euler steps in the numerical integration of (4-5), with unit time step. Natural questions are whether the discrete iteration is stable, and whether this can be established by the Lyapunov function used in the continuous time studies. A preliminary exploration of these questions is now done for the simplest example of a single link, single source network, comparing the discrete dynamics (7-8), with the simulation of the continuous dynamics obtained with the Matlab ODE solvers. The values used where $c = 1$ for link capacity, $\log(x)$ for the utility function, $\gamma = 0.1$, and initially $\alpha = 1$. Simulations shown in Figure 4 indicate a close approximation between both solutions, and indeed it appears that the discrete iteration is stable as well. Nevertheless, when we plot the Lyapunov function simulation in Figure 5, we notice a transient increase around $t = 10$. This means that our $V(b,p)$ is not a Lyapunov function of the discrete dynamics, and cannot be used to establish stability. If we repeat the simulation using the smaller value $\alpha = 0.1$, this difficulty disappears and now both solutions have a decreasing trajectory for $V(b,p)$.

This raises the question as to whether a stability proof could be derived for the discrete system based on the current Lyapunov function, but introducing bounds in the parameters α, γ (as was done in [7] for the first order algorithms). Alternatively, we could seek another Lyapunov function to get stronger results for the discrete case. Notice that non-strictly decreasing functions as this one often do not behave well under discretization.

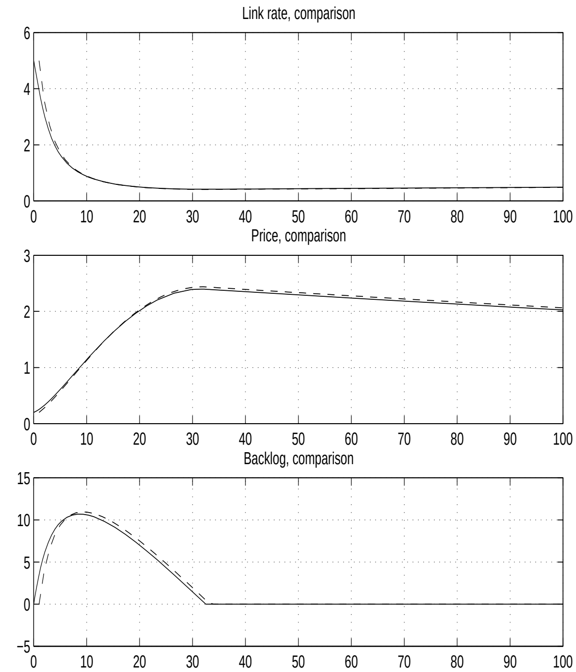

Figure 4: Comparison of continuous (solid) and discrete (dashed) dynamics.

These questions remain open for further research. Another comment is that the model (7-8) assumes prices are updated at the same rates as the buffer, i.e. at every packet. It is probably more reasonable to model the price dynamics as slower, in which case the buffer dynamics would be close to the continuous limit.

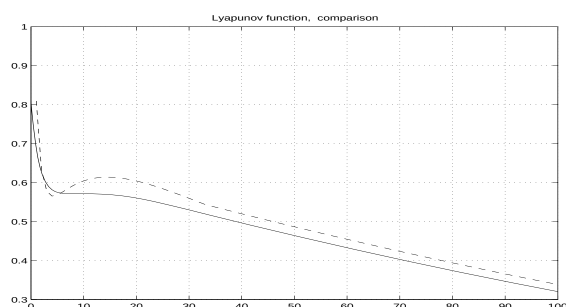

Figure 5: Lyapunov function evolution, continuous (solid) and discrete (dashed) dynamics

5 Conclusion

We have employed a continuous time, deterministic model as a way of analyzing stability for the distributed congestion control algorithm of [8]. With this model we have been able to prove global asymptotic stability for the general multi-link, multi-source case, under mild assumptions, and accounting for all nonlinearities.

An important factor that has *not* been accounted for in the above analysis is the effect of delay. It is intuitively clear from classical considerations that such a system with two integrators in the loop will have limited stability margins to delay, so that stability can only be accommodated by slowing down the response (choosing small step parameters γ, α_l). Along these lines, in [10] we employ linearized models to obtain parameter design ranges consistent with stability with respect to a uniform delay. Interesting questions for future research are to obtain such delay stability proofs with nonlinear models, and generalizations to systems with unequal delays, where parameter selection might be performed by sources using measured real-time values of the round trip delay.

Acknowledgement

I would like to thank Steven Low for initial guidance in this topic.

References

[1] C. Courcoubetis, V. Siris, and G.D.Stamoulis, "Integration of Pricing and Flow Control for Available Bit Rate Services in ATM Networks", *Proceedings of Globecom '96*, London UK.

[2] V. Jacobson, "Congestion avoidance and control", *Proc. ACM SIGCOMM '88*, pp 314-329.

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

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

[5] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness, and stability". *Journal of Operations Research Society*, 49(3), pp. 237-252, March 1998.

[6] H.K. Khalil. *Nonlinear Systems.* Prentice-Hall, 1996.

[7] 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.

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

[9] S. Low, L. Peterson and L. Wang, "Understanding Vegas: a duality model", to appear in *Proceedings of ACM Sigmetrics*, June 2001. http://netlab.caltech.edu/pub.html.

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