Multi-Path TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet

Huaizhong Han, Srinivas Shakkottai, *Student Member, IEEE*, C. V. Hollot, *Fellow, IEEE*, R. Srikant, *Fellow, IEEE*, and Don Towsley, *Fellow, IEEE*

***Abstract*—We consider the problem of congestion-aware multi-path routing in the Internet. Currently, Internet routing protocols select only a single path between a source and a destination. However, due to many policy routing decisions, single-path routing may limit the achievable throughput. In this paper, we envision a scenario where multi-path routing is enabled in the Internet to take advantage of path diversity. Using minimal congestion feedback signals from the routers, we present a class of algorithms that can be implemented at the sources to stably and optimally split the flow between each source-destination pair. We then show that the connection-level throughput region of such multi-path routing/congestion control algorithms can be larger than that of a single-path congestion control scheme.**

***Index Terms*—Congestion control, multipath routing, Nyquist stability, overlay networks.**

I. Introduction

In RECENT years, there has been a great deal of interest in congestion control in the Internet. Kelly [10]–[12] proposed a framework in which congestion control can be viewed as a mechanism for fair resource allocation in a network of elastic users, such as the Internet. This framework, and more generally, differential equation models of congestion control, can be used to study the stability of congestion control and active queue management (AQM) schemes using control-theoretic methods [5], [8], [13]–[15], [18], [19], [22], [23]. For a comprehensive review of these techniques, see [21].

In most prior models, it has been assumed that each user is assigned a single path between its source and destination. The user then reacts to congestion on its path. However, congestion may be caused indirectly due to inefficiencies in the routing protocol itself. For example, BGP is primarily a policy-based protocol and, depending upon the policy, can sometimes choose a low bandwidth path for a source, even when an alternate high bandwidth path is available. In this paper, we consider networks where multiple paths are available to each user between its source and destination, and the user can direct its flow along these paths using source routing. The amount of flow on each path is determined by the user in response to congestion indications from the routers on the path. Currently, source routing is not supported in routers in the Internet. However, two scenarios exist in which this might become possible. The first is when we overlay the network with routers that allow source routing. Consider the scenario depicted in Fig. 1, which shows a network of ISPs connected to each other by means of peering points. In this network of ISP clouds, depending on the policy employed by the ISPs, a connection from ISP2 to ISP4 may be routed via peering point *P4* even though more bandwidth may be available on a different path, say via ISP5, through peering points P1 and P3. This presents an opportunity for overlay networking to improve the service provided to the end users in the following manner: suppose that one installs overlay routers at the peering points and allows source routing at these overlay routers. Further, if the provider of the overlay routing service buys bandwidth from the ISPs, then one can create a logical network as shown in Fig. 2. This would allow us to provide a service where data transfer can simultaneously take place over multiple routes in the overlay network. Of course, it is not necessary to always place the overlay routers at the peering points. Any set of overlay routers which allow source routing would provide some

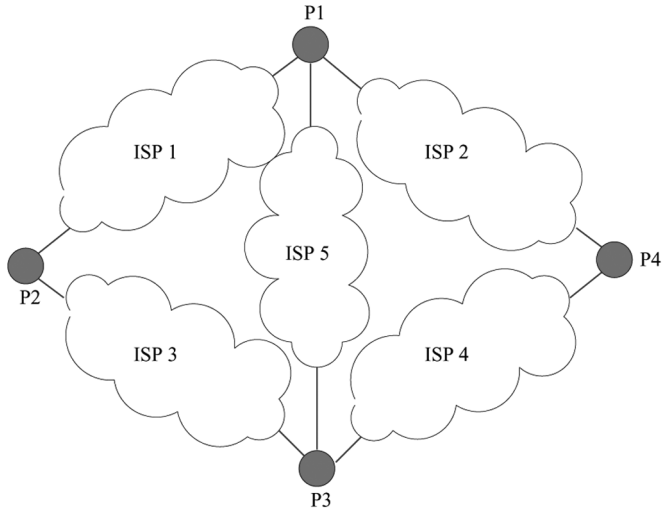


Fig. 1. A network of ISPs connected via peering points, denoted by *P1* through *P4*.

Manuscript received March 2, 2004; revised July 7, 2005; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor F. Paganini. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under Grant F30602-00-2-0542, the National Science Foundation under Grants ANI-0085848, ANI-0125979, CNS 05-19922, and CNS 05-19691, and the Air Force Office of Scientific Research under Grant F49620-01-1-0365.

H. Han, C. V. Hollot, and D. Towsley are with the University of Massachusetts, Amherst, MA 01003 USA (e-mail: hhan@ecs.umass.edu; hollot@ecs.umass.edu; towsley@cs.umass.edu).

S. Shakkottai and R. Srikant are with the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: sshakkot@uiuc.edu, rsrikant@uiuc.edu)

Digital Object Identifier 10.1109/TNET.2006.886738

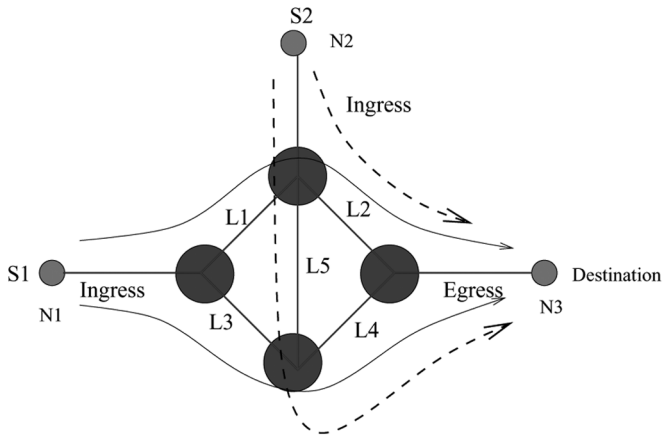


Fig. 2. A logical network formed from the network in Fig. 1 by overlaying routers and using virtual links through the ISP clouds. In this example, two sources S_1 and S_2 are transferring data to a single destination using two paths each.

path diversity which can only improve network performance. The end users would have to run a version of TCP that allows for multiple paths to be used. This requires slight changes to the protocol stack so as to allow each end-host to be associated with multiple routes and to divide the packets sent to that host amongst the different routes in an appropriate manner. The routers in the network would be unchanged—they would continue to provide feedback on each link in the traditional manner either by dropping or marking packets. The change would be in how the sender responds to feedback to alter the congestion window on each path being used.

A slightly different case arises when multi-homing is allowed. Under this scenario, a user could subscribe to two or more ISPs. The user thus obtains a diversity of paths, although each provider would be using the traditional routing algorithms. As far as path diversity is concerned, the multi-homing case looks very similar to the overlay case. A study of both scenarios is given in [1]. Our algorithm applies to either scenario—the only requirement is the ability to address these paths separately at the end hosts.

Several questions immediately arise: (i) if an overlay topology is used, where should these overlay routers be placed given an existing physical network topology?, (ii) if multi-homing is used, how much path diversity is obtained?, and (iii) given the availability of multiple routes between each source-destination pair, how does one design stable congestion control algorithms that exploit the multi-path routing capability? We are interested in the third question in this paper. This question was answered by Kelly *et al.* [12] in the case where there are no round-trip delays. There are some technical difficulties in the proof in [12] due to the possible existence of multiple equilibria which have been answered in the discrete-time context in [16]. In this paper, we derive a stability condition when there is feedback delay in obtaining the congestion information. Our approach is to modify the utility function in [12] to ensure a unique equilibrium point and then linearize the system around the equilibrium point. The multi-path routing problem was also considered in [17], [24] using a duality model; however, due to the lack of strict concavity of the objective function, only a heuristic was provided to solve the problem.

We refer to a data transfer between two nodes as a *flow*. By "node," we mean a source or destination for network traffic. Each flow may use multiple *paths* to transfer data. Obviously many flows can exist between a node-pair, but it is not necessary that all of them share the same path-set. An example of this would be when some users subscribe to the multi-path service but others do not. In Section II, we review the utility function formulation in [12], and present a congestion control scheme that can be used to maximize the net system utility. In Section III, we find a local stability condition for the proposed control scheme. We also show that the sufficient condition for the local stability of congestion-control using single-path controllers determined by Vinnicombe [23] is a special case of our result. In Section IV, we extend the proof to general window-based controllers and also consider the stability issues when there are arrivals and departures of users (as opposed to a fixed number of users considered in Sections II andIII). Using a proof technique developed in [2], we show the following result: if the flows can be split among the available routes such that the total load on each link is less than the link capacity, then our algorithm will automatically find such a split and ensure connection-level stability. Finally, we verify the results through simulations in Section V.

II. Multi-Path Controller

Consider a network consisting of a set of links and a set of routes, where each route is a collection of links. There could be multiple routes between each source-destination pair. Our goal is to find an implementable, decentralized congestion control algorithm for such a network with multi-path routing. Each user n is identified in terms of its source-destination pair as well as the set of routes available to it, the index r being used to denote a route. So we have a set of routes $\mathcal{R}(n)$ associated with user n. Similarly, associated with a route r is a set of routes $s(r)$ all associated with the same user, and with identical source and destination. We note that $r \in s(r)$. Sometimes, to avoid introducing additional notation, we also use $s(r)$ to denote the source that uses route r. For instance, in Fig. 2 suppose there were two users S_1 and S_3 between the node pair (N_1, N_3). Let S_1 subscribe to the multi-path service. Then $\mathcal{R}(S_1) = \{r_1, r_2\}$, where r_1 uses L_1 and L_2, and r_2 uses L_3 and L_4. Also $s(r_1) = s(r_2) = \{r_1, r_2\}$. On the other hand let S_3 subscribe to only a single path service. Then $\mathcal{R}(S_3) = \{r_3\}$ where r_3 uses L_1 and L_2 and $s(r_3) = \{r_3\}$. We emphasize that although r_1 and r_3 use the same links and are between the same node pair, they are indexed separately as they belong to the path-sets of different users.

In this paper, we consider the total system utility given by

$$\mathcal{U}(x) = \sum_n \left(w_n \log \left(\sum_{m \in \mathcal{R}(n)} x_m \right) + \varepsilon \sum_{m \in \mathcal{R}(n)} \log(x_m) \right) - \sum_l \int_0^{\sum_{k:l \in k} x_k} f_l(y) dy. \quad (1)$$

Here we denote the transmission rate on route m by x_m, and w_n is a weighting constant for all routes associated with user n. Also $f_l(y)$ is the price associated with link l, when the arrival rate to the link is y, which may be considered to be the cost

associated with congestion on the link y. The price functions are assumed to be strictly increasing functions with $f_l(0) = 0$ for all l. There is a small difference between the above utility function and the multi-path utility function of [12]. We have added a term containing ε to ensure that the net utility function is strictly concave and hence has a unique maximum.

We propose the following controller, which is a natural generalization of the single-path controller in [23], to control the flow on route r:

$$\dot{x}_r(t) = \kappa_r x_r(t-\tau_r)\left(w_r - \left(\sum_{m\in s(r)} x_m(t)\right) q_r(t)\right) + \kappa_r \varepsilon \sum_{m\in s(r)} x_m(t), \quad (2)$$

where τ_r is the round trip time of route r and $w_r = w_n$ for all routes r associated with user n. The term $q_r(t)$ is the estimate of the route price at time t. This price is the sum of the individual link prices p_l on that route. The link price in turn is some function f_l of the arrival rate at the link. Taking delays into account, we have

$$q_r(t) = \sum_{l\in r} f_l\left(\sum_{k:l\in k} x_k(t - d_1(l,k) - d_2(l,r))\right)$$
$$p_l(t) = f_l\left(\sum_{k:l\in k} x_k(t - d_1(l,k))\right).$$

Thus,

$$q_r(t) = \sum_{l\in r} p_l\left(t - d_2(l,r)\right).$$

Here $d_1(l,r)$ denotes the delay from source $s(r)$ to link l on route r and $d_2(l,r)$ denotes the delay from link l to source $s(r)$.

Letting $q_r(t)$ denote the marking probability on route r, we can interpret the congestion controller in (2) as follows. For each acknowledgment received, the rate is increased by $\kappa_r w_r$ and for every marked packet received, the rate is decreased by a factor $\kappa_r\left(\sum_{m\in s(r)} x_m(t)\right)$. Note that he expression $\kappa_r\left(\sum_{m\in s(r)} x_m(t)\right)$ is easily calculated at the user end. The term containing ε is an artifact of the fact that we would like the controller to have a unique equilibrium point to which all trajectories converge. The presence of the ε term also ensures that a nonzero rate is used on all paths allowing us automatically to probe the price on all paths. Otherwise, an additional probing protocol would be required to probe each high-price path to know when its price drops significantly to allow transmission on the path. Implementation of this controller would be by means of a window-based controller that is obtained by simply multiplying the transmission rate x_r by the RTT τ_r on that route. Strictly speaking, q_r cannot be interpreted as marking probability since it is the sum of prices on route r. The probability a packet is not marked on a route should be the product of the probabilities of not being marked on each link on the route. Thus, our model so far is only applicable to a scenario where q_r is very small. However, we can also deal with the more general case as we show in Section IV. Before we proceed further, we make the following comment on our congestion feedback model. Typically congestion feedback is generated at the router based on the queue length or the virtual queue length at the router. We do not explicitly include queue length in our model and assume that the congestion cost is directly a function of the arrival rate. Kelly argues that such an assumption is realistic [10] if there are stochastic disturbances and the queue length hits zero several times within an RTT. Regimes under which this assumption is valid have been identified in [6]. Alternatively, we can assume that congestion feedback is indeed based on the instantaneous arrival rate. However, the arrival rate cannot be measured instantaneously and one has to resort to some averaging to estimate the arrival rate. This can be modeled as in [23], and our results can be extended to include a model for the estimator of the arrival rate, but we do not do so here for the sake of simplicity.

Proceeding in a manner identical to Kelly *et al.* [12] we can easily show that the controller is stable and maximizes the net system utility in the absence of feedback delays. To show stability in the presence of delays, we linearize the system around its equilibrium point and obtain conditions on the gains $\{\kappa_r\}$ to ensure local stability. Linearization of (2) about $(\hat{x}_r, \hat{q}_r)$ yields

$$\dot{x}_r(t) = -\kappa_r\varepsilon\hat{x}_r^{-1}\sum_{m\in s(r)}\hat{x}_m x_r(t-\tau_r) + \kappa_r\varepsilon\sum_{m\in s(r)} x_m(t) - \kappa_r\hat{x}_r\left(\hat{q}_r\sum_{m\in s(r)} x_m(t) + \sum_{m\in s(r)}\hat{x}_m q_r(t)\right). \quad (3)$$

Taking the Laplace transform (and denoting the variables in the Laplace domain by $x(s)$ and $q(s)$) produces

$$(sI + (KXQ - \varepsilon K)(I+J))x(s) = -\varepsilon KX^{-1}X_\Sigma E(s)x(s) - KXX_\Sigma q(s) \quad (4)$$

where $X = \mathrm{diag}\{\hat{x}_r\}$, $Q = \mathrm{diag}\{\hat{q}_r\}$, $X_\Sigma = \mathrm{diag}\left\{\sum_{m\in s(r)}\hat{x}_m\right\}$, and $K = \mathrm{diag}\{\kappa_r\}$.[1] The matrix $I + J$ is block diagonal where each block $(I+J)_n$ is a matrix of *ones*, with the size of each block being equal to the number of routes available to user n. The linearized link dynamics using rate feedback are described by

$$q(s) = E(s)R^T(-s)F_pR(s)x(s) \quad (5)$$

where $E(s) = \mathrm{diag}\{e^{-s\tau_r}\}$, $F_p = \mathrm{diag}\{f_i'\}$. $R(s)$ is the routing matrix, which is defined by

$$R_{l,r}(s) = \begin{cases}\exp(-sd_1(l,r)) & \text{if route } r \text{ contains link } l\\ 0 & \text{else.}\end{cases}$$

Combining this with the rate dynamics above describes a negative feedback loop with loop return ratio function

$$L(s) = \left(sI + (KXQ - \varepsilon K)(\sqrt{KXX_\Sigma})^{-1}\times(I+J)\sqrt{KXX_\Sigma}\right)^{-1} \times\left(\varepsilon KX^{-1}X_\Sigma E(s) + E(s)\sqrt{KXX_\Sigma}R^T(-s)\times F_pR(s)\sqrt{KXX_\Sigma}\right). \quad (6)$$

[1]The notation $\mathrm{diag}\{x_r\}$ is used to denote a diagonal matrix with its r th element being the scalar x_r, or a block diagonal matrix with its r th block being the matrix x_r.

Our main result, which we discuss in the next section, gives conditions on $\{\kappa_r\}$ in terms of the RTTs, for which the linearized congestion control system is stable.

III. Stability Condition for the Multi-Path Controller

We use the generalized Nyquist Condition [7] in order to study the stability of the congestion control loop. In the context of our problem, a sufficient condition for stability would be:

- The internal dynamics of the multi-path controller

$$(sI+(KXQ-\varepsilon K)(I+J))x(s) = -\varepsilon K X^{-1} X_\Sigma E(s) x(s)$$

  is exponentially stable. We denote the return ratio associated with the above by $L_{\rm int}(s)$.
- The eigen-locus of the loop return ratio, $L(s)$ in (6), crosses the real axis to the right of the point $-1+j0$.

The first condition ensures that the internal dynamics (which is the open loop controller for $L(s)$) system has no poles in the closed right-half complex plane.

This means that when we consider the complete controller with return ratio $L(s)$, the Nyquist D-contour, which encircles the entire right half plane, encloses no open-loop poles. The generalized Nyquist condition says that the number of poles of $L(s)$ in the RHP is the difference between the number of open loop poles in the RHP (equal to the RHP poles of $L_{\rm int}(s)$, which we have set to zero by the first condition) and number of encirclements of the critical point $-1+j0$ by the eigen-locus of return-ratio (which we set to zero by the second condition). Thus, the number of poles of $L(s)$ in the closed RHP is zero, which ensures stability.

We first consider the internal dynamics $L_{\rm int}(s)$. The transfer function of the open loop controller associated with $L_{\rm int}(s)$ is $(sI+(KXQ-\varepsilon K)(I+J))^{-1}$, which has poles at zero and at $s=\kappa_r(\varepsilon-\hat{x}_r\hat{q}_r)$. So by choosing

$$\varepsilon < \hat{x}_r \hat{q}_r \tag{7}$$

we may ensure that the open loop poles of $L_{\rm int}(s)$ are in the closed left-half complex plane.

We use the Nyquist condition in order to ensure that the poles of $L_{\rm int}(s)$ are in the open LHP. The condition we obtain is

$$\kappa_r \varepsilon \frac{\hat{x}_{\Sigma n}}{\hat{x}_r}\tau_r < \frac{\pi}{2} \tag{8}$$

for all routes r, where $n=s(r)$ is the user that controls the flow over route r. The proof of the above uses the idea of the field of a matrix (more on this topic is given below) and is given in the appendix. We now derive a sufficient condition for the stability of the linearized multi-path control system given by (4) and (5). We will show that the linear system is stable if, for each user n, there exists $\hat{\tau}_n > 0$ such that

$$\kappa_r \hat{\tau}_n \hat{x}_{\Sigma n} \sum_{l\in r} f_l'(\hat{y}_l)\hat{y}_l + \kappa_r \varepsilon \hat{\tau}_n \max_n \max_{i\in\mathcal{R}(n)} \left\{\frac{\hat{x}_{\Sigma n}}{\hat{x}_i}\right\} \le \frac{1}{\frac{2}{\pi}+\max_{i\in\mathcal{R}(n)}\left|\frac{\tau_{ni}}{\hat{\tau}_n}-1\right|} \tag{9}$$

for all routes r associated with user n. Here τ_{ni} is the RTT associated with the i th route of user n and $\hat{y}_l = \sum_{k:l\in k}\hat{x}_k$ (i.e., the total equilibrium rate on link l).

The question arises as to whether we can select parameters to simultaneously satisfy conditions (7) –(9). Also, at first sight, it seems that $\frac{\hat{x}_{\Sigma n}}{\hat{x}_r}$ could be arbitrarily large, since $\hat{x}_r$ could be arbitrarily small on some routes. However, we make the following observations about the stability condition:

- Notice that the equilibrium condition ($\dot{x}_r = 0$) yields

$$\varepsilon\frac{\hat{x}_{\Sigma n}}{\hat{x}_r} = \hat{q}_r\hat{x}_{\Sigma n} - w_r < \hat{q}_r\hat{x}_{\Sigma n}.$$

  This implies that (7) is satisfied. We also notice that any κ_r which satisfies (9) automatically satisfies (8). Finally, since we have just seen that $\varepsilon\frac{\hat{x}_{\Sigma n}}{\hat{x}_r}$ is bounded for all ε, it means that κ_r can be chosen small enough to satisfy the stability condition (9).
- To calculate κ_r we can take $\hat{\tau}_n = \max_{i\in\mathcal{R}(n)}\tau_{ni} \triangleq \tau_{n,\max}$. Under this choice, a simpler sufficient condition is given by

$$\kappa_r\left(\hat{x}_{\Sigma n} f_l'(\hat{y}_l)\hat{y}_l + \pi/2\right) \le \frac{1}{\tau_{n,\max}\left(\frac{2}{\pi}+1\right)}$$

  for all links l.

To establish the above result, we make use of the concept of the field of a matrix. The field of a matrix M is defined as $\mathcal{F}(M) = \{x^T M x : x\in\mathbb{C} \text{ and } x^T x = 1\}$. It is obvious that the eigenvalues of a matrix lie in its field. There are some simple properties of matrix fields which we use [9]:

- The eigenvalues $\lambda(MN)$ of the product of the two square matrices M and N, where N is positive semi-definite, lie in the algebraic product of the fields of M and N, i.e., $\lambda(MN)\subseteq\mathcal{F}(M)\mathcal{F}(N)$. Similarly, if $0\notin\mathcal{F}(M)$ and N is as before, $\lambda(M^{-1}N)\subseteq\frac{\mathcal{F}(N)}{\mathcal{F}(M)}$.
- The field of the sum of two matrices is contained in the sum of their fields, i.e., if A and B are two matrices, $\mathcal{F}(A+B)\subset\mathcal{F}(A)+\mathcal{F}(B)$.
- The matrix field of a normal matrix is the convex combination of its eigenvalues.
- The matrix field of a block diagonal matrix is the convex hull of the union of the matrix fields of each block.

Now, consider the return ratio considered in (6). An equivalent return ratio function for this feedback loop is

$$\begin{aligned} L(s) = \Big(sI + (KXQ-\varepsilon K)\left(\sqrt{KXX_\Sigma}\right)^{-1} \\ \times(I+J)\sqrt{KXX_\Sigma}\Big)^{-1} \times E(s) \\ \times\left(\varepsilon K X^{-1}X_\Sigma + \sqrt{KXX_\Sigma}R^T(-s)\right. \\ \left.\times F_p R(s)\sqrt{KXX_\Sigma}\right). \end{aligned} \tag{10}$$

The nth block of $(KXQ-\varepsilon K)\left(\sqrt{KXX_\Sigma}\right)^{-1}(I+J)\sqrt{KXX_\Sigma}$ is the outer product

$$(KXQ-\varepsilon K)_n(\sqrt{KXX_\Sigma})_n^{-1}(I+J)_n(\sqrt{KXX_\Sigma})_n = b_n a_n^T$$

where

$$a_n \triangleq (I+J)_n(\sqrt{KXX_\Sigma})_n = [\sqrt{\kappa_{n1}\hat{x}_{n1}\hat{x}_{\Sigma n}} \quad \cdots \quad \sqrt{\kappa_{n\zeta_1}\hat{x}_{n\zeta_2}\hat{x}_{\Sigma n}}]^T \tag{11}$$

and

$$b_n \triangleq (KXQ-\varepsilon K)_n(\sqrt{KXX_\Sigma})_n^{-1} = \begin{bmatrix} \frac{\kappa_{n1}\hat{x}_{n1}\hat{q}_{n1}-\kappa_{n1}\varepsilon}{\sqrt{\kappa_{n1}\hat{x}_{n1}\hat{x}_{\Sigma n}}} \\ \frac{\kappa_{n2}\hat{x}_{n2}\hat{q}_{n2}-\kappa_{n2}\varepsilon}{\sqrt{\kappa_{n2}\hat{x}_{n2}\hat{x}_{\Sigma n}}} \\ \cdots \\ \frac{\kappa_{n\zeta_n}\hat{x}_{n\zeta_n}\hat{q}_{n\zeta_n}-\kappa_{n\zeta_n}\varepsilon}{\sqrt{\kappa_{n\zeta_n}\hat{x}_{n\zeta_n}\hat{x}_{\Sigma n}}} \end{bmatrix}. \tag{12}$$

A straightforward evaluation gives

$$\left(sI+(KXQ-\varepsilon K)(\sqrt{KXX_\Sigma})^{-1}(I+J)\sqrt{KXX_\Sigma}\right)_n^{-1} = \left(sI+b_na_n^T\right)^{-1} = \frac{1}{s\left(s+b_n^Ta_n\right)}\left(\left(s+b_n^Ta_n\right)I-b_na_n^T\right)$$

so that

$$\left(sI+(KXQ-\varepsilon K)\left(\sqrt{KXX_\Sigma}\right)^{-1}(I+J)\sqrt{KXX_\Sigma}\right)^{-1} = \mathrm{diag}_{\mathrm{block}}\left(\frac{1}{s\left(s+b_n^Ta_n\right)}\left(\left(s+b_n^Ta_n\right)I-b_na_n^T\right)\right)$$

and

$$L(s)=\mathrm{diag}_{\mathrm{block}}\left(\frac{1}{s\left(s+b_n^Ta_n\right)}\left(\left(s+b_n^Ta_n\right)I-b_na_n^T\right)\right)E(s) \times\left(\varepsilon KX^{-1}X_\Sigma+\sqrt{KXX_\Sigma}R^T(-s)F_pR(s)\sqrt{KXX_\Sigma}\right).$$

Because $\sqrt{KXX_\Sigma}R^T(-j\omega)F_pR(j\omega)\sqrt{KXX_\Sigma}$ is positive semi-definite, there exists a unitary matrix $V(j\omega)$ such that $\sqrt{KXX_\Sigma}R^T(-j\omega)F_pR(j\omega)\sqrt{KXX_\Sigma} = V(j\omega)\mathrm{diag}\{\gamma_i\}V^H(j\omega)$ where $\gamma_i \geq 0$. Then

$$\lambda(L(j\omega)) \subset Co\left\{\bigcup_n \mathcal{F}\left(\frac{1}{j\omega\left(j\omega+b_n^Ta_n\right)}\left(\left(j\omega+b_n^Ta_n\right)I-b_na_n^T\right)E_n\right)\right\} \times\mathcal{F}\left(\varepsilon KX^{-1}X_\Sigma+V(j\omega)\mathrm{diag}\{\gamma_i\}V^H(j\omega)\right)$$

and

$$\lambda(L(j\omega)) \subset Co\left\{\bigcup_n \mathcal{F}\left(\frac{1}{j\omega\left(j\omega+b_n^Ta_n\right)}\left(\left(j\omega+b_n^Ta_n\right)I-b_na_n^T\right)E_n\right)\right\} \times\left(Co\left\{\frac{\kappa_i\varepsilon\hat{x}_{\Sigma j}}{\hat{x}_i}\right\}+Co\{\gamma_i\}\right)$$

where we use Co to denote the convex hull. Using an equilibrium constraint, we now show that the b_n are proportional to the a_n. Indeed, since $KX(W-X_\Sigma Q)+\varepsilon KX_\Sigma = 0$, then

$$(KXQ-\varepsilon K)\left(\sqrt{KXX_\Sigma}\right)^{-1} = \left(QX_\Sigma^{-1}-\varepsilon(XX_\Sigma)^{-1}\right)\sqrt{KXX_\Sigma} = WX_\Sigma^{-2}\sqrt{KXX_\Sigma}.$$

If all the weights w_r associated with the nth user are equal to $\bar{w}_n$, then, recalling the definitions of a_n and b_n

$$b_n = (KXQ-\varepsilon K)_n(\sqrt{KXX_\Sigma})_n^{-1} = (WX_\Sigma^{-2})_n(I+J)_n(\sqrt{KXX_\Sigma})_n = \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n.$$

Consequently,

$$\lambda(L(j\omega)) \subset Co\left\{\bigcup_n \mathcal{F}\left(\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n\right)}\left(\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n\right)I-\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_na_n^T\right)E_n\right)\right\} \times\left(Co\left\{\frac{\kappa_i\varepsilon\hat{x}_{\Sigma j}}{\hat{x}_i}\right\}+Co\{\gamma_i\}\right). \tag{13}$$

Lemma 1: (Proof straightforward) If $\alpha, \beta \in \mathbf{R}$, then

$$|e^{-j\alpha\omega}-e^{-j\beta\omega}| \leq \omega|\alpha-\beta|$$

for all $\omega \geq 0$. □

Lemma 2: (Proof given in Appendix) Let $e^{-s\tau_{ni}}$ denote the ith routing delay element in block E_n. Then, given constant $\hat{\tau}_n > 0$

$$\mathcal{F}\left(\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n\right)}\left((j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n)I-\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_na_n^T\right)E_n\right) \subset Co\left\{\frac{e^{-j\omega\hat{\tau}_n}}{j\omega}, \frac{e^{-j\omega\hat{\tau}_n}}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n}\right\}+\mathrm{disk}\left(\max_i|\tau_{ni}-\hat{\tau}_n|\right).$$

□

Since $\frac{e^{-j\omega\hat{\tau}_n}}{j\omega}$ crosses the negative real axis of the complex plane at $-\frac{2\hat{\tau}_n}{\pi}$, and $\frac{e^{-j\omega\hat{\tau}_n}}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n}$ crosses the negative real axis to the right of $-\frac{2\hat{\tau}_n}{\pi}$, then the following is immediate.

Lemma 3: For $\hat{\tau}_n > 0$

$$\bigcup_\omega \mathcal{F}\left(\frac{1}{j\omega(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n)}\left((j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^Ta_n)I-\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_na_n^T\right)E_n\right)$$

intersects the real axis to the right of

$$-\frac{2\hat{\tau}_n}{\pi}-\max_{i\in\mathcal{R}(n)}|\tau_{ni}-\hat{\tau}_n|.$$

□

Theorem 4: The feedback system associated with loop transfer function $L(s)$ in (10) is stable if for each user n, there exists a positive number $\hat{\tau}_n$ such that

$$\kappa_r \hat{\tau}_n \hat{x}_{\Sigma n} \sum_{l \in r} f_l'(\hat{y}_l)\hat{y}_l + \kappa_r \varepsilon \hat{\tau}_n \max_n \max_{i \in \mathcal{R}(n)} \left\{ \frac{\hat{x}_{\Sigma n}}{\hat{x}_i} \right\} \leq \frac{1}{\frac{2}{\pi} + \max_{i \in \mathcal{R}(n)} \left| \frac{\tau_{ni}}{\hat{\tau}_n} - 1 \right|}$$

for all routes r belonging to the path set of the nth user.

Proof: Recall from (13)

$$\lambda(L(j\omega)) \subset Co\left\{ \bigcup_n \mathcal{F}\left(\frac{1}{j\omega\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right)} \times \left(\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right) I - \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n a_n^T \right) E_n \right) \right\} \times \left(Co\{\frac{\kappa_i \varepsilon \hat{x}_{\Sigma j}}{\hat{x}_i}\} + Co\{\gamma_i\} \right).$$

Lemma 3 shows

$$\bigcup_\omega \mathcal{F}\left(\frac{1}{j\omega\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right)} \times \left(\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right) I - \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n a_n^T \right) E_n \right)$$

intersects the real axis to the right of $-\frac{2\hat{\tau}_n}{\pi} - \max_{i \in \mathcal{R}(n)} |\tau_{ni} - \hat{\tau}_n|$. Because both $\frac{\kappa_i \varepsilon \hat{x}_{\Sigma j}}{\hat{x}_i}$ and γ_i are positive, then

$$\left(\max_i \{\gamma_i\} + \max_n \max_{i \in \mathcal{R}(n)} \left\{ \frac{\kappa_i \varepsilon \hat{x}_{\Sigma n}}{\hat{x}_i} \right\} \right) \times \left(\frac{2\hat{\tau}_n}{\pi} + \max_{i \in \mathcal{R}(n)} |\tau_{ni} + \hat{\tau}_n| \right) \leq 1$$

implies

$$Co\left\{ \bigcup_n \mathcal{F}\left(\frac{1}{j\omega\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right)} \times \left(\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n a_n^T\right) I - \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n a_n^T \right) E_n \right) \right\} \times \left(Co\left\{ \frac{\kappa_i \varepsilon \hat{x}_{\Sigma j}}{\hat{x}_i} \right\} + Co\{\gamma_i\} \right)$$

is to the right of $-1+j0$. So, $\lambda(L(j\omega))$ is to the right of $-1+j0$ and, by the Generalized Nyquist Theorem, the system is stable. The proof is completed by noting that γ_i is an eigenvalue of $\sqrt{KXX_\Sigma} R^T(-j\omega) F_p R(j\omega) \sqrt{KXX_\Sigma}$ and

$$\lambda\left(\sqrt{KXX_\Sigma} R^T(-j\omega) F_p R(j\omega) \sqrt{KXX_\Sigma} \right) \leq \left\| KX_\Sigma R^T(-j\omega) F_p \mathrm{diag}\{y_l\} \right\|_\infty = \max_n \max_{r \in \mathcal{R}(n)} \kappa_r \hat{x}_{\Sigma n} \sum_{l \in r} f_l'(\hat{y}_l)\hat{y}_l.$$

□

Corollary: If all routes sharing the same source-destination pair have the same time delay: i.e., given n, $\hat{\tau}_{ni}$ = constant for all i, then the feedback system associated with loop transfer function $L(s)$ in (10) is stable if

$$\kappa_r \hat{\tau}_n \hat{x}_{\Sigma n} \sum_{l \in r} f_l'(\hat{y}_l)\hat{y}_l + \kappa_r \varepsilon \hat{\tau}_n \leq \frac{\pi}{2}.$$

Proof: Take $\hat{\tau}_n = \hat{\tau}_{ni}$. □

Note that for $\varepsilon = 0$, this is identical to

$$\frac{\kappa_r w_r \tau_r}{\hat{q}_r} \sum_{l \in r} f_l'(\hat{y}_l)\hat{y}_l \leq \frac{\pi}{2} \tag{14}$$

which is a sufficient condition for the local stability of congestion-control using single-path controllers determined by Vinnicombe [23].

IV. Extensions and Remarks

In this section, we discuss some extensions of the algorithm presented in the previous section and also comment on the stability of the network under file arrivals and departures.

A. Multi-Path TCP: A General Class of Controllers

The stability condition derived previously can be extended to a general class of window-based algorithms of which TCP-Reno is a special case. Let the congestion window of a window-based controller be denoted by $cwnd$. Note that in the fluid model we may interchangeably use either the window-based or rate-based approaches by merely dividing the window size by RTT. For example, the equivalent window-based controller to the controller (2) is

$$\frac{1}{\tau_r}\frac{d}{dt} cwnd_r(t) = \kappa_r \left(w_r \frac{cwnd_r(t-\tau_r)}{\tau_r} - \left(\sum_{m \in s(r)} \frac{cwnd_m(t)}{\tau_m} \right) \frac{cwnd_r(t-\tau_r)}{\tau_r} \left(\sum_{l \in r} f_l(t) \right) \right) + \kappa_r \varepsilon \sum_{m \in s(r)} \frac{cwnd_m(t)}{\tau_m}.$$

In this section we will show how the stability condition extends to a class of window-based controllers, which are similar to Scalable-TCP-like algorithms studied in [23].

Consider the multi-path window-based controller

$$\frac{d}{dt} cwnd_r(t) = \frac{cwnd(t-\tau_r)}{\tau_r} \left[a \left(\sum_{m \in s(r)} cwnd_m \right)^\eta (1 - q_r(t)) - b \left(\sum_{m \in s(r)} cwnd_m \right)^\mu q_r(t) \right] + \varepsilon \sum_{m \in s(r)} cwnd_m(t) \tag{15}$$

where $\mu - \eta > 0$. Note that the above controller can be interpreted as follows:

- The term containing ε is added to ensure that the equilibrium point is unique.
- $(1-q_r(t))\frac{cwnd(t-\tau_r)}{\tau_r}$ is the rate of reception of acks for unmarked packets. Thus, for each unmarked packet, the window size is increased by $a\left(\sum_{m\in s(r)} cwnd_m\right)^{\eta}$. The term $\varepsilon\sum_{m\in s(r)} cwnd_m(t)$ also causes the window size to be increased, ensuring that it never goes to zero. As mentioned earlier, this enables the probing of paths for bandwidth.
- The rate of reception of acks for marked packets is $q_r(t)\frac{cwnd(t-\tau_r)}{\tau_r}$. Thus, upon receiving an ack for a marked packet, the congestion windows is decreased by $b\left(\sum_{m\in s(r)} cwnd_m\right)^{\mu}$.

We may choose a, b, η and μ to reflect the behavior of different TCP flavors.

The equivalent rate-based controller is

$$\tau_r\dot{x}_r = x_r(t-\tau_r)\left(a(1-q_r(t))\left(\sum_{m\in s(r)} x_m\tau_m\right)^{\eta} - bq_r(t)\left(\sum_{m\in s(r)} x_m\tau_m\right)^{\mu}\right) + \varepsilon\sum_{m\in s(r)} x_m\tau_m. \quad (16)$$

Linearizing about the equilibrium value $(\hat{x}_r, \hat{q}_r)$ and defining $\sum_{m\in s(r)}\hat{x}_m\tau_m \triangleq \chi_r$ we obtain (all other notation is identical to that used previously)

$$\tau_r\dot{x}_r = -\varepsilon\hat{x}_r^{-1}\chi_r x_r(t-\tau_r) + \varepsilon\sum_{m\in s(r)} x_m\tau_m - \hat{x}_r\left((a\chi_r^{\eta}+b\chi_r^{\mu})q_r + \left(-\eta a(1-\hat{q}_r)\chi_r^{\eta-1} + \mu b\hat{q}_r\chi_r^{\mu-1}\right)\sum_{m\in s(r)} x_m\tau_m\right). \quad (17)$$

After taking Laplace transforms, we may write the above in matrix form as

$$(sI + (X\tilde{Q} - \varepsilon I)(I+J))Tx(s) = -X\tilde{X}_{\Sigma}q(s) - \varepsilon X^{-1}X_{\Sigma}E(s)x(s) \quad (18)$$

where

$$\begin{aligned}\tilde{Q} &= \operatorname{diag}\left(-\eta a(1-\hat{q}_r)\chi_r^{\eta-1} + \mu b\hat{q}_r\chi_r^{\mu-1}\right)\\ \tilde{X}_{\Sigma} &= \operatorname{diag}\left(a\chi_r^{\eta}+b\chi_r^{\mu}\right)\\ X_{\Sigma} &= \operatorname{diag}(\chi_r)\ (\text{Note}: \chi_r \text{ is identical for all } r\in s(r))\\ T &= \operatorname{diag}(\tau_r). \end{aligned} \quad (19)$$

Now, we may take $\sqrt{T}x(s) \triangleq z(s)$. Stability of $z(s)$ implies the stability of $x(s)$. Then we have

$$\begin{aligned}&(sI + (X\tilde{Q} - \varepsilon I)(I+J))\sqrt{T}z(s)\\ &\quad = -X\tilde{X}_{\Sigma}q(s) - \varepsilon X^{-1}X_{\Sigma}E(s)\sqrt{T}^{-1}z(s)\\ &\text{and } q(s) = E(s)R^T(-s)F_pR(s)\sqrt{T}^{-1}z(s).\end{aligned}$$

We see that the above expression is similar to (4). We take the return ratio (after combining with link dynamics) as

$$L(s) = \left(sI + (X\tilde{Q} - \varepsilon I)(\sqrt{X\tilde{X}_{\Sigma}})^{-1}(I+J)\sqrt{X\tilde{X}_{\Sigma}}\right)^{-1} \times \left(\varepsilon(XT)^{-1}X_{\Sigma}E(s) + E(s)\sqrt{T^{-1}X\tilde{X}_{\Sigma}}R^T(-s) \times F_pR(s)\sqrt{X\tilde{X}_{\Sigma}T^{-1}}\right). \quad (20)$$

Proceeding as in the previous section and noting that $\sqrt{T^{-1}X\tilde{X}_{\Sigma}}R^T(-s)F_pR(s)\sqrt{X\tilde{X}_{\Sigma}T^{-1}}$ is positive semi-definite we obtain the stability condition for the Scalable-TCP like controller as there must exist a positive number $\hat{\tau}_n$ such that

$$\frac{\hat{\tau}_n}{\tau_r}(a\chi_r^{\eta}+b\chi_r^{\mu})\sum_{l\in r} f_l'(\hat{y}_l)\hat{y}_l + \varepsilon\frac{\hat{\tau}_n}{\tau_r}\max_n\max_{i\in\mathcal{R}(n)}\left\{\frac{\chi_n}{\hat{x}_i}\right\} \le \frac{1}{\frac{2}{\pi}+\max_{i\in\mathcal{R}(n)}\left|\frac{\tau_{ni}}{\hat{\tau}_n}-1\right|}$$

for all routes r belonging to the path set of the nth user.

B. Multiplicative Marking Probabilities

So far we have considered additive prices, which approximates marking probabilities only if they are very small. An interesting situation arises when the marking probability is not so low that we may approximate it as the sum of marking probabilities along the path. We show that the stability condition is unchanged when we consider the actual marking probabilities. Our approach is identical to [23]. In general the end-to-end marking probability q_r on a route r depends on the link marking probabilities, f_l, in the following manner:

$$q_r(t) = 1 - \prod_{l:l\in r}(1 - f_l(t - d_2(l,r)),$$

which linearizes to

$$q_r(t) = \sum_{l:l\in r}\frac{1-\hat{q}_r}{1-\hat{f}_l}f_l(t-(t-d_2(l,r)) \quad \text{and} \quad \hat{q}_r = 1 - \prod_{l:l\in r}(1-\hat{f}_l).$$

This translates to a change in $R(s)$ and we now have

$$\tilde{R}_{l,r}(s) = \begin{cases}\exp(-sd_1(l,r))\frac{1-\hat{q}_r}{1-\hat{f}_l} & \text{if route } r \text{ contains link } l\\ 0 & \text{else.}\end{cases}$$

The proof is unchanged and since

$$\frac{1-\hat{q}_r}{1-\hat{f}_l} = \frac{\text{probability no marking on a route}}{\text{probability no marking on a particular link of the route}} \le 1$$

the stability condition remains unchanged.

C. A Note on Connection-Level Analysis

We briefly consider arrivals and departures of congestion-controlled users in the network. As in [2], we assume that congestion control takes place at a much faster time scale than the time scale at which arrivals and departures of users occur in the network. Suppose that $\mathcal{L}$ is the set of links in the network, $\mathcal{N}$ is

the set of users (with $n \in \mathcal{N}$) in the network at some instant, the vector of rates $\bar{x}$ allocated to the users and their routes at that time instant is assumed to be a solution to the following maximization problem:

$$\max_{\bar{x}} \sum_r w_n \log \left(\sum_{m \in \mathcal{R}(n)} x_m \right) \tag{21}$$

subject to

$$\sum_{j: l \in j} x_j \leq C_l, \quad l \in \mathcal{L}$$
$$x_j \geq 0, \quad \forall\, j.$$

Note that this problem is different from the problem of maximizing (1). As in [12], (1) can be viewed as the penalty function formulation of (21). Further, as in [13], by appropriately choosing the functions $f_l(\cdot)$, and letting $\varepsilon \to 0$, the solutions to the two problems can be made identical. It is straightforward to show that if both $\bar{x}^*$ and $\bar{y}^*$ maximize (21), then

$$\sum_{\mu \in \mathcal{R}(n)} x^*_\mu = \sum_{\xi \in \mathcal{R}(n)} y^*_\xi, \quad \forall\, n \in \mathcal{N}.$$

The reason for stating the above result is that, as in the single-path case considered in [2], the departure rate of a call is determined by the amount of resources it is allocated by the network and this is unique. Then we can use the same type of analysis in [2] and show the following result.

Theorem 5: Let λ_n denote the arrival rate of sessions of type n. Let $\frac{1}{\mu_n}$ be the mean number of bits in the files of type n. Define $\rho_n = \frac{\lambda_n}{\mu_n}$ to be the offered load. Suppose that the file size distribution is exponential and independent across the users, and the file arrival processes are independent, Poisson processes. The connection-level model of the network is stable[2] if there exists $\{\tilde{\rho}_m\}$ such that $\sum_{m \in \mathcal{R}(n)} \tilde{\rho}_m = \rho_n$ and

$$\sum_{j: l \in j} \tilde{\rho}_j < C_l, \quad \forall\, l \in \mathcal{L}.$$

□

We do not provide the proof of this theorem here since it is identical to the proof in [2], with only minor differences. The interesting observation in this paper is that the multi-path stability at the congestion time scale automatically splits the traffic among the various routers to stabilize the system at the connection level. The theorem also shows that the throughput region of a network with multi-path routing can be larger than one with single-path routing. For example, consider two parallel links of unit capacity, say link 1 and link 2 and two types of users. Suppose that users of type 1 bring a workload of 0.7 and users of type 2 bring a workload of 1.1. If users of type 1 use only link 1 and users of type 2 use only link 2, then the network will be unstable. On the other hand, if users of type 2 are allowed to use two routes with the routes being links 1 and 2, then the network is stable from the previous theorem.

[2]By stability, we mean stochastic stability here, not Lyapunov stability as in the previous sections. We refer the interested reader to [2] for details.

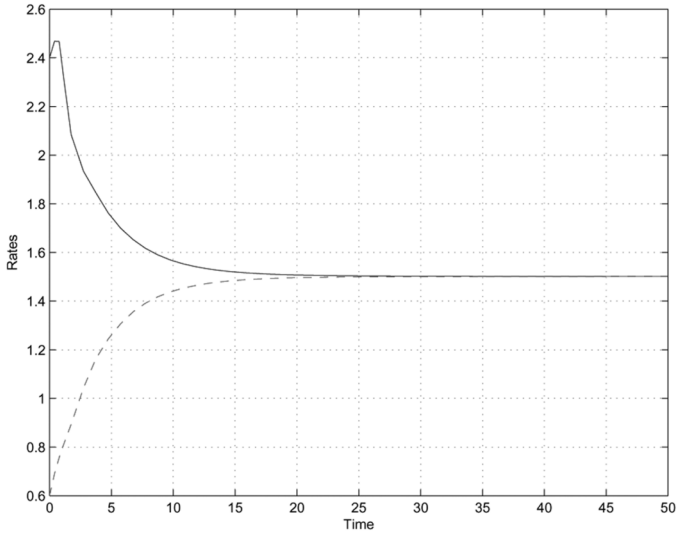


Fig. 3. Example. The evolution of the user rates as a function of time. The solid line corresponds to User 1 and the dashed line corresponds to User 2.

V. Simulation

In Section III, we derived a condition for the stability of the multi-path routing scheme. Here we show through simulation that when the conditions are met, the network does indeed reach a stable equilibrium. In other words, even though the system of congestion controllers is nonlinear, the linear analysis seems to provide a reasonable guideline for choosing the congestion control parameters. Our simulations are at a fluid level, where we ignore the fine packet-level detail. Extensive prior experience has shown that fluid models capture the packet-level model very well. In particular, the stability aspects of the packet model are predicted well by fluid models [3], [4], [8], [19], [20].

Consider the network shown in Fig. 2. Let the link capacity be 2 Mb/s on links L_1, L_2 and L_5, and 1 Mb/s on L_3 and L_4. The rest of the links (the access and egress) are assumed to be very high capacity links. The one-way propagation delay was taken to be 0.1, 0.2, 0.2, 0.2 and 0 seconds on Links L_1, L_2, L_3, L_4, and L_5, respectively. Let the path sets available to the two classes of users be as indicated in the figure. The marking function is taken to be $f_l(y) = \left(\frac{y}{C_l}\right)^\beta$, where C_l is the capacity of Link l and β is taken to be 6. This marking function has been extensively used in [23].

1) Experiment 1: The objective of this experiment is to verify that the linearized stability condition implies global stability for a wide range of initial conditions. We let two persistent users (one of each class) S_1 and S_2 access this network. This experiment was carried out using Simulink, with the initial conditions being swept from 1.5 (the equilibrium value) to 3.0 (the maximum possible value) for User 1, while User 2's initial conditions were the complement (1.5 to 0). We define convergence time to be the time required for both users to reach 95% of the final value. Using the linear stability condition, we chose the following values for the controller gains:

$$\kappa_{11} = 0.098, \quad \kappa_{12} = 0.098, \quad \kappa_{21} = 0.218, \quad \kappa_{22} = 0.218.$$

We provide an example of how the rates for each user look for one of the initial conditions in Fig. 3, where the total user rates

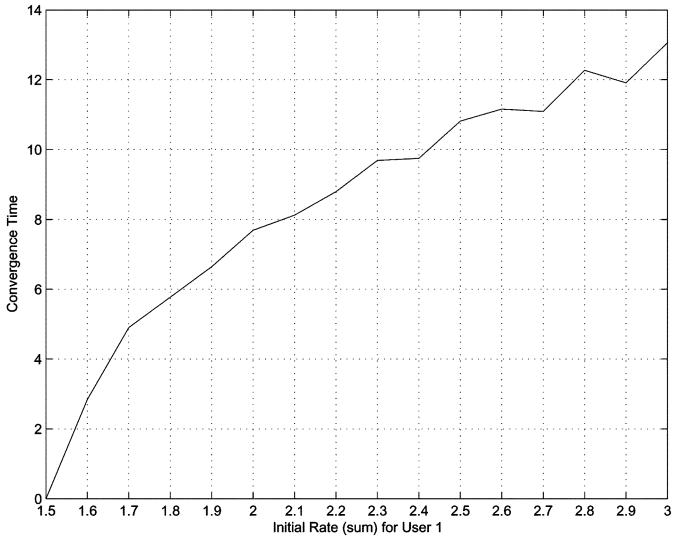


Fig. 4. Experiment 1. The convergence time for the system of two users plotted as a function of the initial conditions of User 1. The convergence time is finite for all initial conditions, implying that the linearized stability condition is sufficient.

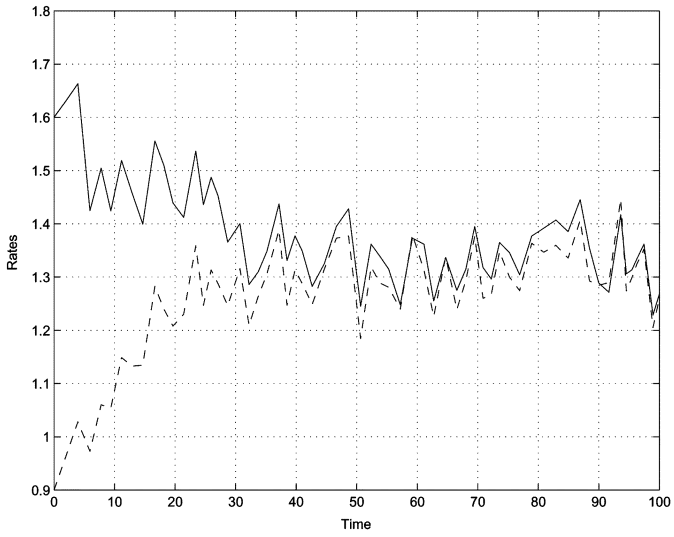


Fig. 5. Experiment 2. The evolution of the users' rates as a function of time in the presence of short-lived flows.

converge to their equilibrium values for this choice of controller gains.

Fig. 4 shows the convergence time for different initial conditions. The maximum possible distance from the equilibrium of 1.5 is when User 1 gets the whole capacity of 3.0, while User 2 gets nothing, while the minimum possible is when both users start off at the equilibrium condition of 1.5. The convergence time in all cases is finite, thus verifying our analytical results.

2) Experiment 2: We now introduce noise sources on links L_2 and L_4 to simulate short-lived flows occupying a maximum of 25% of the link capacity with a mean equal to 12.5% of the link capacity. This simulation is also carried out using Simulink at a fluid level. The simulation results are shown in Fig. 5. As one expects, due to the randomness in the network, the rates of the two long-lived flows do not converge. However, the rates do not grow unbounded and they seem to oscillate around a mean rate

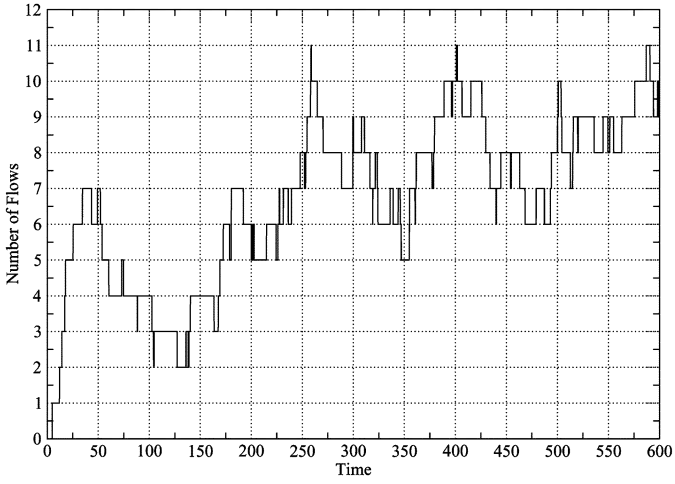


Fig. 6. Experiment 3. The number of file transfers in progress, as a function of time.

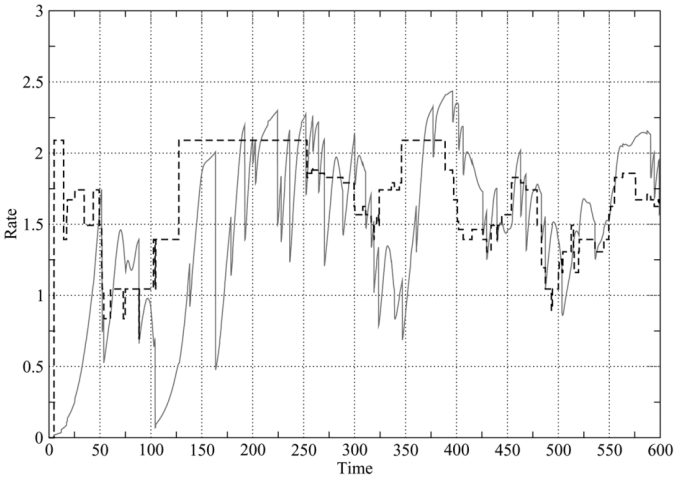


Fig. 7. Experiment 3. The total rate achieved by file type 1 (solid line) and the theoretical rate (dotted).

of about 1.3, which is what is expected when the link capacities are reduced by 12.5%. In this sense, the system is stable even in the presence of stochastic disturbances.

3) Experiment 3: In this experiment, we simulate the network with file arrivals and departures, satisfying the condition of Theorem 6. This simulation was carried out in C, again with the fluid approximation being used. In Fig. 2, we now think of S_1 as files of Type 1 and S_2 as files of Type 2. For each of the two file types, we simulate Poisson arrival processes, with exponentially distributed file sizes. The arrival rates were taken to be 0.1 and 0.035 files/second for S_1 and S_2, respectively, and the mean file size for both file types was taken to be 20 Mbits. Since there are arrivals and departures in the network, the number of files of each type can vary from zero to any arbitrarily large number, but due to the fact that the arrival rates and mean file sizes are chosen to satisfy the stability condition, the probability of a large number of simultaneous file transfers should be very small.

In Fig. 6, we show the total number of file transfers taking place in the network as a function of time. As the figure shows, the number of file transfers remains bounded, thus verifying connection-level stability. In Figs. 7 and 8, we plot the total flow

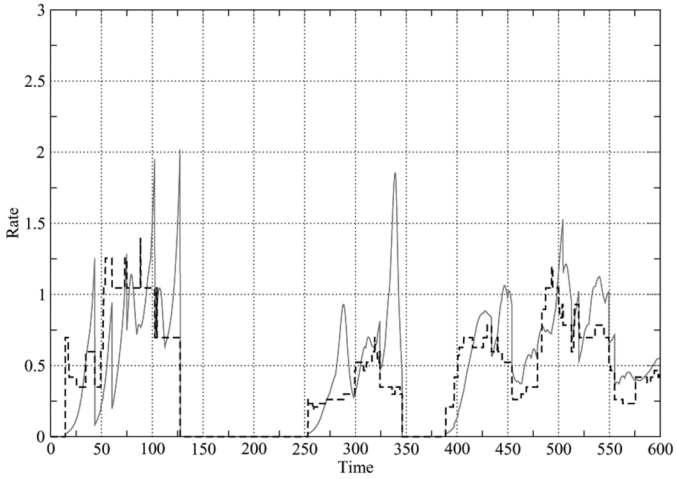


Fig. 8. Experiment 3. The total rate achieved by file type 2 (solid line) and the theoretical rate (dotted).

rate achieved by each file type and compare it to its theoretical value.

The theoretical values are computed by maximizing the net utility in (1) and using the number of files of each type in the network at each time instant. Notice that, whenever there is a new arrival or departure in the system (seen from Fig. 6), there is transient period over which the theoretical rate and the achieved rate for each file type are different. However, given sufficient time (roughly 10–15 RTTs), the achieved rates converge to the theoretical rate. One can draw the conclusion from the simulation that the throughput achieved by the network matches the theoretical throughput reasonably well, even though the calculation assumes that the congestion controllers reach equilibrium instantaneously.

VI. Conclusion

In this paper, we have studied congestion control algorithms for networks where multi-path routing is possible using an overlay network of routers, multi-homing or some other means. Our main contributions are:

- We showed that the decentralized control framework of the single path case extends to the multi-path case in a natural manner, and proved a sufficient linearized stability condition for such systems.
- The condition obtained in [2] states that the total average load through each link must not exceed the capacity of the link for the system to remain stable at the connection time scale. We showed that an appropriately modified version of this result remains valid in the multi-path case as well.
- We showed through simulation that the linear stability conditions seem to be sufficient for global stability. We also verified through simulation that the time-scale separation assumption between connection-level dynamics and congestion-level dynamics is reasonable.

Appendix

The proof of condition (8) is provided below. Define $D \triangleq \text{diag}\{d_i\} \triangleq (KXQ - \varepsilon K)$. The condition (7) $\varepsilon < \hat{x}_r \hat{q}_r$ implies that $d_i > 0$. Then we have

$$\begin{aligned}\lambda(L_{\text{int}}(j\omega)) &= \lambda\left(\varepsilon(j\omega I + D(I+J))^{-1} K X^{-1} X_\Sigma E(j\omega)\right) \\ &= \lambda\left(\varepsilon D^{\frac{1}{2}}\left(j\omega I + D^{\frac{1}{2}}(I+J)D^{\frac{1}{2}}\right)^{-1}\right. \\ &\qquad \left.\times D^{-\frac{1}{2}} K X^{-1} X_\Sigma E(j\omega)\right) \\ &= \lambda\left(\varepsilon\left(j\omega I + D^{\frac{1}{2}}(I+J)D^{\frac{1}{2}}\right)^{-1}\right. \\ &\qquad \left.\times D^{-\frac{1}{2}} K X^{-1} X_\Sigma E(j\omega) D^{\frac{1}{2}}\right)\end{aligned}$$

since eigenvalues do not change under matrix commutation. Because K, X, X_Σ, $E(j\omega)$ are diagonal matrices, the above expression is the same as

$$= \lambda\left(\left(j\omega I + D^{\frac{1}{2}}(I+J)D^{\frac{1}{2}}\right)^{-1} \varepsilon K X^{-1} X_\Sigma E(j\omega)\right).$$

Because

$$\lambda(A^{-1}B) \subset \frac{\mathcal{F}(B)}{\mathcal{F}(A)}$$

if $0 \notin \mathcal{F}(A)$ and B is positive semi-definite, we have

$$\lambda(L_{\text{int}}(j\omega)) \subset \frac{\mathcal{F}(\varepsilon K X^{-1} X_\Sigma E(j\omega))}{\mathcal{F}(j\omega I + D^{\frac{1}{2}}(I+J)D^{\frac{1}{2}})}.$$

Because $D^{\frac{1}{2}}(I+J)D^{\frac{1}{2}}$ is positive semi-definite, its eigenvalues are $\{0, \ldots, 0, \sum_{i\in\mathcal{R}(n)} d_i\}$, where $\mathcal{R}(n)$ is the path set for user n. So

$$\lambda(L_{\text{int}}(j\omega)) \subset \frac{Co\left\{\varepsilon\kappa_r \hat{x}_r^{-1} \sum_{m\in s(r)} \hat{x}_m e^{-j\omega\tau_r}\right\}}{Co\left\{j\omega, j\omega + \sum_{i\in\mathcal{R}(n)} d_i\right\}}.$$

Therefore, the point of intersection of the eigen-locus $\lambda(L_{\text{int}}(j\omega))$ and real axis is to the right of

$$\frac{\varepsilon\kappa_r \hat{x}_r^{-1} \sum_{m\in s(r)} \hat{x}_m e^{-j\omega\tau_r}}{j\omega}.$$

Thus,

$$\varepsilon\kappa_r \hat{x}_r^{-1} \tau_r \sum_{m\in s(r)} \hat{x}_m < \frac{\pi}{2}$$

guarantees that $\lambda(L_{\text{int}}(j\omega))$ is to the right of $-1 + j0$. Hence, the system is stable by the generalized Nyquist criterion. □

Lemma 2: Let $e^{j\omega\tau_{ni}}$ denote the ith routing delay element in block E_n. Then, given constant $\hat{\tau}_n > 0$,

$$\begin{aligned}&\mathcal{F}\left(\frac{1}{j\omega\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right)}\left(\left(j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n\right) I\right.\right. \\ &\qquad \left.\left. - \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n a_n^T\right) E_n\right) \\ &\subset Co\left\{\frac{e^{j\omega\hat{\tau}_n}}{j\omega}, \frac{e^{j\omega\hat{\tau}_n}}{j\omega + \frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2} a_n^T a_n}\right\} + \text{disk}\left(\max_i |\tau_{ni} - \hat{\tau}_n|\right).\end{aligned}$$

Proof: There exists a unitary matrix U satisfying

$$a_n a_n^T = \text{Udiag}\left\{0,\ldots,0,a_n^T a_n\right\} U^T$$

so that

$$\begin{aligned}
&\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)}\left(\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)I-\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n a_n^T\right)\\
&=\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)}\left(\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)I\right.\\
&\quad\left.-\text{Udiag}\left\{0,\ldots,0,\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right\}U^T\right)\\
&=\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)}\text{Udiag}\left\{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n,\ldots,j\omega\right.\\
&\quad\left.+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n,j\omega\right\}U^T\\
&=\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T.
\end{aligned}$$

So,

$$\begin{aligned}
&\mathcal{F}\left(\frac{1}{j\omega\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)}\left(\left(j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n\right)I\right.\right.\\
&\qquad\left.\left.-\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n a_n^T\right)E_n\right)\\
&\quad=\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T E_n\right).
\end{aligned}$$

Given $\hat{\tau}_n > 0$,

$$\begin{aligned}
&\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T E_n\right)\\
&\quad=\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T\right.\\
&\qquad\left.\times\left(E_n-e^{j\omega\hat{\tau}_n}I+e^{j\omega\hat{\tau}_n}I\right)\right)\\
&\quad\subset\mathcal{F}\left(\text{Udiag}\left\{\frac{e^{j\omega\hat{\tau}_n}}{j\omega},\ldots,\frac{e^{j\omega\hat{\tau}_n}}{j\omega},\frac{e^{j\omega\hat{\tau}_n}}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T\right)\\
&\qquad+\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T\right.\\
&\qquad\left.\times\left(E_n-e^{j\omega\hat{\tau}_n}I\right)\right)\\
&\quad=Co\left\{\frac{e^{j\omega\hat{\tau}_n}}{j\omega},\frac{e^{j\omega\hat{\tau}_n}}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}\\
&\qquad+\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T\right.\\
&\qquad\left.\times\left(E_n-e^{j\omega\hat{\tau}_n}I\right)\right).
\end{aligned}$$

From Lemma 1

$$\begin{aligned}
\left\|E_n-e^{j\omega\hat{\tau}_n}I\right\|_2&=\max_i\left\{\left|e^{j\omega\tau_{ni}}-e^{-j\omega\hat{\tau}_n}\right|\right\}\\
&\le\omega\max_i|\tau_{ni}-\hat{\tau}_n|.
\end{aligned}$$

Then, for $x^H x=1$,

$$\begin{aligned}
&\left|x^H\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}U^T\right.\right.\\
&\qquad\left.\left.\times\left(E_n-e^{j\omega\hat{\tau}_n}I\right)\right)x\right|\\
&\quad\le\left\|\text{diag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}\right\|_2\\
&\qquad\times\left\|E_n-e^{j\omega\hat{\tau}_n}I\right\|_2\\
&\quad\le\max_i|\tau_{ni}-\hat{\tau}_n|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathcal{F}\left(\text{Udiag}\left\{\frac{1}{j\omega},\ldots,\frac{1}{j\omega},\frac{1}{j\omega+\frac{\bar{w}_n}{\hat{x}_{\Sigma n}^2}a_n^T a_n}\right\}\right.\\
&\qquad\left.\times U^T\left(E_n-e^{j\omega\hat{\tau}_n}I\right)\right)\subset\text{disk}\left(\max_i|\tau_{ni}-\hat{\tau}_n|\right).
\end{aligned}$$

□

REFERENCES

[1] A. Akella, J. Pang, B. Maggs, S. Seshan, and A. Shaikh, "A comparison of overlay routing and multihoming route control," in *Proc. ACM SIGCOMM*, Portland, OR, Feb. 2004.

[2] T. Bonald and L. Massoulie, "Impact of fairness on Internet performance," in *Proc. ACM SIGMETRICS*, 2001.

[3] T. Bu and D. Towsley, "Fixed point approximation for TCP behavior in an AQM network," in *Proc. ACM SIGMETRICS*, 2001.

[4] S. Deb, S. Shakkottai, and R. Srikant, Stability and convergence of TCP-like congestion controllers in a many-flows regime University of Illinois, Urbana, Tech. Rep., 2002, a shorter version appeared in the *Proc. IEEE INFOCOM 2003*.

[5] S. Deb and R. Srikant, "Global stability of congestion controllers for the Internet," *IEEE Trans. Autom. Contr.*, vol. 48, no. 6, pp. 1055–1060, Jun. 2003.

[6] ——, "Rate-based versus queue-based models of congestion control," *Proc. ACM SIGMETRICS*, 2004.

[7] C. A. Desoer and Y. T. Wang, "On the generalized Nyquist stability criterion," *IEEE Trans. Autom. Contr.*, vol. 25, no. 2, pp. 187–196, Apr. 1980.

[8] C. V. Hollot, V. Misra, D. Towsley, and W. Gong, "On designing improved controllers for AQM routers supporting TCP flows," in *Proc. IEEE INFOCOM*, Anchorage, AK, Apr. 2001, pp. 1726–1734.

[9] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[10] F. P. Kelly, "Models for a self-managed Internet," *Philos. Trans. Royal Soc.*, vol. A358, pp. 2335–2348, 2000.

[11] ——, "Mathematical modelling of the Internet," in *Mathematics Unlimited—2001 and Beyond*B. Engquist and W. Schmid, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 685–702.

[12] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[13] S. Kunniyur and R. Srikant, "A time-scale decomposition approach to adaptive ECN marking," *IEEE Trans. Autom. Contr.*, vol. 47, no. 6, pp. 882–894, Jun. 2002.

[14] ——, "Stable, scalable, fair congestion control and AQM schemes that achieve high utilization in the Internet," *IEEE Trans. Autom. Contr.*, vol. 48, no. 11, pp. 2024–2029, Nov. 2003.

[15] ——, "Analysis and design of an adaptive virtual queue algorithm for active queue management," *IEEE Trans. Netw.*, vol. 12, no. 2, pp. 286–299, Apr. 2004.

[16] X. Lin and N. B. Shroff, "The multipath utility maximization problem," in *Proc. 41st Allerton Conf. Communications, Control and Computing*, Oct. 2003.
[17] S. H. Low, "Optimization flow control with on-line measurement or multiple paths," in *Proc. 16th Int. Teletraffic Congress*, Edinburgh, U.K., 1999.
[18] S. H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Control Syst. Mag.*, vol. 22, no. 1, pp. 28–43, Feb. 2002.
[19] V. Misra, W. Gong, and D. Towsley, "A fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED," in *Proc. ACM SIGCOMM*, Stockholm, Sweden, Sep. 2000.
[20] S. Shakkottai and R. Srikant, "Mean FDE models for Internet congestion control," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1050–1072, Jun. 2001, a shorter version appeared in the *Proc. IEEE INFOCOM 2002* under the title "How good are fluid models of Internet congestion control?".
[21] R. Srikant, *The Mathematics of Internet Congestion Control*. Berlin, Germany: Birkhauser, 2003.
[22] G. Vinnicombe, On the stability of end-to-end congestion control for the Internet Univ. Cambridge, Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR.398, 2001 [Online]. Available: http://www.eng.cam.ac.uk/~gv
[23] ——, "On the stability of networks operating TCP-like congestion control," in *Proc. IFAC World Congress*, Barcelona, Spain, 2002.
[24] W.-H. Wang, M. Palaniswami, and S. H. Low, "Optimal flow control and routing in multi-path networks," *Performance Evaluation*, vol. 52, pp. 119–132, Apr. 2003.

Huaizhong Han received the B.S. and M.S. degrees from the University of Science and Technology of China in 1998 and 2001, respectively. Since Fall 2001, he has been with the University of Massachusetts, Amherst, where he is a Research Assistant working toward the Ph.D. degree in electrical and computer engineering.

His research interests include modeling and control of computer networks and nonlinear systems.

Srinivas Shakkottai (S'00) received the B.Eng. degree in electronics and communication engineering from Bangalore University, India, in 2001, and the M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 2003. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign.

His research interests include the design and analysis of peer-to-peer systems, pricing approaches to resource allocation in fixed and wireless networks, game theory, congestion control in the Internet and measurement and analysis of Internet data.

C.V. Hollot (S'79–M'82–F'04) received the Ph.D. degree in electrical engineering from the University of Rochester, Rochester, NY, in 1984.

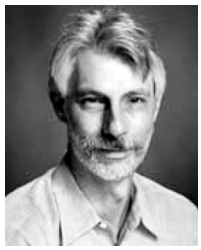
Since 1984, he has been with the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst. He has served as Associate Editor for several control journals. His research interests are in the theory and application of feedback control.

Dr. Hollot received the National Science Foundation Presidential Young Investigator Award in 1988.

R. Srikant (S'90–M'91–SM'01–F'06) received the B.Tech. degree from the Indian Institute of Technology, Madras, in 1985, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1988 and 1991, respectively, all in electrical engineering.

He was a Member of Technical Staff at AT&T Bell Laboratories from 1991 to 1995. He is currently with the University of Illinois, where he is a Professor in the Department of Electrical and Computer Engineering and a Research Professor in the Coordinated Science Laboratory. His research interests include communication networks, stochastic processes, queueing theory, information theory, and game theory.

Dr. Srikant was an associate editor of *Automatica*, and is currently an associate editor of the IEEE/ACM TRANSACTIONS ON NETWORKING and IEEE TRANS. AUTOM. CONTR.. He is also on the editorial boards of special issues of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and IEEE TRANSACTIONS ON INFORMATION THEORY. He was the chair of the 2002 IEEE Computer Communications Workshop in Santa Fe, NM, and will be a program co-chair of IEEE INFOCOM 2007.



Don Towsley (M'78–SM'93–F'95) received the B.A. degree in physics and the Ph.D. degree in computer science from the University of Texas, Austin, in 1971 and 1975, respectively.

From 1976 to 1985, he was a Member of the Faculty of the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, where he is currently a Distinguished Professor in the Department of Computer Science. He has held visiting positions at IBM T. J. Watson Research Center, Yorktown Heights, NY (1982–1983), Laboratoire MASI, Paris, France (1989–1990), INRIA, Sophia Antipolis, France (1996), and AT&T Labs Research, Florham Park, NJ (1997). His research interests include networks, multimedia systems, and performance evaluation.

Dr. Towsley has served on the editorial boards of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE/ACM TRANSACTIONS ON NETWORKING, and is currently the Editor-in-Chief of the IEEE/ACM TRANSACTIONS ON NETWORKING. He currently serves on the editorial boards of *Performance Evaluation* and *Journal of the ACM*. He received the 1998 IEEE Communications Society William Bennet Paper Award and three Best Conference Paper Awards from ACM SIGMETRICS. He was a Program Co-Chair of the joint ACM SIGMETRICS and PERFORMANCE 1992 Conference. He is a member of ORSA and Chair of IFIP Working Group 7.3. He is a Fellow of the ACM.