Stability of end-to-end algorithms for joint routing and rate control

Frank Kelly
Statistical Laboratory
University of Cambridge
Cambridge, U.K.
f.p.kelly@statslab.cam.ac.uk

Thomas Voice
Statistical Laboratory
University of Cambridge
Cambridge, U.K.
t.d.voice@statslab.cam.ac.uk

ABSTRACT

Dynamic multi-path routing has the potential to improve the reliability and performance of a communication network, but carries a risk. Routing needs to respond quickly to achieve the potential benefits, but not so quickly that the network is destabilized. This paper studies how rapidly routing can respond, without compromising stability.

We present a sufficient condition for the local stability of end-to-end algorithms for joint routing and rate control. The network model considered allows an arbitrary interconnection of sources and resources, and heterogeneous propagation delays. The sufficient condition we present is decentralized: the responsiveness of each route is restricted by the round-trip time of that route alone, and not by the round-trip times of other routes. Our results suggest that stable, scalable load-sharing across paths, based on end-to-end measurements, can be achieved on the same rapid time-scale as rate control, namely the time-scale of round-trip times.

Categories and Subject Descriptors

C.2.2 [**Computer-Communication Networks**]: Network Protocols—*congestion control, routing protocols*

General Terms

Algorithms, Theory

Keywords

Internet, dynamic routing, scalable TCP

1. INTRODUCTION

Historically, the primary purpose of IP routing has been to maintain connectivity in the presence of topology changes and network failures. IP routing typically chooses the shortest path to the destination, based on simple metrics like hop count or distance. While the simplicity of this approach has made IP routing highly scalable, there has long been a desire to improve the reliability and performance of the Internet through the use of routing metrics that are more sensitive to congestion [28]. More recently there has also been increasing interest in multi-path routing, motivated by applications to ad-hoc networks [4, 11] and overlay TCP [7], and by perceived problems with current routing protocols [21, 29]. But despite the potential advantages of dynamic routing, it has in the past been difficult to deploy in packet-based networks like the Internet because of potential instability, manifested as routing oscillations [26].

In recent years theoreticians have developed a framework that allows a congestion control algorithm such as Jacobson's TCP [9] to be interpreted as a distributed mechanism solving a global optimization problem: for reviews see [12, 19, 20, 22]. The framework is based on fluid-flow models, and the form of the optimization problem makes explicit the equilibrium resource allocation policy of the algorithm, which can often be restated in terms of a fairness criterion. And the dynamics of the fluid-flow models allow the machinery of control theory to be used to study stability, and to develop rate control algorithms that scale to arbitrary capacities [14, 19, 22].

Han *et al.* [7] have used the framework to study multi-path routing in the Internet. They have presented an algorithm that can be implemented at sources to optimally split the flow between each source-destination pair, and they develop a sufficient condition for the local stability of the algorithm. The condition is decentralized in the sense that the gain parameters for the routes serving a particular source-destination pair are restricted by the round-trip times of those routes, and not by round-trip times elsewhere in the network.

In this paper we improve on this result, and present an algorithm with a sufficient condition for local stability that is decentralized in the stronger sense that the gain parameter for each route is restricted by the round-trip time of that route, but not by the round-trip times of other routes, even those other routes serving the same source-destination pair. The novel feature of our scheme is that the control exerted by a source over its available route flow rates is treated similarly to link congestion feedback. This allows us to apply established single-path techniques to our multi-path model. The condition we derive is conceptually simpler than that of [7], and is less demanding in the case where the round-trip times associated with a source-destination pair are highly heterogeneous. The sufficient condition is a generalization of Vinnicombe's [23] original condition for the single path case, and depends explicitly on the fairness criterion implemented by the algorithm. The sufficient condition constrains the speed of routing adaptation to essentially the same time-scale as is allowed for rate control.

In this paper we are modelling networks with *path diversity*, i.e. systems where at least some source-destination pairs have access to two or more different routes. At the minimum, we assume that, for these source-destination pairs, the source can send different flows addressed to the same destination over different routes, for example, through different Internet service providers, or different initial wireless links. We note that explicit support for edge routing is not currently available in the Internet, but it is enough for our purpose that by some means or other it is possible to create some path diversity.

A helpful distinction, developed by Zhu *et al.* [29], is to view routing information as separated into its structural and dynamic components, the former being concerned with the existence of links, and the latter with the quality of paths across the network. We suppose that the routes available, however discovered, are fixed on the time-scale we are considering. Our focus in this paper is the stability or otherwise of the system's response to dynamic information.

More formally, we suppose the network comprises an interconnection of a set of *sources*, S, with a set of *resources*, J. Each source $s \in S$ identifies a unique source-destination pair. Associated with each source is a collection of *routes*, each route being a set of resources. If a source s transmits along a route r, then we write $r \in s$. Likewise, if a route r uses a resource j we write $j \in r$. For a route r we let $s(r)$ be the (unique) source such that $r \in s(r)$. We let R denote the set of all routes.

We make no assumptions about whether the routes $r \in s$ are disjoint. Clearly the ability to generate resource disjoint routes will assist in the construction of highly robust end-to-end communication for the source-destination pair labelled by s, but our model also covers the case where some or all of the routes $r \in s$ share some path segments.

In our model a route r has associated with it a flow rate $x_r(t) \geq 0$, which represents a dynamic fluid approximation to the rate at which the source $s(r)$ is sending packets along route r at time t.

For each route r and resource $j \in r$, let T_{rj} denote the propagation delay from $s(r)$ to j, *i.e.* the length of time it takes for a packet to travel from source $s(r)$ to resource j along route r. Let T_{jr} denote the propagation delay from j to $s(r)$, *i.e.* the length of time it takes for congestion control feedback to reach $s(r)$ from resource j along route r. In the protocols we shall be considering, a packet must reach its destination before an acknowledgement containing congestion feedback is returned to its source. Further, we assume queueing delays are negligible. Thus for all $j \in r$, $T_{rj} + T_{jr} = T_r$, the round trip time for route r.

Use the notation $a = (b)_c^+$ to mean $a = b$ if $c > 0$ and $a = \max(0, b)$ if $c = 0$.

We are now ready to introduce our fluid-flow model of joint routing and rate control:

$$\frac{d}{dt} x_r(t) = \kappa_r x_r(t) \left(1 - \frac{\lambda_r(t)}{U'_{s(r)}(y_{s(r)}(t))} \right)^+_{x_r(t)} \tag{1}$$

where

$$\lambda_r(t) = \sum_{j \in r} \mu_j(t - T_{jr}), \tag{2}$$

$$\mu_j(t) = p_j(z_j(t)), \quad z_j(t) = \sum_{r: j \in r} x_r(t - T_{rj}) \tag{3}$$

and

$$y_s(t) = \sum_{r \in s} x_r(t - T_r). \tag{4}$$

Here and throughout we assume that, unless otherwise specified, j ranges over the set J, r ranges over the set R, and s ranges over the set S.

We motivate (1-4) as follows. The flow through resource j at time t, $z_j(t)$, comes from routes r that pass through resource j; and the flow that resource j sees at time t on route r left its source a time T_{rj} earlier. If we suppose that resource j adds a price $p_j(z_j)$ onto packets when the total flow through resource j is z_j, then we obtain (3). The total price accumulated by a single packet on route r, and returned to the source $s(r)$ via an acknowledgement received at time t, is given by (2). Finally (1) corresponds to a rate control algorithm for the flow on route r that comprises two components: a steady increase at rate proportional to $\kappa_r x_r(t)$; and a steady decrease at a rate depending upon both the price signals arriving back from route r, and the total rate of acknowledgements $y_{s(r)}(t)$ over all routes serving the source for route r. We shall see that the functions $U_s, s \in S$, appearing in (1) determine how resources are shared. And later, in Section 3, we shall reinterpret p_j as the drop or mark probability at resource j rather than a price.

Although $y_s(t)$ can be interpreted as the rate of acknowledgements received by source s, an alternative, and possibly more practical, implementation of (4) would be that each packet sent along a route $r \in s$ is marked with the flow rate or window size for r at time of sending. The source then computes $y_s(t)$, according to (4), from the values recorded in returning acknowledgements. Later we shall consider an alternative scheme, where each packet sent by s is marked with the total flow rate over all $r \in s$, and where, when an acknowledgement packet is returned, x_r is updated according to

$$\frac{d}{dt} x_r(t) = \kappa_r x_r(t) \left(1 - \frac{\lambda_r(t)}{U'_{s(r)}\left(\sum_{a \in s(r)} x_a(t - T_r)\right)} \right)^+_{x_r(t)} ; \tag{5}$$

here the sum $\sum_{a \in s(r)} x_a(t - T_r)$ is just the total flow rate recorded in a returning acknowledgement. We shall see that this alternative scheme has similar stability properties as those we prove for (1-4).

Under mild assumptions we shall establish that the system (1-4) is locally stable about an equilibrium point, provided the gain parameter κ_r on each route $r \in R$ satisfies a simple sufficient condition.

As an example of the results in Section 2, suppose that

$$U_s(y_s) = \frac{w_s y_s^{1-\alpha}}{1-\alpha},$$

so that the resource shares obtained by different sources are weighted α-fair [18]. When $w_s = 1, s \in S$, the cases $\alpha \to 0$, $\alpha \to 1$ and $\alpha \to \infty$ correspond respectively to an allocation which achieves maximum throughput, is *proportionally fair* or is *max-min fair* [18, 22]. TCP fairness, in the case where each source has just a single route, corresponds to the choice $\alpha = 2$ with w_s the reciprocal of the square of the (single) round trip time for source s [17, 22].

Further suppose that

$$p_j(z_j) = \left(\frac{z_j}{C_j}\right)^{\beta}, \tag{6}$$

for constant C_j representing the capacity of resource j. Then the sufficient condition for local stability that we obtain is satisfied if, for each $r \in R$,

$$\kappa_r T_r(\alpha + \beta) < \frac{\pi}{2}. \tag{7}$$

Thus we have a sufficient condition for local stability that restricts the gain parameter κ_r on route r by the round-trip time T_r of route r, but not by other round-trip times, even those of other routes serving the same source. The condition has a straightforward dependence on the fairness criterion, described by the parameter α, as well as the resource responsiveness, described by the parameter β. Notably, the condition does not depend upon the size of the flow rates x or the congestion feedback λ, the number of resources on routes, the number of flows on routes or the network topology.

The organization of the paper is as follows. In Section 2 we present the formal analysis of the above model. In particular, we show that the use of delayed information described in (4) corresponds, under linearization, to the introduction of a fictitious link for each source-destination pair into the single route model, and hence leads to a straightforward sufficient condition for stability. The model analysed in Section 2 allows a fairly general form for the functions U_s, but is, in some respects, oversimplified. Section 3 considers a more specialized model that better approximates a network like the Internet, where congestion is indicated by a dropped or marked packet. We propose a routing extension of scalable TCP [14], a variant of TCP with attractive scaling properties that we show are inherited by our multi-path extension. Section 4 concludes with a brief discussion of the implications of our results for the division of routing functionality between layers of the network architecture.

2. ANALYSIS

We shall show, in Theorem 1, that an equilibrium point of the system (1-4) solves an optimization problem. We shall then discuss the global stability of the system in the case where there are no propagation delays, before proving our main result, Theorem 2, on the local stability of the system with propagation delays.

First we introduce matrices to succinctly express the relationships between sources, routes and resources. Let $A_{jr} = 1$ if $j \in r$, so that resource j lies on route r, and set $A_{jr} = 0$ otherwise. This defines a 0–1 matrix $A = (A_{jr}, j \in J, r \in R)$. Set $H_{sr} = 1$ if $r \in s$, so that route r serves source s, and set $H_{sr} = 0$ otherwise. This defines a 0–1 matrix $H = (H_{sr}, s \in S, r \in R)$.

Next we describe our regularity assumptions. Assume that the function $U_s(y_s)$, $y_s \geq 0$, is a increasing function of y_s, twice continuously differentiable, with $U_s'(y_s) \to 0$ as $y_s \uparrow \infty$ and $U_s''(y_s) > 0$ for $y_s > 0$. Thus $U_s(\cdot)$ is strictly concave. Assume that the function $p_j(z_j)$, $z_j \geq 0$, is a non-negative function of z_j, continuously differentiable with $p_j'(z_j) > 0$ for $z_j > 0$. Let $C_j(z_j)$ be defined by

$$C_j(z_j) = \int_0^{z_j} p_j(u)\, du.$$

From our assumptions on $p_j(\cdot)$, the function $C_j(\cdot)$ is strictly convex. The function $C_j(\cdot)$ will in general be parameterised by the capacity C_j of resource j, as for example if $p_j(\cdot)$ is given by the form (6).

THEOREM 1. *If the vector $x = (x_r, r \in R)$ solves the optimization problem*

$$\begin{array}{ll} \textit{maximize} & \sum_{s\in S} U_s(y_s) - \sum_{j \in J} C_j(z_j) \\ \textit{where} & y = Hx \quad z = Ax \\ \textit{over} & x \geq 0, \end{array} \tag{8}$$

then x is an equilibrium point of the system (1-4).

PROOF. The objective function of the optimization problem (8) is differentiable, and so it is maximized at $(x_r, r \in R)$ if and only if, for each $r \in R$,

$$x_r \geq 0, \quad U'_{s(r)}\left(\sum_{r:r\in s} x_r\right) - \sum_{j \in r} p_j\left(\sum_{a:j\in a} x_a\right) \geq 0 \tag{9}$$

and

$$x_r \cdot \left(U'_{s(r)}\left(\sum_{r:r\in s} x_r\right) - \sum_{j\in r} p_j\left(\sum_{a:j\in a} x_a\right)\right) = 0. \tag{10}$$

But condition (10) implies that the derivative (1) is zero, after substituting $x_r(t) = x_r, t \geq 0$, into (2-4). □

Remark 1. The optimization problem (8) has a long history in connection with road transport networks [1, 27], as well as communication networks [2, 6]. By our assumptions on the functions $U_s(\cdot)$ and $C_j(\cdot)$, there exists a solution to the optimization problem. At an optimum $x = (x_r, r \in R)$ is not necessarily unique, but, by the strict concavity of the functions $U_s(\cdot)$ and the strict convexity of the functions $C_j(\cdot)$, the vectors $y = Hx$ and $z = Ax$ are unique. If X is the set of optima x, then X is the intersection of an affine space with the orthant,

$$X = \{x : Hx = y, Ax = z\} \cap \{x : x \geq 0\},$$

and is compact.

Remark 2. The set X does not exhaust the equilibria of the system (1-4). For example, $x(t) = 0, t \geq 0$, is also an equilibrium point. The difficulty is that if $x_r(\bar{t}) = 0$ then $x_r(t) = 0$ for $t > \bar{t}$: the trajectory $x(t)$ can thus become trapped in the face $\{x : x_r = 0\}$. Call an equilibrium *spurious* if it is not also an optimum of the optimization problem. To understand better the issue, consider the dynamical system, evolving on $\{x : x_r \geq \epsilon, r \in R\}$, defined by

$$\frac{d}{dt}\, x_r(t) = \kappa_r(x(t)) \left(U'_{s(r)}(y_{s(r)}(t)) - \lambda_r(t)\right)^+_{x_r(t)-\epsilon} \qquad (11)$$

with (2-4), where $T_r = 0, r \in R$, and ϵ is a small positive constant. Assume that $\kappa_r(x)$ is a positive, continuous function bounded away from zero on the set $\{x : x_r \geq \epsilon, r \in R\}$. Let X_ϵ be the set of optima to the amended optimization problem (8), with $\{x \geq 0\}$ replaced by $\{x : x_r \geq \epsilon, r \in R\}$. Following [13], rewrite the objective function of (8) as

$$\mathcal{U}(x) = \sum_{s \in S} U_s\Big(\sum_{r \in s} x_r\Big) - \sum_{j \in J} \int_0^{\sum_{r:j\in r} x_r} p_j(u)du,$$

and, for the dynamical system (11), calculate

$$\frac{d}{dt}\, \mathcal{U}\big(x(t)\big) = \sum_{r \in R} \frac{\partial \mathcal{U}}{\partial x_r} \cdot \frac{d}{dt}\, x_r(t)$$

$$= \sum_{r \in R} \kappa_r(x(t)) \left(\left(U'_{s(r)}(y_{s(r)}(t)) - \lambda_r(t)\right)^2\right)^+_{x_r(t)-\epsilon}.$$

Hence outside any neighbourhood of X_ϵ

$$\frac{d}{dt}\mathcal{U}\big(x(t)\big) > 0$$

and is bounded away from zero. This is enough to ensure that any trajectory of the dynamical system (11) converges to the set X_ϵ. Thus, at least when there are no propagation delays, the amended system (11) can avoid being trapped at faces. The amendment, which prevents $x_r(t)$ dropping below a low level ϵ, can be interpreted as follows: even if a route appears too expensive, a low level of probing should take place, in case the price of the route changes. Any low, but positive, level of probing[1] is sufficient to rule out spurious equilibria, and henceforth we do not explicitly incorporate the parameter ϵ within our model (1-4).

We now turn to our main concern, the local stability of the system (1-4).

Define an equilibrium point x to be *interior* if it satisfies (9-10), and if for each route r either one or other of the inequalities (9) is strict; we thus rule out the possible degeneracy that *both* terms in the product (10) might vanish. At an interior equilibrium point x it is possible for a route r not to be used, i.e. $x_r = 0$, but there then exists a neighbourhood of x such that within this neighbourhood $x_r > 0$ implies $\dot{x}_r < 0$.

We next establish a sufficient condition for the local stability of $(y(t), z(t))$ near any given interior equilibrium point x. Let $y = Hx, z = Ax$, $U''_s = U''_s(y_s)$, $\mu_j = p_j(z_j), p'_j = p'_j(z_j), j \in J$, and $\lambda = A\mu$. Let $T_{max} = \max(T_r, r \in R)$; this parameter is needed to describe an initial condition of the system (1-4), but will not be part of the sufficient condition for local stability.

[1]We note that probing is likely to be important for structural, as well as dynamic, aspects of routing.

THEOREM 2. *Let x be an interior equilibrium point, and suppose that for each $r \in R$,*

$$\frac{\kappa_r T_r}{\lambda_r} \left(-U''_{s(r)} y_{s(r)} + \sum_{j \in r} z_j p'_j\right) < \frac{\pi}{2}. \qquad (12)$$

Then there exists a neighbourhood $\mathcal{N}$ of x such that for any initial trajectory $(x(t), t \in (-T_{max}, 0))$ lying within the neighbourhood $\mathcal{N}$, $(y(t), z(t))$ converge exponentially as $t \to \infty$ to the unique solution (y, z) to the optimization problem (8).

PROOF. Initially assume that $x_r > 0$ for $r \in R$, and thus that

$$U'_{s(r)}\left(\sum_{r:r \in s} x_r\right) = \sum_{j \in r} p_j \left(\sum_{a:j \in a} x_a\right) \qquad (13)$$

for each $r \in R$. Later we shall see that the assumption is without loss of generality.

Let $x_r(t) = x_r + u_r(t)$, $y_s(t) = y_s + v_s(t)$, $z_j(t) = z_j + w_j(t)$. Then, linearizing the system (1-4) about x, and using the relation (13), we obtain the equations

$$\frac{d}{dt}\, u_r(t) = -\frac{\kappa_r x_r}{\lambda_r} \left(-U''_{s(r)}\; v_{s(r)}(t) + \sum_{j \in r} p'_j w_j(t - T_{jr})\right), \qquad (14)$$

$$v_s(t) = \sum_{r:r \in s} u_r(t - T_r), \qquad (15)$$

$$w_j(t) = \sum_{r:j \in r} u_r(t - T_{rj}). \qquad (16)$$

Let us overload notation and write $u_r(\omega), v_s(\omega), w_j(\omega)$ for the Laplace transforms of $u_r(t), v_s(t), w_j(t)$ respectively. We may deduce from (14-16),

$$\omega u_r(\omega) = -\frac{\kappa_r x_r}{\lambda_r} \left(-U''_{s(r)}\; v_s(\omega) + \sum_{j \in r} p'_j e^{-\omega T_{jr}} w_j(\omega)\right),$$

$$v_s(\omega) = \sum_{r:r \in s} e^{-\omega T_r} u_r(\omega),$$

$$w_j(\omega) = \sum_{r:j \in r} e^{-\omega T_{rj}} u_r(\omega).$$

We calculate that

$$\begin{pmatrix} v(\omega) \\ w(\omega) \end{pmatrix} = -P^{-1} R(-\omega)^T X(\omega) R(\omega) P \begin{pmatrix} v(\omega) \\ w(\omega) \end{pmatrix}, \quad (17)$$

where $X(\omega)$ is an $|R| \times |R|$ diagonal matrix with entries $X_{rr}(\omega) = e^{-\omega T_r}/(\omega T_r)$, and P is a $(|S| + |J|) \times (|S| + |J|)$

diagonal matrix with entries $P_{ss} = 1$, $P_{jj} = (p'_j)^{\frac{1}{2}}$, and $R(\omega)$ is a $|R| \times (|S| + |J|)$ matrix where

$$R_{rs}(\omega) = \left(-U''_{s(r)} T_r \frac{\kappa_r x_r}{\lambda_r}\right)^{\frac{1}{2}}, r \in s$$

$$R_{rj}(\omega) = e^{-\omega T_{jr}} \left(\frac{\kappa_r x_r}{\lambda_r} T_r p'_j\right)^{\frac{1}{2}}, \; j \in r$$

and all other entries are 0.

The matrix $G(\omega) = P^{-1} R(-\omega)^T X(\omega) R(\omega) P$, which appears in (17) is called the *return ratio* for (v, w). From the generalized Nyquist stability criterion [5, 8] it is sufficient to prove that the eigenvalues of the return ratio $G(\omega)$ do not encircle the point -1 for $\omega = i\theta$, $-\infty < \theta < \infty$, in order to deduce that $(v(t), w(t)) \to 0$ exponentially as $t \to \infty$. Note, we are not, at this stage, interested in the asymptotic behaviour of $u(t)$.

If λ is an eigenvalue of the return ratio then we can find a unit vector z such that

$$\lambda z = R(i\theta)^\dagger X(i\theta) R(i\theta) z,$$

where $\dagger$ represents the matrix conjugate, and hence

$$\lambda = z^\dagger R(i\theta)^\dagger X(i\theta) R(i\theta) z.$$

If $d = R(i\theta) z$ then, since X is diagonal,

$$\lambda = \sum_r |d_r|^2 X_{rr}(i\theta) = \sum_r |d_r|^2 \frac{e^{-i\theta T_r}}{i\theta T_r}.$$

Hence $\lambda = K\zeta$, where $K = \|R(i\theta)z\|^2$ and ζ lies in the convex hull of

$$\left\{\frac{e^{-i\theta T_r}}{i\theta T_r} : r \in s, s \in S\right\}.$$

The convex hull includes the point $-2/\pi$ on its boundary (at $\theta T_r = \pi/2$), but contains no point on the real axis to the left of $-2/\pi$ [23], and hence if λ is real then $\lambda \geq (-2/\pi)K$.

Next we bound K. Let Q be the $(|S| + |J|) \times (|S| + |J|)$ diagonal matrix taking values $Q_{ss} = y_s \sqrt{-U''_s}$ and $Q_{jj} = z_j \sqrt{p'_j}$, let $\rho(\cdot)$ denote the spectral radius, and $\|\cdot\|_\infty$ the maximum row sum matrix norm. Then

$$\begin{aligned} K &= z^\dagger R(i\theta)^\dagger R(i\theta) z \\ &\leq \rho(R(i\theta)^\dagger R(i\theta)) \\ &= \rho(Q^{-1} R(i\theta)^\dagger R(i\theta) Q) \\ &\leq \|Q^{-1} R(i\theta)^\dagger R(i\theta) Q\|_\infty \\ &< \frac{\pi}{2}, \end{aligned}$$

the last inequality following from (12).

So we have that $\lambda > -1$ for any real eigenvalue λ. Thus, when the loci of the eigenvalues of $G(i\theta)$ for $-\infty < \theta < \infty$ cross the real axis, they do so to the right of -1. Hence the loci of the eigenvalues of $G(i\theta)$ cannot encircle -1, the generalized Nyquist stability criterion is satisfied and the system (14-16) is stable, in the sense that $v_s(t) \to 0$, $w_j(t) \to 0$ exponentially, for all s, j, as $t \to \infty$. There remains the possibility that $x(t)$ might hit a boundary of the positive orthant, and invalidate the linearization (14-16). To rule out this possibility, note that there exists an open neighbourhood of x, say $\mathcal{N}'$, such that whilst $x(t) \in \mathcal{N}'$, the linearization is valid, and so $y_s(t) \to y_s$ and $z_j(t) \to z_j$ exponentially. Now $(\dot{x}_r(t), t > 0)$ is defined by (1-4) as a function of $(y(t), z(t), t > -T_{max})$. Thus, whilst $x(t) \in \mathcal{N}'$, $\dot{x}_r$ decays exponentially to 0 for all r and therefore, the total distance $x_r(t)$ can travel from $x_r(0)$, whilst remaining in $\mathcal{N}'$, is bounded by

$$\gamma \max_{t \in (-T_{max}, 0)} \|(y(t), z(t)) - (y, z)\|$$

for some γ. Hence we can pick an open subset, $\mathcal{N} \subset \mathcal{N}'$ such that if $x(t) \in \mathcal{N}$ for $t \in (-T_{max}, 0)$ then $x(t) \in \mathcal{N}'$ for all t. Furthermore, if $x(t) \in \mathcal{N}'$ for all t then, since $\dot{x}_r$ decays exponentially to 0, $x(t)$ must be Cauchy, and must therefore tend to a limit. Thus $\mathcal{N}$ is as required.

Finally we shall relax the assumption that $x_r > 0$ for all r. Since x is an interior equilibrium point, $x_r = 0$ implies $\dot{x}_r(t) < 0$. Thus there is a neighbourhood of x, say $\mathcal{M}$, such that, on $\mathcal{M}$, the linearization of (1-4) coincides with the case where we discard all r such that $x_r = 0$. Therefore, as above, we may choose an open neighbourhood $\mathcal{N} \subset \mathcal{M}$ such that for any initial trajectory $(x(t), t \in (-T_{max}, 0))$ lying within the neighbourhood $\mathcal{N}$, $(y(t), z(t))$ converge exponentially as $t \to \infty$ to the unique solution (y, z) to the optimization problem (8). $\square$

Remark 1. The linearization (14-16) is similar to that arising in the treatment by Johari and Tan [10], Massoulié [16] and Vinnicombe [23] of the case where each source-destination pair has a single route. Comparing our linearization with theirs, it is as if we transform our model by treating each route as arising from a separate source, and for each $s \in S$ we add a fictitious link $l(s)$ to each of the routes $r \in s$, with $T_{l(s)r} = 0$ and $p_{l(s)}(y_s) = -U'_s(y_s)$. A complication is the non-uniqueness of x, and hence our need to approach the result via the convergence of (y, z).

Remark 2. Theorem 2 remains valid if equation (1) is replaced by (5). This corresponds to the added links discussed in remark 1 having the property that $T_{l(s)r} = T_r$ rather than $T_{l(s)r} = 0$: the flows from source s share a fictitious link as they *leave* the source s, rather than as they *return* to source s.

Remark 3. In [7] a system similar to (1-4), is considered, but where, instead of equation (4), $y_s(t) = \sum_{a \in s} x_a(t)$. The sufficient condition for local stability obtained in [7] restricts κ_r by round-trip times of all routes serving $s(r)$, and can be onerous if the round-trip times of the routes serving a source-destination pair are heterogeneous. We have shown that using delayed information, either $x_a(t - T_a)$ or $x_a(t - T_r)$ rather than $x_a(t)$, allows a fully decentralized sufficient condition.

Remark 4. If

$$U_s(y_s) = \frac{w_s y_s^{1-\alpha_s}}{1 - \alpha_s},$$

and

$$p_j(z_j) = \left(\frac{z_j}{C_j}\right)^{\beta_j}$$

then the condition (12) is satisfied if, for each $r \in R$,

$$\kappa_r T_r(\alpha_{s(r)} + \max(\beta_j, j \in r)) < \frac{\pi}{2}.$$

Observe that the importance of the source parameter α_s, relative to the resource parameters β_j, increases as α_s increases. The condition (7) arises as the special case where $\alpha_s = \alpha, s \in S$, and $\beta_j = \beta, j \in J$.

Remark 5. If we consider a network consisting of one source s, one route $r \in s$ and one link $l \in r$, then the linearization of this system is

$$\dot{u}(t) = -\kappa\,(\alpha + \beta)\,u(t - T). \tag{18}$$

The Nyquist stability criterion, applied to the Laplace transform of this differential equation tells us that our system is locally stable if and only if

$$\kappa T\,(\alpha + \beta) < \frac{\pi}{2}.$$

Indeed if this inequality were replaced by equality, then $u(t) = \sin(\pi t/2T)$ solves equation (18), an oscillatory solution with period $4T$. Thus, for this example, our condition is tight.

3. AN EXTENSION OF SCALABLE TCP

In this Section we consider a refinement of the fluid-flow model used earlier. The refinement is intended to better approximate the behaviour of a network like the Internet, where congestion is indicated by a dropped or marked packet, and hence where a single packet crossing the network generates just a single bit of information concerning congestion along its route. The refinement follows [23] and we use it to develop a routing extension of scalable TCP [14], a variant of TCP with certain attractive scaling properties that we show are inherited by our multi-path extension.

The single bit of information is carried back to the source by the acknowledgement stream. The rate at which acknowledgements from route r arrive back at the source $s(r)$ at time t is $x_r(t-T_r)$, since these acknowledgements arise from packets sent a time T_r previously. Let $\lambda_r(t)$ be the proportion of these acknowledgements that indicate congestion. These acknowledgements arise from packets that passed through resource j a time T_{jr} previously: thus

$$\lambda_r(t) = 1 - \prod_{j \in r}(1 - \mu_j(t - T_{jr})). \tag{19}$$

where $\mu_j(t)$ is the proportion of packets marked at resource j at time t, under the approximation that packet marks at different resources are independent. Observe that (2) approximates (19) when the probabilities $\mu_j, j \in J$, are small. The flow on route r that is seen at resource j at time t left the source for route r a time T_{rj} previously: hence

$$\mu_j(t) = p_j\left(\sum_{r: j \in r} x_r(t - T_{rj})\right), \tag{20}$$

as in (3). Whereas in Section 2 the functions $p_j(\cdot)$ were not necessarily bounded, in this Section we presume they are bounded above by 1, in line with their interpretation here as a drop or mark probability rather than a price.

Suppose the sending rate $x_r(t)$ on route r at time t varies according to following algorithm: on receipt of a positive acknowledgement the sending rate is increased by $\bar{a}/T_r$, and on receipt of a negative acknowledgement indicating congestion the sending rate is decreased by $b_r x_r(t)/T_r$. This corresponds to the stable flow control algorithm of [15]: particular choices for $\bar{a}, b_r$ give scalable TCP [14].

We are now ready to define our routing extension. We suppose that the response on receiving a negative acknowledgement is altered. Now, on receipt of a negative acknowledgement the sending rate is decreased by $b_r y_{s(r)}(t)/T_r$, where $y_s(t)$ is given by equation (4). The fluid-flow model becomes

$$\begin{aligned}\frac{d}{dt}\,x_r(t) &= \frac{x_r(t - T_r)}{T_r} \\ &\quad \cdot \left(\bar{a}\,\left(1 - \lambda_r(t)\right) - b_r\,y_{s(r)}(t)\,\lambda_r(t)\right)^+_{x_r(t)}.\end{aligned} \tag{21}$$

Let x be an equilibrium point. Then

$$\sum_{a \in s(r)} x_a = \frac{\bar{a}}{b_r} \cdot \frac{1 - \lambda_r}{\lambda_r} \tag{22}$$

for any route r with $x_r > 0$. Let $x_r(t) = x_r + u_r(t)$, for those routes with $x_r > 0$. Then linearizing about x and using the relation (22) we obtain

$$\begin{aligned}T_r \frac{d}{dt}\,u_r(t) &= -\bar{a}(1 - \lambda_r) \\ &\quad \cdot \left(\frac{x_r}{\sum_{a \in s(r)} x_a} \sum_{a \in s(r)} u_a(t - T_a) + \frac{x_r}{\lambda_r}\nu_r(t)\right)\end{aligned} \tag{23}$$

where

$$\nu_r(t) = \sum_{j \in r} \frac{p_j'}{1 - p_j} \sum_{a: j \in a} u_a(t - T_{aj} - T_{jr}). \tag{24}$$

The method of Theorem 2 may be applied to this linearized system, and we find that local stability of (23-24) is implied by $\|R(i\theta)^\dagger R(i\theta)\|_\infty < 1$ where

$$R_{r,s}(\omega) = \left(\frac{\bar{a}x_r(1 - \lambda_r)}{y_s}\right)^{\frac{1}{2}} \qquad r \in s$$

$$R_{rj}(\omega) = e^{-\omega T_{jr}}\left(\frac{\bar{a}x_r(1 - \lambda_r)p_j'}{\lambda_r(1 - p_j)}\right)^{\frac{1}{2}} \qquad j \in r$$

and all other entries are 0. For this $R(\omega)$, $\|R(i\theta)^\dagger R(i\theta)\|_\infty$ is less than 1 if

$$\bar{a}\frac{1 - \lambda_r}{\lambda_r}\left(\lambda_r + \sum_{j \in r} \frac{z_j p_j'}{1 - p_j}\right) < \frac{\pi}{2}. \tag{25}$$

Suppose that the functions $p_j(\cdot), j \in J$, are given by equation (6). Then

$$\frac{1 - \lambda_r}{\lambda_r} \sum_{j \in r} \frac{z_j p_j'}{1 - p_j} \leq \beta.$$

Thus, from (25), a sufficient condition for local stability is that

$$\bar{a}\,(1+\beta) < \frac{\pi}{2}. \tag{26}$$

The corresponding condition for local stability of scalable TCP [14, 23] is $\bar{a}\beta < \frac{\pi}{2}$, and so the introduction of routing makes stability only a little harder to ensure.

Remark 1. If $b_r = 1/w_s, r \in s$, then from (22) the marking rate λ_r is the *same* on every route $r \in s$ with $x_r > 0$. Further, the total flow rate serving s is proportional to the weight w_s and approximately inversely proportional to this common value of λ_r, corresponding to a resource allocation across source-destination pairs that is approximately weighted proportionally fair [13]. In particular, if a source-destination pair s has more routes across the network then this may aid the network to balance load, and may help the resilience and reliability achieved from the network by the source-destination pair s; but the weight in the fairness criterion, namely w_s, is unaffected by the number of routes to which s has access. The marking rate on unused routes serving s is at least as high as the common value of λ_r on the routes used, provided some probing mechanism ensures a flow can escape from zero if the second term in the product (21) is positive.

Remark 2. As in Section 2, the same condition is sufficient for local stability if, in equation (21), $y_{s(r)} = \sum_{a\in s(r)} x_a(t-T_a)$ is replaced by $\sum_{a\in s(r)} x_a(t-T_r)$. The first sum, $y_{s(r)}$, is the rate of acknowledgements arriving back at the source $s(r)$ at time t, summed over all routes serving $s(r)$. An implementation might record $x_a(t)$ in a packet leaving on route a at time t, copy the value into the acknowledgement for the packet, and thus return it for use by $s(r)$ a time T_a later. In contrast, $\sum_{a\in s(r)} x_a(t-T_r)$ is the aggregate flow rate leaving the source $s(r)$ at time $t-T_r$. An implementation might record the aggregate rate $\sum_{a\in s(r)} x_a(t)$ in a packet leaving $s(r)$ at time t, copy the value into the acknowledgement for the packet, and thus return it to $s(r)$ via route r a time T_r later.

Remark 3. In the model (19), (20), (21) the equation (20) represents the marking probability at a resource as a function of the instantaneous flow through the resource. Suppose instead the marking probability at a resource is a function of an exponentially weighted average of the flow through the resource. Consider the system (19), (21) where, instead of equation (20),

$$\mu_j(t) = p_j(z_j(t)), \quad \delta_j \frac{d}{dt} z_j(t) = \sum_{a:j\in a} x_a(t-T_{aj}) - z_j(t).$$

Voice [25] establishes a decentralized sufficient condition for the local stability of this system, of the same form as that described in this paper, building upon the results of Vinnicombe [23, 24] for the case without routing.

Remark 4. The function (6) is a natural form to be implemented by active queue management [15, 24]: for example, $z_j(t)$ might be estimated as in the previous remark, and packets marked accordingly. For large buffers operating with drop tail, a more reasonable approximation for the proportion of packets overflowing the buffer is [22]

$$p_j(z_j) = [z_j - C_j]^+/z_j.$$

With this form, the sufficient condition (23) for local stability is satisfied if

$$b_r(\lambda_r + M_r) \sum_{a\in s(r)} x_a < \frac{\pi}{2}$$

where $M_r = \sum_{j\in J} I[p_j > 0]A_{jr}$, the number of saturated resources on route r. This is a much less attractive condition than (26): as well as the dependence on M_r, which may not be known at the edge of the network, its dependence on x prevents it scaling to arbitrary flow rates. The network may be stable for certain capacities and flow rates, but may become unstable with larger capacities and flow rates. In contrast the condition (26) is indeed scalable: as for scalable TCP in the single route case, the flow rates x, the marking rates λ, even the network topology, are notable by their absence from the stability condition.

4. CONCLUSION AND DISCUSSION

In this paper we have used a fluid-flow model to analyse the local stability of an end-to-end algorithm for joint routing and rate control. We have seen that stable, scalable load-sharing across paths, based on end-to-end measurements, can be achieved on the same rapid time-scale as rate control.

In the Internet there is generally a single path from a source to a destination, or a pre-determined split of traffic across a set of paths [21], mirroring the layering within the TCP/IP stack, where rate control is part the transport layer but routing is considered to be part of the network layer. The optimization and control framework of this paper sheds light on an aspect of this layering (cf. [3]), and on the possible separation (cf. [29]) of routing information into slowly varying structural information, able to provide a source-destination pair with a collection of available paths, and dynamic information, determined from end-to-end measurements and used, in our proposal, by a source-destination pair to balance load across paths. Our results suggest that while structural information may be provided by the network layer, load-balancing is more naturally part of the transport layer. In particular, we have observed that, for dynamic routing, the key constraint on the responsiveness of each route is the round-trip time of that route, information which is naturally available at sources.

5. ACKNOWLEDGEMENTS

The authors are grateful to Bob Briscoe, Jon Crowcroft, Christophe Diot, Richard Gibbens, Damon Wischik and two anonymous referees for their comments on an earlier draft of this paper. The research of the authors is supported in part by EPSRC grant GR/S86266/01.

6. REFERENCES

[1] M. Beckmann, C.B. McGuire and C.B. Winsten. *Studies in the Economics of Transportation*. Cowles Commission Monograph, Yale University Press, 1956.

[2] D. Bertsekas and R. Gallager. *Data Networks*, 2nd edition. Prentice-Hall, New Jersey, 1992.

[3] M. Chiang. Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control. *IEEE J. Sel. Areas Comm.*, 23:104–116, 2005.

[4] J. Crowcroft, R. Gibbens, F. Kelly, and S. Östring. Modelling incentives for collaboration in mobile ad hoc networks. *Performance Evaluation*, 57:427–439, 2004.

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

[6] S.J. Golestani. *A Unified Theory of Flow Control and Routing in Data Communication Networks*. PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science, Cambridge, MA, 1980.

[7] H. Han, S. Shakkottai, C. Hollot, R. Srikant, and D. Towsley. Overlay TCP for multi-path routing and congestion control. In *ENS-INRIA ARC-TCP Workshop*, Paris, France, 2003.

[8] O.L.R. Jacobs. *Introduction to Control Theory*. Oxford University Press, Oxford, 1993.

[9] V. Jacobson. Congestion avoidance and control. *Computer Communication Review* 18(4): 314–329, 1988.

[10] R. Johari and D.K.H. Tan. End-to-end congestion control for the Internet: delays and stability. *IEEE/ACM Transactions on Networking*, 9:818–832, 2001.

[11] D.B. Johnson and D.A. Maltz. Dynamic source routing in ad hoc wireless networks. In T. Imielinski and H. Korth, editors, *Mobile Computing*, 153–181. Kluwer, 1996.

[12] F. Kelly. Fairness and stability of end-to-end congestion control. *European Journal of Control*, 9:159–176, 2003.

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

[14] T. Kelly. Scalable TCP: improving performance in highspeed wide area networks. *Computer Communication Review*, 32(2):83–91, 2003.

[15] T. Kelly. *Engineering Flow Controls for the Internet*. PhD thesis, Department of Engineering, University of Cambridge, 2004. http://www-lce.eng.cam.ac.uk/∼ctk21/papers/

[16] L. Massoulié. Stability of distributed congestion control with heterogeneous feedback delays. *IEEE Transactions on Automatic Control*, 47:895–902, 2002.

[17] M. Mathis, J. Semke, J. Mahdavi, and T. Ott. The macroscopic behaviour of the TCP congestion avoidance algorithm. *Computer Communication Review*, 27:67–82, 1997.

[18] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8:556–567, 2000.

[19] F. Paganini, Z. Wang, J.C. Doyle, and S.H. Low. Congestion control for high performance, stability and fairness in general networks. *IEEE/ACM Transactions on Networking*, 2005.

[20] A. Papachristodoulou, L. Li, and J.C. Doyle. Methodological frameworks for largescale network analysis and design. *Computer Communication Review*, 34(3):7–20, 2004.

[21] A. Sridharan, R. Guérin, and C. Diot. Achieving near-optimal traffic engineering solutions for current OSPF/IS-IS networks. *IEEE/ACM Transactions on Networking*, 2005.

[22] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.

[23] G. Vinnicombe. On the stability of networks operating TCP-like congestion control. *Proc. IFAC World Congress*, Barcelona, Spain 2002.

[24] G. Vinnicombe. Robust congestion control for the Internet. 2002.

[25] T. Voice. Delay stability results for congestion control algorithms with multi-path routing. 2004. http://www.statslab.cam.ac.uk/∼tdv20

[26] Z. Wang and J. Crowcroft. Analysis of shortest-path routing algorithms in a dynamic network environment. *Computer Communication Review*, 22(2):63–71, 1992.

[27] J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1:325–378, 1952.

[28] S. Yilmaz and I. Matta. On the scalability-performance tradeoffs in MPLS and IP routing. In *Proceedings of SPIE ITCOM'2002: Scalability and Traffic Control in IP Networks*, Boston, MA, 2002.

[29] D. Zhu, M. Gritter, and D.R. Cheriton. Feedback based routing. *Computer Communication Review*, 33(1):71–76, 2003.