Rate control for communication networks: shadow prices, proportional fairness and stability *

F.P. Kelly, A.K. Maulloo and D.K.H. Tan
University of Cambridge, UK

This paper analyses the stability and fairness of two classes of rate control algorithm for communication networks. The algorithms provide natural generalizations to large-scale networks of simple additive increase/multiplicative decrease schemes, and are shown to be stable about a system optimum characterized by a proportional fairness criterion. Stability is established by showing that, with an appropriate formulation of the overall optimization problem, the network's implicit objective function provides a Lyapunov function for the dynamical system defined by the rate control algorithm. The network's optimization problem may be cast in primal or dual form: this leads naturally to two classes of algorithm, which may be interpreted in terms of either congestion indication feedback signals or explicit rates based on shadow prices. Both classes of algorithm may be generalized to include routing control, and provide natural implementations of proportionally fair pricing.

Keywords: ATM network, congestion indication, elastic traffic, Internet, Lyapunov function, proportionally fair pricing, queues, routing, tatonnement.

Introduction

The design and control of modern communication networks raise several issues well suited to study using techniques of operational research such as optimization, network programming and stochastic modelling. In this paper we illustrate this theme, through the presentation and analysis of a mathematical model that arises in connection with the development and deployment of large-scale broadband networks.

In future communication networks there are expected to be applications that are able to modify their data transfer rates according to the available bandwidth within the network. Traffic from such applications is termed *elastic* [1]; a typical current example is TCP traffic over the Internet [2], and future examples may

---

*Correspondence: Frank Kelly, Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, UK. Email: f.p.kelly@statslab.cam.ac.uk .

include the controlled-load service of the Internet Engineering Task Force [3] and the Available Bit Rate transfer capability of ATM (asynchronous transfer mode) networks [4].

The key issue we address in this paper concerns how the available bandwidth within the network should be shared between competing streams of elastic traffic; in particular, we present a tractable mathematical model and use it to analyse the stability and fairness of a class of rate control algorithms. Traditionally stability has been considered an engineering issue, requiring an analysis of randomness and feedback operating on fast time-scales, while fairness has been considered an economic issue, involving static comparisons of utility. In future networks the intelligence embedded in end-systems, acting on behalf of human users, is likely to lessen the distinction between engineering and economic issues and increase the importance of an interdisciplinary view. (This general theme was the subject of the 1996 Blackett Memorial Lecture; further aspects are developed elsewhere [5].)

There is a substantial literature on rate control algorithms, recently reviewed by Hernandez-Valencia et al. [6]. Key early papers of Jacobson [2] and Chiu and Jain [7] identified the advantages of adaptive schemes that either increase flows linearly or decrease flows multiplicatively, depending on the absence or presence of congestion. Important recent papers of Bolot and Shankar [8], Fendick et al. [9] and Bonomi et al. [10] have analysed the stability of networks with a single bottleneck resource, where congestion is signalled by the build-up of a queue at the bottleneck's buffer, and where propagation delays are significant. (In wide-area networks propagation times may be significant in comparison with queueing times: for a transatlantic link of 600 Megabits per second, ten million bits may be in flight between queues.) The framework we adopt in this paper is simpler than that analysed by these authors in that we directly model only rates and not queue lengths, but more complex in that we model a network with an arbitrary number of bottleneck resources. Theoretical work [11], [12] on queues serving the superposition of a large number of streams indicates circumstances when the busy period preceding a buffer overflow may be relatively short, and several authors have argued the advantages of preventing queue build-up through the bounding of rates (see, for example, Charny et al. [13]).

Any discussion of the performance of a rate control scheme must address the issue of fairness, since there exist situations where a given scheme might maximize network throughput, for example, while denying access to some users. The most commonly discussed fairness criterion is that of max-min fairness: loosely, a set of rates is max-min fair if no rate may be increased without simultaneously decreasing another rate which is already smaller. In a network with a single bottleneck resource max-min fairness implies an equal share of the resource for each flow through it. Mazumdar et al. [14] have pointed out that from a game-theoretic standpoint such an allocation is not special, and have advocated instead the Nash bargaining solution, from cooperative game theory, as capturing natural assumptions as to what constitutes fairness.

The need for networks to operate in a public (and therefore potentially non-cooperative) environment has stimulated work on charging schemes for broadband networks: see Kelly [15] for a scheme based on time and volume measurements for non-elastic traffic, MacKie-Mason and Varian [16] for a description of a 'smart market' based on a per-packet charge when the network is congested, and the collection edited by McKnight and Bailey [17] for several further papers and references. Kelly [18] describes a model for elastic traffic in which a user chooses the charge per unit time that the user is willing to pay; thereafter the user's rate is determined by the network according to a proportional fairness criterion applied to the rate per unit charge. It was shown that a system optimum is achieved when users' choices of charges and the network's choice of allocated rates are in equilibrium. There remained the question of how the proportional fairness criterion could be implemented in a large-scale network. In this paper we show that simple rate control algorithms, using additive increase/multiplicative decrease rules or explicit rates based on resource shadow prices, can provide stable convergence to proportional fairness per unit charge, even in the presence of random effects and delays.

Mechanisms by which supply and demand reach equilibrium have, of course, long been a central concern of economists, and there exists a substantial body of theory on the stability of what are termed tatonnement processes [19], [20], [21]. From this viewpoint the rate control algorithms described in this paper are particular embodiments of a 'Walrasian auctioneer', searching for market clearing prices. The 'Walrasian auctioneer' of tatonnement theory is usually considered a rather implausible construct; we show that the structure of a communication network provides a natural context within which to investigate the consequences for a tatonnement process of stochastic perturbations and time lags.

The organization of the paper is as follows. In the next section we describe our basic model of a network, describe two classes of rate control algorithm, and provide an outline of our results. Detailed proofs are provided in the next two sections, following which we illustrate our theoretical results through a discussion of some numerical examples. We then consider user adaptation and routing, and finally conclude with some remarks on open issues.

Outline of results

The basic model

Consider a network with a set J of *resources*, and let C_j be the finite capacity of resource j, for $j \in J$. Let a *route* r be a non-empty subset of J, and write R for the set of possible routes. Set $A_{jr} = 1$ if $j \in r$, so that resource j lies on route r, and set $A_{jr} = 0$ otherwise. This defines a $0-1$ matrix $A = (A_{jr}, j \in J, r \in R)$.

Associate a route r with a user, and suppose that if a rate x_r is allocated to user r then this has utility $U_r(x_r)$ to the user. Assume that the utility $U_r(x_r)$ is an increasing, strictly concave and continuously differentiable function of x_r over the range $x_r \geq 0$ (following Shenker [1], we call traffic that leads to such a utility function *elastic* traffic). Assume further that utilities are additive, so that the aggregate utility of rates $x = (x_r, r \in R)$ is $\sum_{r \in R} U_r(x_r)$. Let $U = (U_r(\cdot), r \in R)$ and $C = (C_j, j \in J)$. Under this model the system optimal rates solve the following problem.

$SYSTEM(U, A, C)$:

$$\begin{array}{ll} \text{maximize} & \sum_{r \in R} U_r(x_r) \\ \text{subject to} & Ax \leq C \\ \text{over} & x \geq 0. \end{array}$$

While this optimization problem is mathematically fairly tractable (with a strictly concave objective function and a convex feasible region), it involves utilities U that are unlikely to be known by the network. We are thus led to consider two simpler problems.

Suppose that user r may choose an amount to pay per unit time, w_r, and receives in return a flow x_r proportional to w_r, say $x_r = w_r / \lambda_r$, where λ_r could be regarded as a charge per unit flow for user r. Then the utility maximization problem for user r is as follows.

$USER_r(U_r; \lambda_r)$:

$$\begin{array}{ll} \text{maximize} & U_r\left(\frac{w_r}{\lambda_r}\right) - w_r \\ \text{over} & w_r \geq 0. \end{array}$$

Suppose next that the network knows the vector $w = (w_r, r \in R)$, and attempts to maximize the function $\sum_r w_r \log x_r$. The network's optimization problem is then as follows.

$NETWORK(A, C; w)$:

$$\begin{array}{ll} \text{maximize} & \sum_{r \in R} w_r \log x_r \\ \text{subject to} & Ax \leq C \\ \text{over} & x \geq 0. \end{array}$$

It is known [18] that there always exist vectors $\lambda = (\lambda_r, r \in R)$, $w = (w_r, r \in R)$ and $x = (x_r, r \in R)$, satisfying $w_r = \lambda_r x_r$ for $r \in R$, such that w_r solves $USER_r(U_r; \lambda_r)$ for $r \in R$ and x solves $NETWORK(A, C; w)$; further, the vector x is then the unique solution to $SYSTEM(U, A, C)$.

A vector of rates $x = (x_r, r \in R)$ is *proportionally fair* if it is feasible, that is $x \geq 0$ and $Ax \leq C$, and if for any other feasible vector x^*, the aggregate of proportional changes is zero or negative:

$$\sum_{r \in R} \frac{x_r^* - x_r}{x_r} \leq 0. \tag{1}$$

If $w_r = 1, r \in R$, then a vector of rates x solves $NETWORK(A, C; w)$ if and only if it is proportionally fair. Such a vector is also the Nash bargaining solution (satisfying certain axioms of fairness [22]), and, as such, has been advocated in the context of telecommunications by Mazumdar et al. [14].

A vector x is such that the *rates per unit charge* are proportionally fair if x is feasible, and if for any other feasible vector x^*

$$\sum_{r \in R} w_r \frac{x_r^* - x_r}{x_r} \leq 0. \tag{2}$$

The relationship between the conditions (1) and (2) is well illustrated when $w_r, r \in R$, are all integral. For each $r \in R$, replace the single user r by w_r identical sub-users, construct the proportionally fair allocation over the resulting $\sum_r w_r$ users, and provide to user r the aggregate rate allocated to its w_r sub-users; then the resulting rates *per unit charge* are proportionally fair. This construction also illustrates the need to adapt the notion of fairness to a non-cooperative context, where it is possible for a single user to represent itself as several distinct users. It is straightforward to check [18] that a vector of rates x solves $NETWORK(A, C; w)$ if and only if the rates per unit charge are proportionally fair.

We note in passing that if, for a fixed set of users and arbitrary parameters $w = (w_r, r \in R)$, the network solves $NETWORK(A, C; w)$, then the resulting rates $x = (x_r, r \in R)$ solve a variant of the problem $SYSTEM(U, A, C)$, with a weighted objective function $\sum_r \alpha_r U_r(x_r)$ where $\alpha_r = w_r/(x_r U_r'(x_r))$ for $r \in R$. Thus a choice of the parameters $w = (w_r, r \in R)$ by the network (rather than by users) corresponds to an implicit weighting by the network of the relative utilities of different users, with weights related to the users' various marginal utilities.

Under the decomposition of the problem $SYSTEM(U, A, C)$ into the problems $NETWORK(A, C; w)$ and $USER_r(U_r; \lambda_r)$, $r \in R$, the utility function $U_r(x_r)$ is not required by the network, and only appears in the optimization problem faced by user r. The Lagrangian [23] for the problem $NETWORK(A, C; w)$ is

$$L(x, z; \mu) = \sum_{r \in R} w_r \log x_r + \mu^T (C - Ax - z)$$

where $z \geq 0$ is a vector of slack variables and μ is a vector of Lagrange multipliers (or shadow prices). Then

$$\frac{\partial L}{\partial x_r} = \frac{w_r}{x_r} - \sum_{j \in r} \mu_j,$$

and so the unique optimum to the primal problem is given by

$$x_r = \frac{w_r}{\sum_{j \in r} \mu_j} \tag{3}$$

where $(x_r, r \in R)$, $(\mu_j, j \in J)$ solve

$$\mu \geq 0, \quad Ax \leq C, \quad \mu^T(C - Ax) = 0 \tag{4}$$

and relation (3). Furthermore the associated dual problem quickly reduces, after elision of terms not dependent on the shadow prices μ, to the following problem.

$DUAL(A, C; w)$**:**

$$\begin{array}{ll} \text{maximize} & \sum_{r \in R} w_r \log(\sum_{j \in r} \mu_j) - \sum_{j \in J} \mu_j C_j \\ \text{over} & \mu \geq 0. \end{array}$$

While the problems $NETWORK(A, C; w)$ and $DUAL(A, C; w)$ are mathematically tractable, it would be difficult to implement a solution in any centralized manner. A centralized processor, even if it were itself completely reliable and could cope with the complexity of the computational task involved, would have its lines of communication through the network vulnerable to delays and failures. Rather, interest focuses on algorithms which are decentralized and of a simple form: the challenge is to understand how such algorithms can be designed so that the network as a whole reacts intelligently to perturbations. Next we describe two simple classes of decentralized algorithm, designed to implement solutions to relaxations of the problems $NETWORK(A, C; w)$ and $DUAL(A, C; w)$.

A primal algorithm

Consider the system of differential equations

$$\frac{d}{dt} x_r(t) = \kappa\Big(w_r - x_r(t) \sum_{j \in r} \mu_j(t)\Big) \tag{5}$$

where

$$\mu_j(t) = p_j\Big(\sum_{s: j \in s} x_s(t)\Big). \tag{6}$$

(Here and throughout we assume that, unless otherwise specified, r ranges over the set R and j ranges over the set J.) We may motivate the relations (5)–(6) in several ways. For example, suppose that $p_j(y)$ is a price charged by resource j, per unit flow through resource j, when the total flow through resource j is y. Then by adjusting the flow on route r, $x_r(t)$, in accordance with equations (5)–(6), the network attempts to equalize the aggregate cost of this flow, $x_r(t) \sum_{j \in r} \mu_j(t)$, with a target value w_r, for every $r \in R$. (For an enlightening description of the technological implementation of such algorithms in an ATM network, see Courcoubetis et al. [24].)

For an alternative motivation, suppose that resource j generates a continuous stream of feedback signals at rate $p_j(y)$ when the total flow through resource j is y. Suppose further that when resource j generates a feedback signal, a copy is sent to each user r whose route passes through resource j, where it is interpreted as a congestion indicator requiring some reduction in the flow x_r. Then equation (5) corresponds to a response by user r that comprises two components: a steady increase at rate proportional to w_r, and a multiplicative decrease at rate proportional to the stream of feedback signals received. (For early discussions of algorithms with additive increase and multiplicative decrease see Chiu and Jain [7] and Jacobson [2]; Hernandez-Valencia et al. [6] review several algorithms based on congestion indication feedback.)

Later we establish that under mild regularity conditions on the functions p_j, $j \in J$, the expression

$$\mathcal{U}(x) = \sum_{r \in R} w_r \log x_r - \sum_{j \in J} \int_0^{\sum_{s:j \in s} x_s} p_j(y)dy \tag{7}$$

provides a Lyapunov function for the system of differential equations (5)–(6), and we deduce that the vector x maximizing $\mathcal{U}(x)$ is a stable point of the system, to which all trajectories converge.

The functions $p_j, j \in J$, may be chosen so that the maximization of the Lyapunov function $\mathcal{U}(x)$ arbitrarily closely approximates the optimization problem $NETWORK(A, C; w)$, and, in this sense, is a relaxation of the network problem. In our penultimate section we shall see that certain relaxations correspond naturally to a system objective which takes into account loss or delays, as well as flow rates.

The Lyapunov function (7) thus provides an enlightening analysis of the global stability of the system (5)–(6), and of the relationship between this system and the problem $NETWORK(A, C; w)$. However, the system (5)–(6) has omitted to model two important aspects of decentralized systems, namely stochastic perturbations, and time lags. We analyze these aspects by considering small perturbations to the stable point x.

Stochastic perturbations within the network may well arise from a resource's method of sensing its load. Equation (6) represents the response $\mu_j(t)$ of resource j as a continuous function of a load, $y = \sum_{s:j \in s} x_s$, which is assumed known. In practice a resource may assess its load by an error-prone measurement mechanism, and then choose a feedback signal from a small set of possible signals. (See Hernandez-Valencia et al. [6] and Bonomi et al. [10] for more detailed descriptions of binary feedback and congestion indication rate control algorithms.) In the next section we describe how such mechanisms motivate various stochastic models of the network. One particular model takes the form

$$dx_r(t) = \kappa\Big(w_r dt - x_r(t)\sum_{j\in r}\Big(\mu_j(t)dt + \mu_j(t)^{1/2}\varepsilon_j^{1/2}dB_j(t)\Big)\Big) \tag{8}$$

where $B_j(t)$ is a standard Brownian motion, representing stochastic effects at resource j, and ε_j is a scaling parameter for these effects. If the scaling parameters $\varepsilon_j, j \in J$, are small then the stochastic differential equation (8) has, as solution, a multidimensional Ornstein–Uhlenbeck process, centred on the stable point x of the differential equations (5)–(6). The stationary distribution for $(x_r(t), r \in R)$ is a multivariate normal distribution, with covariance matrix that can be explicitly calculated in terms of the parameters of the network.

Similarly we shall describe a model incorporating time lags that generalizes equations (5)–(6), and shall analyse its behaviour close to the stable point x. Our models of stochastic effects and of time-lags provide important insights into the behaviour of the network, and allow us to quantify the various relationships and trade-offs between speed of convergence, the magnitude of fluctuations about the equilibrium point, and the stability of the network.

A dual algorithm

The equations (5)–(6) represent a system where rates vary gradually, and shadow prices are given as functions of the rates. Next we consider a system where shadow prices vary gradually, with rates given as functions of the shadow prices. Let

$$\frac{d}{dt}\,\mu_j(t) = \kappa\Big(\sum_{r:j\in r} x_r(t) - q_j\big(\mu_j(t)\big)\Big) \tag{9}$$

where

$$x_r(t) = \frac{w_r}{\sum_{k\in r}\mu_k(t)}. \tag{10}$$

The relationship between the algorithm (9)–(10) and the problem $DUAL(A, C; w)$ parallels that between the primal algorithm (5)–(6) and the problem $NETWORK(A, C; w)$, and, again, we may motivate the algorithm in several ways. For example, suppose that $q_j(\eta)$ is the flow through resource j which generates a price at resource j of η. Then an economist would describe the right hand side of equation (9) as the vector of excess demand at prices $(\mu_j(t), j \in J)$, and would recognise equations (9)–(10) as a tatonnement process by which prices adjust according to supply and demand (Varian [21], Chapter 21).

Later we establish that under mild regularity conditions on the functions q_j, $j \in J$, the expression

$$\mathcal{V}(\mu) = \sum_{r\in R} w_r \log\Big(\sum_{j\in r}\mu_j\Big) - \sum_{j\in J}\int_0^{\mu_j} q_j(\eta)d\eta \tag{11}$$

provides a Lyapunov function for the system of differential equations (9)–(10), and we deduce that the vector μ maximizing $\mathcal{V}(\mu)$ is a stable point of the system, to which all trajectories converge. Further, by appropriate choice of the functions $q_j, j \in J$, the maximization of the function $\mathcal{V}(\mu)$ can arbitrarily approximate the problem $DUAL(A, C; w)$.

We consider stochastic perturbations of system (9)–(10), with a typical example taking the form

$$d\mu_j(t) = \kappa\Big(\sum_{r:j\in r}\Big(x_r(t)dt + x_r(t)^{1/2}\varepsilon_r^{1/2}dB_r(t)\Big) - q_j\Big(\mu_j(t)\Big)dt\Big) \tag{12}$$

where $B_r(t)$ is a standard Brownian motion, representing stochastic effects associated with the flow on route r. If the scaling parameters $\varepsilon_r, r \in R$, are small then the stationary distribution for $(\mu_j(t), j \in J)$ is centred on the stable point μ of the differential equations (9)–(10), with a covariance matrix that can be explicitly calculated in terms of the parameters of the network. Also it is possible to analyse the stability of the model (9)–(10) when time-lags are introduced.

User adaptation

Our analyses of the primal algorithm (5)–(6) and the dual algorithm (9)–(10) assume that the parameters $(w_r, r \in R)$ chosen by the users are fixed, at least on the time scales concerned in the analyses. With increasing intelligence embedded in end-systems, users may in the future be able to vary the parameters $(w_r, r \in R)$ even within these short time scales. Both the algorithms may be extended to this situation.

Suppose that user r is able to monitor its rate $x_r(t)$ continuously, and to vary smoothly the parameter $w_r(t)$ so as to track accurately the optimum to $USER_r(U_r; \lambda_r(t))$, where $\lambda_r(t) = w_r(t)/x_r(t)$ is the charge per unit flow to user r at time t. Then, using revised Lyapunov functions, stability of both the primal and dual algorithms may again be established.

Our next sections provide detailed proofs of the various results outlined above, together with some numerical illustrations. In our penultimate section we shall look again at the system decomposition relating the problems $SYSTEM(U, A, C)$ and $NETWORK(A, C; w)$, and extend the discussion to include routing control.

A primal algorithm

In this section we establish the global stability of the primal algorithm (5)–(6), determine the rate of convergence, and, by considering perturbations about the stable point, model stochastic effects and time lags.

Global stability

Let the function $\mathcal{U}(x)$ be defined by equation (7) where $w_r > 0$, $r \in R$, and suppose that, for $j \in J$, the function $p_j(y)$, $y \geq 0$, is a non-negative, continuous, increasing function of y, not identically zero.

Theorem 1 *The strictly concave function* $\mathcal{U}(x)$ *is a Lyapunov function for the system of differential equations (5)–(6). The unique value* x *maximizing* $\mathcal{U}(x)$ *is a stable point of the system, to which all trajectories converge.*

Proof. The assumptions on $w_r > 0, r \in R$, and $p_j, j \in J$, ensure that $\mathcal{U}(x)$ is strictly concave on $x \geq 0$ with an interior maximum; the maximizing value of x is thus unique. Observe that

$$\frac{\partial}{\partial x_r}\mathcal{U}(x) = \frac{w_r}{x_r} - \sum_{j \in r} p_j\Big(\sum_{s: j \in s} x_s\Big); \tag{13}$$

setting these derivatives to zero identifies the maximum. Further

$$\begin{aligned}\frac{d}{dt}\mathcal{U}(x(t)) &= \sum_{r \in R} \frac{\partial \mathcal{U}}{\partial x_r} \cdot \frac{d}{dt} x_r(t) \\ &= \kappa \sum_{r \in R} \frac{1}{x_r(t)} \Big(w_r - x_r(t) \sum_{j \in r} p_j\Big(\sum_{s: j \in s} x_s(t)\Big)\Big)^2,\end{aligned}$$

establishing that $\mathcal{U}(x(t))$ is strictly increasing with t, unless $x(t) = x$, the unique x maximizing $\mathcal{U}(x)$. The function $\mathcal{U}(x)$ is thus a Lyapunov function for the system (5)–(6), and the theorem follows ([25], Chapter 5). □

Define the continuous functions $p_j(y) = (y - C_j + \varepsilon)^+/\varepsilon^2$ for $j \in J$. Then, as $\varepsilon \to 0$, the maximization of the Lyapunov function $\mathcal{U}(x)$ approximates arbitrarily closely the primal problem $NETWORK(A, C; w)$; in particular the vector x maximizing $\mathcal{U}(x)$ approaches the solution x to relations (3) and (4). Note, however, that the derivative $p_j'(y)$ may become arbitrarily large as the approximation improves.

Rate of convergence

We have seen, in Theorem 1, that the system (5)–(6) converges to a unique stable point: next we investigate the rate of convergence, by linearization about the stable point.

Let x identify the unique vector maximizing $\mathcal{U}(x)$, let $\mu_j = p_j(\sum_{s: j \in s} x_s)$, and suppose p_j is differentiable at this point, with derivative p_j'. Let $x_r(t) =$

$x_r + x_r^{1/2} y_r(t)$. Then, linearizing the system (5)–(6) about x, we obtain

$$\begin{aligned}\frac{d}{dt} y_r(t) &= -\kappa(y_r(t) \sum_{j \in r} \mu_j + x_r^{1/2} \sum_{j \in r} p_j' \sum_{s:j \in s} x_s^{1/2} y_s(t)) \\ &= -\kappa \left(\frac{w_r}{x_r} y_r(t) + x_r^{1/2} \sum_j \sum_s p_j' A_{jr} A_{js} x_s^{1/2} y_s(t) \right).\end{aligned}$$

We may write this in matrix form as

$$\frac{d}{dt} y(t) = -\kappa(WX^{-1} + X^{1/2} A^T P' A X^{1/2}) y(t) \tag{14}$$

where $X = \mathrm{diag}(x_r, r \in R)$, $W = \mathrm{diag}(w_r, r \in R)$ and $P' = \mathrm{diag}(p_j', j \in J)$.

Let

$$\Gamma^T \Phi \Gamma = WX^{-1} + X^{1/2} A^T P' A X^{1/2} \tag{15}$$

where Γ is an orthogonal matrix, $\Gamma^T \Gamma = I$, and $\Phi = \mathrm{diag}(\phi_r, r \in R)$ is the matrix of eigenvalues, necessarily positive, of the real, symmetric, positive definite matrix (15). Then

$$\frac{d}{dt} y(t) = -\kappa \Gamma^T \Phi \Gamma y(t), \tag{16}$$

and thus the rate of convergence to the stable point is determined by the smallest eigenvalue, $\phi_r, r \in R$, of the matrix (15). Note that speed of convergence increases both with the gain parameter κ and with the magnitude of the derivatives P'; we shall see that this conclusion requires qualification in the presence of either stochastic effects or of time-lags.

Our early assumption that $p_j(y)$, $j \in J$, are increasing functions is convenient and often natural: it implies that $\mathcal{U}$ is strictly concave with an interior maximum. If the functions $p_j(y)$, $j \in J$, are not increasing, then $\mathcal{U}(x)$ may not have an interior maximum or it may have multiple stationary points: we describe an example later. Provided $p_j(y)$, $j \in J$, are differentiable at a stationary point, the local behaviour near the stationary point is described by the equation (16).

Stochastic analysis

Next we consider a stochastic perturbation of the linearized equation (16). Let

$$dy(t) = -\kappa \Big(\Gamma^T \Phi \Gamma y(t) dt + F dB(t) \Big) \tag{17}$$

where F is an arbitrary $|R| \times |I|$ matrix and $B(t) = (B_i(t), i \in I)$ is a collection of independent standard Brownian motions, extended to $-\infty < t < \infty$. (Later we describe how the modelling of different sources of randomness may lead to various explicit forms for the matrix F.)

The stationary solution to the system (17) is

$$y(t) = -\kappa \int_{-\infty}^{t} e^{-\kappa(t-\tau)\Gamma^T \Phi \Gamma} F dB(\tau), \tag{18}$$

as can be checked by differentiating both sides of equation (18) with respect to t. The solution (18) is a linear transformation of the Gaussian process $(B(\tau), \tau < t)$; hence $y(t)$ has a multivariate normal distribution, $y(t) \sim N(0, \Sigma)$, where

$$\begin{aligned} \Sigma &= \mathbb{E}\Big[y(t)y(t)^T\Big] \\ &= \kappa^2 \int_{-\infty}^{0} e^{\kappa\tau\Gamma^T\Phi\Gamma} F F^T e^{\kappa\tau\Gamma^T\Phi\Gamma} d\tau \\ &= \kappa\Gamma^T \left[\int_{-\infty}^{0} e^{\tau\Phi}\Gamma F F^T \Gamma^T e^{\tau\Phi} d\tau\right] \Gamma. \end{aligned}$$

Define the symmetric matrix $[\Gamma F; \Phi]$ by

$$\begin{aligned} [\Gamma F; \Phi]_{rs} &= \left[\int_{-\infty}^{0} e^{\tau\Phi}\Gamma F F^T \Gamma^T e^{\tau\Phi} d\tau\right]_{rs} \\ &= \frac{[\Gamma F F^T \Gamma^T]_{rs}}{\phi_r + \phi_s}. \end{aligned}$$

Then

$$\Sigma = \kappa\Gamma^T [\Gamma F; \Phi] \Gamma. \tag{19}$$

Note that the covariance matrix increases linearly with the gain parameter κ; as κ increases, the faster convergence to equilibrium described by relation (16) is at the cost of a greater spread at equilibrium. Varying the derivatives P' has a more subtle effect, through relation (15) and the construction (19), on the covariance matrix; broadly, as P' increases, not only is convergence to equilibrium faster, but also the spread at equilibrium decreases. However, we shall see later that, in the presence of time-lags, increasing P' may compromise stability.

We next illustrate how various sources of randomness may lead to different covariance structures.

Congestion indication with joint feedback. Consider the following stochastic version of the system (5)–(6). Let $(N_j(\tau), \tau \geq 0)$, for $j \in J$, be a collection of independent unit rate Poisson processes, and let

$$dx_r(t) = \kappa \left(w_r dt - x_r(t) \sum_{j \in r} \varepsilon_j dN_j \left(\varepsilon_j^{-1} \int_0^t \mu_j(\tau) d\tau \right) \right), \tag{20}$$

where the functions $\mu_j(\cdot)$, for $j \in J$, are given by equation (6). Equation (20) would describe the following model: resource j generates feedback signals indicating congestion as a time-dependent Poisson process at rate $\varepsilon_j^{-1}\mu_j(t)$; when resource j generates a feedback signal, a copy is sent to each user r whose route passes through resource j; and user r reacts to such a feedback signal by reducing $x_r(t)$ by an amount $\kappa\varepsilon_j x_r(t)$.

Now as $\varepsilon \to 0$, the normalized Poisson process $((\varepsilon N_j(\tau/\varepsilon) - \tau)\varepsilon^{-1/2}, \tau \geq 0)$ converges in distribution to a standard Brownian motion. This motivates the approximation, valid when ε_j are small, of the Poisson driving equation (20) by its Brownian version

$$dx_r(t) = \kappa\left(w_r dt - x_r(t)\sum_{j\in r}\left(\mu_j(t)dt + \varepsilon_j^{1/2}\mu_j(t)^{1/2}dB_j(t)\right)\right)$$

where $(B_j(t), t \geq 0)$, for $j \in J$, are a collection of independent standard Brownian motions.

The corresponding Brownian version of the linearized system (16) is just equation (17) where $B(t) = (B_j(t), j \in J)$, and F is an $|R| \times |J|$ matrix with elements

$$F_{rj} = \varepsilon_j^{1/2}\mu_j^{1/2}A_{jr}x_r^{1/2}. \tag{21}$$

Then

$$FF^T = X^{1/2}A^T EPAX^{1/2}, \tag{22}$$

where $E = \mathrm{diag}(\varepsilon_j, j \in J)$ and $P = \mathrm{diag}(\mu_j, j \in J)$, and hence the stationary covariance matrix Σ may be calculated from expression (19).

Congestion indication with individual feedback. Consider next the Poisson driving equation

$$dx_r(t) = \kappa\left(w_r dt - \sum_{j\in r}\varepsilon_j dN_{jr}\left(\varepsilon_j^{-1}\int_0^t x_r(\tau)\mu_j(\tau)d\tau\right)\right) \tag{23}$$

where $(N_{jr}(\tau), \tau \geq 0)$, for $j \in J, r \in R$, are a collection of independent unit rate Poisson processes. This would describe the following model: feedback signals from resource j to user r arise at rate $\varepsilon_j^{-1}x_r(t)\mu_j(t)$; and user r reacts to such a feedback signal by reducing $x_r(t)$ by an amount $\kappa\varepsilon_j$. The Brownian approximation, valid when ε_j are small, becomes

$$dx_r(t) = \kappa\Big(w_r dt - x_r(t)\sum_{j\in r}\Big(\mu_j(t)dt + \varepsilon_j^{1/2}x_r(t)^{-1/2}\mu_j(t)^{1/2}dB_{jr}(t)\Big)\Big),$$

whose linearization is equation (17) where the $|R| \times |J|.|R|$ matrix F is given by

$$F_{r,(j,s)} = \varepsilon_j^{1/2}\mu_j^{1/2}A_{jr}I[r = s]. \tag{24}$$

Thus FF^T is the matrix $\mathrm{diag}(\sum_{j\in r}\mu_j\varepsilon_j, r \in R)$, and the stationary covariance matrix Σ may be calculated from expression (19). Later we provide a numerical illustration of this calculation, and contrast the results derived from the forms (21) and (24).

Source fluctuations. Consider the Brownian driving equation

$$dx_r(t) = \kappa\Big(w_r dt - x_r(t)\sum_{j\in r}\mu_j(t)dt + \varepsilon_r^{1/2}x_r(t)^{1/2}dB_r(t)\Big),$$

which might correspond to fluctuations arising at sources, rather than within the network. For this system the $|R| \times |R|$ matrix F is the diagonal matrix $\mathrm{diag}(\varepsilon_r^{1/2}, r \in R)$.

Time lags

Consider next the lagged, discrete time system

$$x_r[t+1] = x_r[t] + \kappa\Big(w_r - x_r[t]\sum_{j\in r}\mu_j[t - d(j,r)]\Big) \tag{25}$$

where

$$\mu_j[t] = p_j\Big(\sum_{s:j\in s} x_s[t - d(j,s)]\Big), \tag{26}$$

and $d(j,r)$, $j \in J, r \in R$, are non-negative integers. This might correspond to a model of congestion indication with joint feedback, where there is a delay of $d(j,r)$ between resource j generating a feedback signal and user r receiving it, and another delay of $d(j,r)$ between user r changing its rate and the altered flow reaching resource j. Say that a vector x is an *equilibrium point* of the system (25)–(26) if $x_r[t] = x_r$, for $t = \ldots, 0, 1, 2, \ldots$, satisfies these equations.

Theorem 2 *The vector x maximizing the strictly concave function $\mathcal{U}(x)$ is the unique equilibrium point of the system (25)–(26).*

Proof. The vector x is an equilibrium point if and only if it solves

$$w_r = x_r \sum_{j\in r} p_j\Big(\sum_{s:j\in s} x_s\Big).$$

But this is precisely the stationarity condition implied by the partial derivatives (13) of the function $\mathcal{U}(x)$, a strictly concave function with a unique maximum. □

For small enough values of κ the equilibrium point will be asymptotically stable, since if we replace κ by $\kappa\delta$ in equation (25) and let $x_r(t) = x_r[t/\delta]$, then as $\delta \to 0$ we may approximate arbitrarily closely a solution to equations (5)–(6). But for small values of κ convergence to the equilibrium point is slow, and so it is of interest to investigate the local stability of the equilibrium point for general values of κ.

Let $\mu_j = p_j[\sum_{s:j\in s} x_s]$, and suppose p_j is differentiable at this point, with derivative p'_j. Let $x_r[t] = x_r + x_r^{1/2} y_r[t]$. Then, linearizing the system (25)–(26) about x, we obtain

$$\begin{aligned}
&y_r[t+1] \qquad (27)\\
&\quad = y_r[t] - \kappa\Big(y_r[t]\sum_{j\in r}\mu_j + x_r^{1/2}\sum_{j\in r} p'_j \sum_{s:j\in s} x_s^{1/2} y_s[t-d(j,r)-d(j,s)]\Big)\\
&\quad = y_r[t] - \kappa\left(\frac{w_r}{x_r}\, y_r[t] + \sum_j\sum_s p'_j A_{jr} A_{js} x_r^{1/2} x_s^{1/2} y_s[t-d(j,r)-d(j,s)]\right).
\end{aligned}$$

Define the $|R|\times|R|$ matrices $(L[d], d = 0,1,\ldots,D)$ by

$$(L[d])_{rs} = \sum_j p'_j A_{jr} A_{js} x_r^{1/2} x_s^{1/2} I[d(j,r)+d(j,s)=d]$$

where $D = \max_{j,r,s}\{d(j,r)+d(j,s)\}$. Thus

$$\sum_{d=0}^{D} L[d] = X^{1/2} A^T P' A X^{1/2},$$

the second term of the key matrix (15). Define the vector $y[t] = (y_r[t], r\in R)$. Then we can rewrite equation (28) in the matrix form

$$\begin{pmatrix} y[t+1] \\ y[t] \\ \vdots \\ y[t-D+1] \end{pmatrix} = L \begin{pmatrix} y[t] \\ y[t-1] \\ \vdots \\ y[t-D] \end{pmatrix} \qquad (28)$$

where

$$L = \begin{pmatrix} I-\kappa(WX^{-1}+L[0]) & -\kappa L[1] & -\kappa L[2] & \ldots & -\kappa L[D] \\ I & 0 & 0 & \ldots & 0 \\ 0 & I & 0 & & \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}. \qquad (29)$$

The equilibrium point x of the system (25)–(26) is stable if and only if the spectral radius of the matrix L is less than unity. Recall that in our model of stochastic effects, increasing the derivatives P' had broadly the same reductive effect on the covariance matrix (19) as *decreasing* the gain parameter κ; in contrast the destabilising effect on the matrix (29) of increasing P' is broadly the same as *increasing* κ.

For simplicity of notation we have used the same gain parameter κ for each $r\in R$. If κ is replaced by κ_r in equation (25), then we again obtain relations (28)–(29), but now with κ interpreted as the matrix $\text{diag}(\kappa_r, r\in R)$. An interesting topic concerns how the time delays within a network affect the choice of gain parameters; we might for example study the problem of choosing $\mathrm{diag}(\kappa_r, r \in R)$ in order to minimize the spectral radius of the matrix L.

There exist other natural discrete time versions of the equations (5)–(6), and these too may be analysed in a similar manner. For example, consider the method of repeated substitution, $x_r[t+1] = w_r / \sum_{j\in r} \mu_j[t-d(j,r)]$ or its damped version

$$x_r[t+1] = (1-\kappa)x_r[t] + \kappa \frac{w_r}{\sum_{j\in r} \mu_j[t-d(j,r)]}$$

where $\mu_j[t]$ is again given by equation (26). Then the linearized relations (28)–(29) are altered in that the top row of the matrix (29) becomes

$$(I - \kappa(I + XW^{-1}L[0]), -\kappa XW^{-1}L[1], \ldots, -\kappa XW^{-1}L[D]).$$

A dual algorithm

In this section we investigate the stability of the dual algorithm (9)–(10), including a perturbation analysis of stochastic effects and time lags. Finally we note that the system (9)–(10) is just one example of a dual algorithm, and consider variants that share the Lyapunov function (11).

Global stability

Let the function $\mathcal{V}(\mu)$ be defined by equation (11), where $w_r > 0$, $r \in R$, and suppose that, for $j \in J, q_j(0) = 0$ and $q_j(\eta), \eta \geq 0$, is a continuous, strictly increasing function of η.

Theorem 3 *The strictly concave function $\mathcal{V}(\mu)$ is a Lyapunov function for the system of differential equations (9)–(10). The unique value μ maximizing $\mathcal{V}(\mu)$ is a stable point of the system, to which all trajectories converge.*

Proof. The assumptions on $w_r > 0, r \in R$, and on $q_j, j \in J$, ensure that $\mathcal{V}(\mu)$ is strictly concave on $\mu \geq 0$ with an interior maximum; the maximizing value of μ is thus unique, and is determined by setting the derivatives

$$\frac{\partial}{\partial \mu_j} \mathcal{V}(\mu) = \sum_{r:j\in r} \frac{w_r}{\sum_{k\in r} \mu_k} - q_j(\mu_j) \tag{30}$$

to zero. Also,

$$\begin{aligned} \frac{d}{dt} \mathcal{V}(\mu(t)) &= \sum_{j\in J} \frac{\partial \mathcal{V}}{\partial \mu_j} \cdot \frac{d}{dt} \mu_j(t) \\ &= \kappa \sum_{j\in J} \Big(\sum_{r:j\in r} \frac{w_r}{\sum_{k\in r} \mu_k(t)} - q_j(\mu_j(t)) \Big)^2, \end{aligned}$$

establishing that $\mathcal{V}(\mu(t))$ is strictly increasing with t, unless $\mu(t) = \mu$, the unique value μ maximizing $\mathcal{V}(\mu)$. The function $\mathcal{V}(\mu)$ is thus a Lyapunov function for the system (9)–(10), and the theorem follows [25]. □

The maximization of the Lyapunov function $\mathcal{V}(\mu)$ becomes the dual problem if, for $j \in J$, $\eta > 0$, $q_j(\eta) = C_j$. These functions violate the assumption that $q_j(\eta)$ is continuous at $\eta = 0$, but they may be arbitrarily closely approximated, for example by the functions $q_j(\eta) = C_j\eta/(\eta+\varepsilon)$ for small positive ε. Note, however, that the derivative $q_j'(\eta)$ may become arbitrarily large as the approximation improves.

Rate of convergence

Let μ identify the unique vector maximizing $\mathcal{V}(\mu)$, let $x_r = w_r/\sum_{k\in r}\mu_k$, and suppose $q_j(y)$ differentiable at the point $y = \mu_j$, with derivative q_j'. Let $\mu_j(t) = \mu_j + \xi_j(t)$. Then, linearizing the system (9)–(10) about μ, we obtain, after some reduction,

$$\frac{d}{dt}\xi(t) = -\kappa(AXW^{-1}XA^T + Q')\xi(t)$$

where $W = \mathrm{diag}(w_r, r \in R)$ and $Q' = \mathrm{diag}(q_j', j \in J)$. Let

$$\Theta^T\Psi\Theta = AXW^{-1}XA^T + Q' \tag{31}$$

where Θ is an orthogonal matrix, $\Theta^T\Theta = I$, and $\Psi = \mathrm{diag}(\psi_j, j \in J)$ is the matrix of eigenvalues, necessarily non-negative, of the real, symmetric, positive semi-definite matrix (31). Then

$$\frac{d}{dt}\xi(t) = -\kappa\Theta^T\Psi\Theta\xi(t), \tag{32}$$

and thus the rate of convergence to the stable point is determined by the smallest eigenvalue of the matrix (31). Note that speed of convergence increases both with the gain parameter κ and with the magnitude of the derivatives Q'.

Stochastic analysis

Next consider a stochastic perturbation of the linearized equation (32). Let

$$d\xi(t) = -\kappa\left(\Theta^T\Psi\Theta\xi(t)dt - GdB(t)\right) \tag{33}$$

where $B(t) = (B_i(t), i \in I)$ is a collection of independent standard Brownian motions, and G is a $|J| \times |I|$ matrix.

A similar analysis to that of the last section determines the stationary covariance matrix Σ of $\xi(t)$. Define the symmetric matrix $[\Theta G; \Psi]$ by

$$[\Theta G; \Psi]_{jk} = \frac{[\Theta GG^T\Theta^T]_{jk}}{\psi_j + \psi_k}.$$

Then

$$\Sigma = \kappa \Theta^T [\Theta G; \Psi] \Theta. \tag{34}$$

Note that the covariance matrix increases linearly with the gain parameter κ; as κ increases, the faster convergence to equilibrium described by relation (32) is at the cost of a greater spread at equilibrium.

Next we describe an example illustrating how a model of the form (33) might arise.

Shadow prices inferred from fluctuating flow rates. Consider the Poisson driving equation

$$d\mu_j(t) = \kappa \left(\sum_{r:j\in r} \varepsilon_r dN_r \left(\varepsilon_r^{-1} \int_0^t x_r(\tau) d\tau \right) - q_j(\mu_j(t))dt \right)$$

where $(N_r(\tau), \tau \geq 0)$, for $r \in R$, are a collection of independent unit rate Poisson processes. This would describe a model where, on a very fine time-scale, the flow on route r takes the form of a time-dependent Poisson process of rate $x_r(t)/\varepsilon_r$, with each point of the process containing a workload of size ε_r. The Brownian approximation, valid when ε_r are small, becomes

$$d\mu_j(t) = \kappa \left(\sum_{r:j\in r} \left(x_r(t)dt + \varepsilon_r^{1/2} x_r(t)^{1/2} dB_r(t) \right) - q_j(\mu_j(t))dt \right)$$

whose linearization is equation (33) where G is a $|J| \times |R|$ matrix with elements

$$G_{jr} = \varepsilon_r^{1/2} x_r^{1/2} A_{jr};$$

thus $GG^T = AXEA^T$ where $E = \mathrm{diag}(\varepsilon_r, r \in R)$.

Time lags

Consider next the system

$$\mu_j[t+1] = \mu_j[t] + \kappa \Big(\sum_{r:j\in r} x_r[t - d(j,r)] - q_j(\mu_j[t]) \Big) \tag{35}$$

where

$$x_r[t] = \frac{w_r}{\sum_{k\in r} \mu_k[t - d(k,r)]}. \tag{36}$$

A vector μ is an equilibrium point of the system (35)–(36) if $\mu_j[t] = \mu_j$, for $t = \ldots, 0, 1, 2, \ldots$, satisfies these equations.

Theorem 4 *The vector μ maximizing the strictly concave function $\mathcal{V}(\mu)$ is the unique equilibrium point of the system (35)–(36).*

Proof. The vector μ is an equilibrium point if and only if solves

$$\sum_{r:j\in r}\frac{w_r}{\sum_{k\in r}\mu_k}=q_j(\mu_j).$$

But this is precisely the stationarity condition implied by the partial derivatives (30) of the function $\mathcal{V}(\mu)$, a strictly concave function with a unique maximum. The result follows. □

Next we investigate the stability of the equilibrium point. Let $x_r = w_r/\sum_{k\in r}\mu_k$, and suppose q_j is differentiable at the point $y=\mu_j$, with derivative q_j'. Let $\mu_j[t]=\mu_j+\xi_j[t]$. Then, linearizing the system (35)–(36) about μ, we obtain

$$\xi_j[t+1]=\xi_j[t]-\kappa\Big(\sum_{r:j\in r}x_r^2w_r^{-1}\sum_{k\in r}\xi_k[t-d(j,r)-d(k,r)]+q_j'\xi_j[t]\Big). \tag{37}$$

Define the $|J|\times|J|$ matrices $(M[d], d=0,1,\ldots,D)$ by

$$(M[d])_{jk}=\sum_r x_rA_{jr}A_{kr}I[d(j,r)+d(k,r)=d]$$

where now $D=\max_{j,k,r}\{d(j,r)+d(k,r)\}$. Thus

$$\sum_{d=0}^{D}M[d]=AXW^{-1}XA^T.$$

Define the vector $\xi[t]=(\xi_j[t], j\in J)$. Then we can rewrite equation (37) in the matrix form

$$\begin{pmatrix}\xi[t+1]\\ \xi[t]\\ \vdots\\ \xi[t-D+1]\end{pmatrix}=M\begin{pmatrix}\xi[t]\\ \xi[t-1]\\ \vdots\\ \xi[t-D]\end{pmatrix}$$

where

$$M=\begin{pmatrix} I-\kappa(M[0]+Q') & -\kappa M[1] & -\kappa M[2] & \ldots & -\kappa M[D]\\ I & 0 & 0 & \ldots & 0\\ 0 & I & 0 & \ldots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \ldots & 0\end{pmatrix}. \tag{38}$$

The equilibrium point μ of the system (35)–(36) is stable if the spectral radius of the matrix M is less than unity. With stochastic effects, increasing the derivatives Q' has broadly the same reductive effect on the covariance matrix (34) as *decreasing* the gain parameter κ; in contrast the destabilising effect on the matrix (38) of increasing Q' is broadly the same as *increasing* κ.

Variants

Several variants of the primal algorithm (5)–(6) and the dual algorithm (9)–(10) allow a similar analysis. For example, if the right hand side of equation (5) is multiplied by a positive function $f_r(x(t), \mu(t))$ then Theorem 1 remains valid. Similarly, if the right hand side of equation (9) is multiplied by a positive function $f_j(x(t), \mu(t))$ then Theorem 3 remains valid. As a simple example, we could divide the right hand side of equation (5) by w_r, or of equation (9) by $q_j(\mu_j(t))$.

A more subtle variation is obtained if equation (9) is replaced by

$$\frac{d}{dt}\,\mu_j(t) = \kappa\Big(p_j\Big(\sum_{r:j\in r} x_r(t)\Big) - \mu_j(t)\Big), \tag{39}$$

where p_j is the inverse function of q_j, and $x_r(t)$ is again given by equation (10). Note that the expression (39) is of the same sign as expression (9), and so the proof of Theorem 3 goes through as before. Suppose $p_j(y)$ is differentiable at the stable point with derivative p'_j, and let $\mu_j(t) = \mu_j + \xi_j(t)$. Then, linearizing the system (39) about the equilibrium point, we obtain

$$\frac{d}{dt}\,\xi(t) = -\kappa(I + P'AXW^{-1}XA^T)\xi(t)$$

where $P' = \mathrm{diag}(p'_j, j \in J)$, allowing the local convergence properties of the algorithm (39) to be studied.

Examples

In this section we illustrate the results of the last two sections through a discussion of some examples. The first sub-section illustrates how the functions $p_j, j \in J$, may be determined by the detailed stochastic behaviour of resource j; a simple four node network is used to facilitate comparisons between feedback mechanisms. The results of this paper are, of course, intended to apply to large-scale networks, and our second sub-section discusses the behaviour of a dual algorithm in a random network with 100 resources and 1000 routes.

Congestion indication in a four node network

Suppose that the total load y on a resource takes the form, on a very fine time-scale, of a Poisson stream of cells at rate y/ε. Suppose that the time-axis is divided into non-overlapping slots each of length $\tau\varepsilon$, and that a feedback signal is generated for a slot if the total number of cells arriving in that slot exceeds a threshold N. (While there may well be a queue at a resource, we suppose for the moment that the feedback signals are generated by the process just described, rather than, for example, by the queue size exceeding a threshold.) Suppose that when a feedback signal is generated, it is sent to each user r whose route passes through resource j, where it is interpreted as a congestion indicator requiring a reduction in the rate $x_r(t)$ of size $\kappa\varepsilon x_r(t)$. If the probability that a signal is generated in any single slot is small, then this model corresponds to equation (20), with $\varepsilon_j = \varepsilon$ and

$$p_j(y) = \frac{1}{\tau} \sum_{n>N} e^{-y\tau} \frac{(y\tau)^n}{n!}\,. \tag{40}$$

Figure 1: A four-node network.

Consider the network illustrated in Figure 1, where $|J| = |R| = 4$. Let $w_r = 0.0002, r \in R$, suppose $p_j(y)$ is given by equation (40), and choose $N = 128, \tau = 50$, so that the equilibrium point is $x_r = 1.0, r \in R, \mu_j = 0.0001, j \in J$. Then, from relations (19) and (22), the covariance matrix of the rates $(x_r(t), r \in R)$ can be calculated to be

$$\Sigma_1 = \kappa\varepsilon 10^{-2} \begin{pmatrix} 2.4 & 0.8 & -0.8 & 0.8 \\ 0.8 & 2.4 & 0.8 & -0.8 \\ -0.8 & 0.8 & 2.4 & 0.8 \\ 0.8 & -0.8 & 0.8 & 2.4 \end{pmatrix},$$

a matrix whose form we shall discuss shortly.

For a second example, suppose again that a feedback signal is generated by a slot when the total number of cells arriving in that slot exceeds a threshold, N. But now suppose that when a feedback signal is generated at resource j, it is directed at a random route r with probability x_r/y (for example, the signal might be sent to the route responsible for the last cell arriving during the slot that generated the feedback signal). If user r receives a feedback signal, then the rate $x_r(t)$ is reduced by an amount $\kappa\varepsilon$. This model corresponds to equation (23), with $\varepsilon_j = \varepsilon$ and

$$p_j(y) = \frac{1}{\tau y} \sum_{n>N} e^{-y\tau} \frac{(y\tau)^n}{n!} \,. \tag{41}$$

Again let $w_r = 0.0002, r \in R$, and now choose $N = 125, \tau = 50$ so that, now using equation (41), $x_r = 1.0, r \in R, \mu_j = 0.0001, j \in J$, is once again an equilibrium point. Then, from relations (19) and (24), the covariance matrix of the rates $(x_r(t), r \in R)$ can be calculated to be

$$\Sigma_2 = \kappa\varepsilon 10^{-1} \begin{pmatrix} 1.5 & -1.2 & 1.1 & -1.2 \\ -1.2 & 1.5 & -1.2 & 1.1 \\ 1.1 & -1.2 & 1.5 & -1.2 \\ -1.2 & 1.1 & -1.2 & 1.5 \end{pmatrix} .$$

(The function (41) is *not* an increasing function of y for all values of y. But it is increasing at the point $y = 2$, hence the matrix (15) is positive definite and this implies that the equilibrium point is locally stable. We note in passing that function (41) provides an example where, if the parameters $w_r, r \in R$, are set too large, the function $\mathcal{U}(x)$ has no interior maximum and the system (5)–(6) has no equilibrium point).

It is interesting to compare the magnitudes and structures of the matrices Σ_1 and Σ_2. That Σ_2 is larger in magnitude is expected, since with individual feedback there are additional sources of variation in the random choice of which rate is to be reduced by a feedback signal. Note also that rates on routes sharing a node are positively correlated for the joint feedback model. The explanation is that, with joint feedback, congestion indication at a node causes both routes through that node to decrease their rates simultaneously. However, for individual feedback, routes sharing a node are negatively correlated. In this case, a decrease in the flow on a route will allow increases on routes sharing a common node with it.

We have simulated equations (20) and (23), with $\varepsilon = 1.0$, $\kappa = 0.01$, with results that agree well with the matrices Σ_1 and Σ_2.

Suppose next that when the number of cells within a slot exceeds N, each of the cells within the slot causes a feedback signal to be sent to the user responsible for that cell. Suppose a user responds to each feedback signal by reducing its rate by $\kappa\varepsilon$. Then the expected number of feedback signals generated per slot is

$$\sum_{n>N} n e^{-y\tau} \frac{(y\tau)^n}{n!} = y\tau \sum_{n\geq N} e^{-y\tau} \frac{(y\tau)^n}{n!} \,,$$

and so

$$p_j(y) = \sum_{n\geq N} e^{-y\tau} \frac{(y\tau)^n}{n!}, \tag{42}$$

an expression rather similar to the form (40). The covariance structure of the model depends on the size of N relative to the number of routes through a

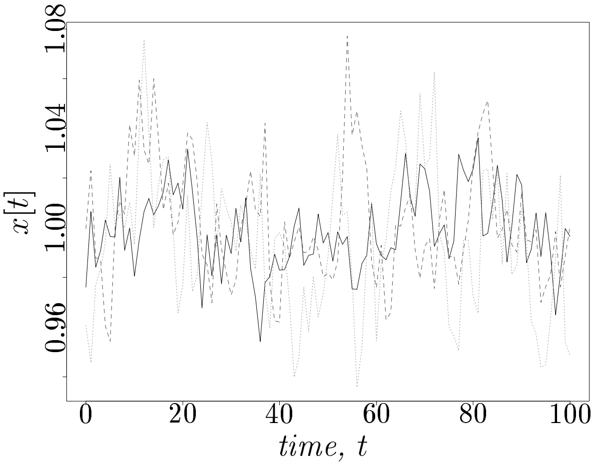


Figure 2: *Rates on three randomly chosen routes.*

resource. If N is large and most or all routes through a resource receive a feedback signal when overload occurs, then the covariance structure resembles that of the joint feedback model; if N is small, and an essentially random set of routes receives a feedback signal, the covariance structure resembles more closely that of the individual feedback model.

Many other mechanisms for signal generation are possible. For example, suppose that cells pass through a buffer which acts as a single server queue, and signals are generated whenever a cell arrives to find the buffer above a threshold level. Under Poisson arrival assumptions the rate of signal generation, and hence $p_j(y)$, may be determined from the analysis of an M/D/1 queue. Note that the buffer may well be a virtual buffer, with service rate lower than that of a real buffer at the resource, in order to signal congestion before the onset of cell loss [26].

Poisson streams and non-overlapping slots allow simple calculations, and are suggestive of the results that may be obtained with more complex models [12]. It is important to note, however, that the main results of earlier sections do not depend upon Poisson assumptions; the derivation of the covariance structure (19) was based on a more general central limit approximation, while Theorems 1 and 3 rely only upon rather weak properties of the functions p_j, q_j, $\mathcal{U}$ or $\mathcal{V}$.

A random network

Next we consider a network where the elements of the matrix A are independent random variables, each taking the value 1 with probability p and the value 0 otherwise, and where the elements of the matrix d are independent random variables uniformly distributed over the set $\{0, 1, \ldots, D\}$. Let $w_r = \sum_j A(j,r)$, and let

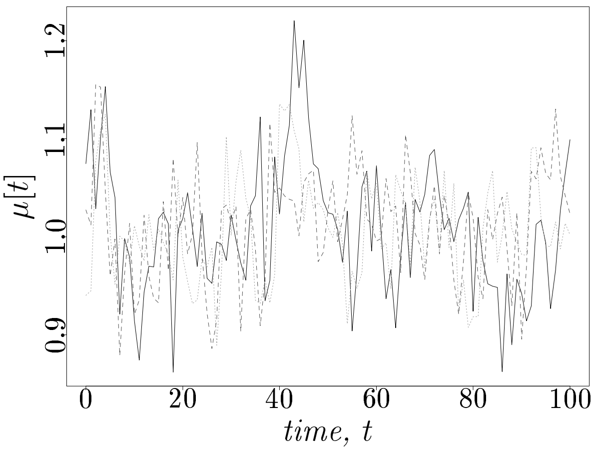


Figure 3: *Shadow prices for three randomly chosen resources.*

$q_j(\eta) = \eta \sum_s A(j,s)$, so that the unique stable point for the system (9)–(10) is $x_r = 1, r \in R, \mu_j = 1, j \in J$.

Consider the system

$$\mu_j[t+1] = \mu_j[t] + \kappa\Big(\sum_{r:j\in r} N_r[t-d(j,r)] - q_j(\mu_j[t])\Big) \tag{43}$$

where $(N_r[t], t = 1, 2, \ldots)$, for $r \in R$, are a collection of independent Poisson random variables, and $N_r[t]$ has mean $x_r[t]$ as defined by equation (36). The system (43) is thus a discrete time version of the dual algorithm that combines both stochastic fluctuations and time lags.

Figures 2 and 3 illustrate the behaviour of five randomly chosen routes, and five randomly chosen resources for the following parameter choices: $J = 100, R = 1000, p = 0.1, D = 10, \kappa = 0.005$. For these parameter choices the average length of a route is 10, the average number of routes through a resource is 100, and the largest time delay between a source and a resource is 10 time units. Note that for this example rates oscillate within a narrower band than shadow prices, and both are relatively well controlled.

In Figure 4, the curve labelled $a = 1$ records the effect of the gain parameter κ on the mean square deviation of shadow prices, σ^2, defined as the expected value of $(\mu_j(t) - 1)^2$ averaged over all resources, $j \in J$. For small values of κ, the relationship is approximately linear, with a slope in good agreement with that predicted by relation (34). However as κ increases, the mean square deviation diverges, with an asymptote at the value of κ (approximately 0.011) at which the spectral radius of the matrix (38) reaches unity and the deterministic time-lagged system becomes unstable.

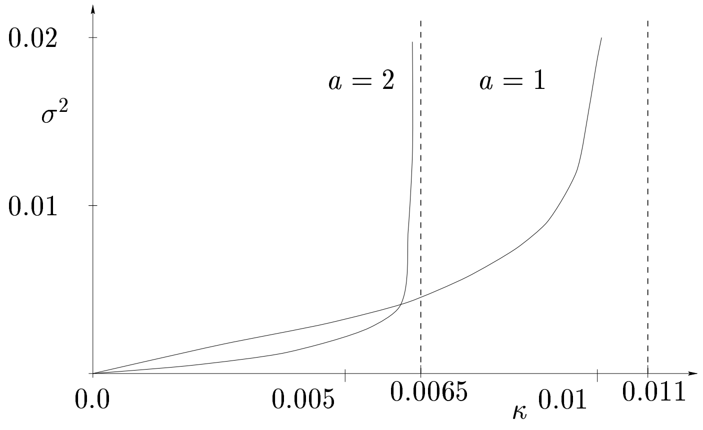


Figure 4: *Relation between the gain parameter κ and the mean square deviation of shadow prices, σ^2, for the resources of the random network. The parameter a labels the sensitivity of the relationship between flow rates and shadow prices at resources.*

Finally let us consider briefly the effect of more general choices for the functions $q_j, j \in J$, describing the relationship between flow rates and shadow prices at resources. Suppose that $q_j(\eta) = (a(\eta - 1) + 1) \sum_s A(j, s)$, so that the unique stable point for the system (9)–(10) is again $x_r = 1, r \in R, \mu_j = 1, j \in J$. The case $a = 1$ is that discussed so far; the case $a = 2$, also illustrated in Figure 4, corresponds to a doubling of the matrix of derivatives Q'. As predicted by our earlier analysis, the effect of increasing a is to reduce variability for smaller values of κ, but also to lower the critical value of κ at which the system becomes unstable.

User adaptation

In this section we consider the stability of systems where users are able to adapt very quickly to their experience of congestion, and illustrate briefly how our methods extend to this situation.

Suppose that user r is able to monitor its rate $x_r(t)$ continuously, and to vary smoothly the parameter $w_r(t)$ so as to track accurately the optimum to $USER_r(U_r; \lambda_r(t))$, where $\lambda_r(t) = w_r(t)/x_r(t)$ is the charge per unit flow to user r at time t . A simple differentiation establishes that the solution to the problem $USER_r(U_r; \lambda_r)$ has $w_r = x_r U_r'(x_r)$, where $x_r = w_r/\lambda_r$. Thus, under accurate tracking by user r of the optimum to $USER_r(U_r; \lambda_r(t))$, the parameter $w_r(t)$ will satisfy

$$w_r(t) = x_r(t) U_r'(x_r(t)), \tag{44}$$

while, for the primal algorithm, $x_r(t)$ evolves according to the revised differential equation

$$\frac{d}{dt} x_r(t) = \kappa\Big(w_r(t) - x_r(t) \sum_{j \in r} \mu_j(t)\Big)$$

where $\mu_j(t)$ is given by equation (6).

We shall establish stability of the revised system, by using a revision of the argument leading to Theorem 1. Consider the revised expression

$$\mathcal{U}(x) = \sum_{r \in R} U_r(x_r) - \sum_{j \in J} \int_0^{\sum_{s:j \in s} x_s} p_j(y) dy.$$

Note that

$$\frac{\partial}{\partial x_r} \mathcal{U}(x) = U_r'(x_r) - \sum_{j \in r} p_j \Big(\sum_{s:j \in s} x_s \Big),$$

and thus

$$\begin{aligned} \frac{d}{dt} \mathcal{U}(x(t)) &= \sum_{r \in R} \frac{\partial \mathcal{U}}{\partial x_r} \cdot \frac{d}{dt} x_r(t) \\ &= \kappa \sum_{r \in R} \frac{1}{x_r(t)} \Big(w_r(t) - x_r(t) \sum_{j \in r} p_j \Big(\sum_{s:j \in s} x_s(t) \Big) \Big)^2, \end{aligned}$$

using relation (44) to substitute for $U_r'(x_r(t))$. Hence $\mathcal{U}(x)$ provides a Lyapunov function for the revised system, and the unique value maximizing $\mathcal{U}(x)$ is a stable point of the system, to which all trajectories converge. Linearization may again be used to investigate behaviour near the stable point: for example, the revised form of equation (14) becomes

$$\frac{d}{dt} y(t) = -\kappa X^{1/2} (A^T P' A - U'') X^{1/2} y(t)$$

where $U'' = \text{diag}(U_r''(x_r), r \in R)$.

A similar analysis is possible for the dual algorithm. Under accurate tracking by user r of the optimum to $USER_r(U_r; \lambda_r(t))$ the parameter $w_r(t)$ will be given by equation (44), where, for the dual algorithm,

$$x_r(t) = \frac{w_r(t)}{\sum_{k \in r} \mu_k(t)}$$

and $\mu_j(t)$ evolves according to the differential equation (9). To find a Lyapunov function for this system, it is helpful to first construct the dual to problem $SYSTEM(U, A, C)$. Let $D_r(\lambda) = x_r$, where x_r is the solution to $\lambda = U_r'(x_r)$, with $D_r(\lambda) = 0$ if $\lambda \geq U_r'(0)$ and $D_r(\lambda) = \infty$ if $\lambda \leq U_r'(\infty)$. Then, after elision of a constant term, the dual of the problem $SYSTEM(U, A, C)$ becomes

$$\begin{array}{ll} \text{maximize} & \sum_{r\in R}\int^{\lambda_r} D_r(\zeta)d\zeta - \sum_{j\in J}\mu_j C_j \\ \text{subject to} & \lambda \le \mu^T A \\ \text{over} & \mu \ge 0 \end{array}$$

where the lower limit in the integral of the function D_r can be chosen to be any fixed value in the range $(U_r'(\infty), U_r'(0))$. We may interpret $D_r(\lambda)$ as the demand of user r when confronted with a price per unit flow of λ; under accurate tracking by user r

$$x_r(t) = D_r\Big(\sum_{j\in r}\mu_j(t)\Big). \tag{45}$$

Consider now the revised Lyapunov function

$$\mathcal{V}(\mu) = \sum_{r\in R}\int^{\sum_{j\in r}\mu_j} D_r(\zeta)d\zeta - \sum_{j\in J}\int_0^{\mu_j} q_j(\eta)d\eta.$$

Note that

$$\frac{\partial}{\partial\mu_j}\mathcal{V}(\mu) = \sum_{r:j\in r} D_r\Big(\sum_{k\in r}\mu_k\Big) - q_j(\mu_j)$$

and thus

$$\begin{aligned} \frac{d}{dt}\mathcal{V}(\mu(t)) &= \sum_{j\in J}\frac{\partial\mathcal{V}}{\partial\mu_j}\cdot\frac{d}{dt}\mu_j(t) \\ &= \kappa\sum_{j\in J}\Big(\sum_{r:j\in r} x_r(t) - q_j(\mu_j(t))\Big)^2, \end{aligned}$$

using equation (9) and relation (45). Hence $\mathcal{V}(\mu)$ provides a Lyapunov function for the revised system, and the unique value maximizing $\mathcal{V}(\mu)$ is a stable point of the system, to which all trajectories converge.

The models considered in this section assume very fast adaptation of the users, indeed so rapid that user r is essentially varying its rate $x_r(t)$ optimally in response to the resource shadow prices $(\mu_j(t), j \in J)$. Interesting questions remain concerning the stability of the system under more general assumptions on users' speed of adaptation.

A more general optimization problem

The optimization problem implicitly solved by the primal algorithm (5)–(6) is not our initial network problem $NETWORK(A, C; w)$, but rather the maximization of the Lyapunov function (7). We begin this section by discussing a possible interpretation of this relaxation of the network problem, where the constraint

$Ax \leq C$ is replaced by penalties, perhaps expressed in terms of delay or loss, that increase as the capacity of a resource is approached. Following this we indicate how the system problem $SYSTEM(U, A, C)$ may be recast both to motivate the relaxation of the network problem, and to allow routing choices.

Delay or loss

Suppose that when a resource is heavily loaded the network incurs some cost, perhaps expressed in terms of delay or loss. Then the optimization of the Lyapunov function (7) might be interpreted in terms of a penalty function

$$C_j(y) = \int_0^y p_j(\eta) d\eta \tag{46}$$

that describes the rate at which cost is incurred at resource j when the load through it is y.

For example, suppose the rate at which cost is incurred at resource j is

$$C_j(y) = \frac{1}{\tau} \sum_{n>N} (n-N) e^{-y\tau} \frac{(y\tau)^n}{n!}$$

when the load is y, that is, ε times the expected number of cells per unit time that exceed a threshold N in the slotted Poisson model of the previous section. Then a simple differentiation establishes that $p_j(y)$, determined by equation (46), is given by equation (42).

Routing

Next we extend the basic model to allow routing choices to be made. Let $s \in S$ now label a user, and suppose s is identified with a subset of R, the routes available to serve the user s. Set $H_{sr} = 1$ if $r \in s$, so that route r serves user s, and set $H_{sr} = 0$ otherwise. This defines a 0–1 matrix $H = (H_{sr}, s \in S, r \in R)$. For each $r \in R$ let $s(r)$ identify a value $s \in S$ such that $H_{sr} = 1$, and suppose this value is unique; view $s(r)$ as the user served by route r.

Now let y_r be the flow on route r, and suppose that resource j incurs a cost $C_j(\sum_{r:j\in r} y_r)$ dependent on the flow through that resource, where $C_j(\cdot)$ is a strictly convex function. Consider the following optimization problems.

$SYSTEM(U, H, A, C)$:

$$\begin{array}{ll} \text{maximize} & \sum_{s\in S} U_s(x_s) - \sum_{j\in J} C_j\Big(\sum_{r:j\in r} y_r\Big) \\ \text{subject to} & Hy = x \\ \text{over} & x, y \geq 0. \end{array}$$

$NETWORK(H, A, C; w)$:

$$\begin{array}{ll} \text{maximize} & \sum_{s \in S} w_s \log\Big(\sum_{r \in s} y_r\Big) - \sum_{j \in J} C_j\Big(\sum_{r:j \in r} y_r\Big) \\ \text{over} & y \geq 0. \end{array}$$

Then, following the approach of Kelly [18], it is possible to show that there exist vectors $\lambda = (\lambda_s, s \in S)$, $w = (w_s, s \in S)$ and $x = (x_s, s \in S)$ satisfying $w_s = \lambda_s x_s$ for $s \in S$, such that w_s solves $USER_s(U_s; \lambda_s)$ for $s \in S$ and x solves $NETWORK(H, A, C; w)$; further x is then the unique vector with the property that there exists a vector y such that (x, y) solves $SYSTEM(U, H, A, C)$.

Thus the relaxation of the network problem may be motivated by a similar relaxation of the overall system problem and both problems may be generalized to include routing choices. Finally we sketch the natural generalisations of the primal and dual algorithms, and their corresponding Lyapunov functions.

Suppose that p_j and C_j are related by equation (46). Generalize the primal algorithm (5)–(6) to become

$$\frac{d}{dt} y_r(t) = \kappa\Big(w_{s(r)} - \Big(\sum_{a \in s(r)} y_a(t)\Big) \sum_{j \in r} \mu_j(t)\Big) \tag{47}$$

(or zero if this expression is negative and $y_r(t) = 0$) where

$$\mu_j(t) = p_j\Big(\sum_{r:j \in r} y_r(t)\Big), \tag{48}$$

and let

$$\mathcal{U}(y) = \sum_{s \in S} w_s \log\Big(\sum_{r \in s} y_r\Big) - \sum_{j \in J} \int_0^{\sum_{r:j \in r} y_r} p_j(y) dy.$$

Then the dynamical system (47)–(48) has the property that

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

unless y solves $NETWORK(H, A, C; w)$.

Similarly the dual algorithm (9)–(10) may be generalized to incorporate a form of least cost routing. For $s \in S$ let

$$\sum_{r \in s} y_r(t) = x_s(t) = \frac{w_s}{\min_{r \in s} \sum_{j \in r} \mu_j(t)},$$

and suppose $y_r(t)$ is only positive on routes r that attain the minimum in the denominator. Then the dynamical system

$$\frac{d}{dt}\mu_j(t) = \kappa\Big(\sum_{r:j \in r} y_r(t) - q_j\big(\mu_j(t)\big)\Big)$$

has the property that $\mathcal{V}(\mu(t))$ is an increasing function of t, where

$$\mathcal{V}(\mu) = \sum_{s \in S} w_s \log\Big(\min_{r \in s} \sum_{j \in r} \mu_j\Big) - \sum_{j \in J} \int_0^{\mu_j} q_j(\eta) d\eta.$$

Thus routing, as well as rate control, may be naturally integrated with proportionally fair pricing.

Concluding remarks

In this paper we have addressed the issue of how available bandwidth within a large-scale broadband network should be shared between competing streams of elastic traffic. An optimization framework leads to a decomposition of the overall system problem into a separate problem for each user, in which the user chooses a charge per unit time that the user is willing to pay, and one for the network; we have shown that two classes of rate control algorithm are naturally associated with the objective functions appearing in, respectively, the primal and dual formulation of the network's problem. In consequence the algorithms provide natural implementations of proportionally fair pricing. We have studied the stability of the algorithms in the presence of stochastic perturbations and time lags, and have illustrated our results with a study of a random network with a hundred resources and a thousand routes. Interesting and challenging questions remain concerning the stability of the entire system under more general assumptions on users' reactions to the rates allocated to them by the network, and when the numbers of users and the amounts of capacity available for elastic traffic vary randomly. An outstanding practical issue concerns how protocols, such as TCP in the Internet or the Available Bit Rate transfer capability of an ATM network, can be adapted to be charge sensitive.

Acknowledgements: Aman Maulloo was supported by the Staff Development Programme, University of Mauritius, and David Tan by St John's College, Cambridge; computing equipment was provided by the EPSRC under grant GR/J31896, and further support was provided by the Commission of the European Communities ACTS project AC039, entitled Charging and Accounting Schemes in Multiservice ATM Networks (CA$hMAN). The authors are grateful to Martin Biddiscombe, Costas Courcoubetis, Jon Crowcroft, Richard Gibbens, James Giles, Meena Lakshmanan, Ravi Mazumdar, Andrew Odlyzko, George D. Stamoulis and Richard Weber for valuable discussions and suggestions.

References

[1] Shenker S (1995). Fundamental design issues for the future Internet. *IEEE Journal Selected Areas Communication*, **13**, 1176–1188.

[2] Jacobson V (1988). Congestion avoidance and control. In *Proc. ACM SIGCOMM '88*, 314–329. An updated version is available via ftp://ftp.ee.lbl.gov/papers/congavoid.ps.Z .

[3] Wroclawski J (1996). Specification of the controlled-load network element service. Internet–Draft. (Internet-Drafts are working documents of the Internet Engineering Task Force - their current status is available by ftp from nic.nordu.net .)

[4] International Telecommunications Union (1996). Recommendation I371: Traffic control and congestion control in B–ISBN. Geneva.

[5] Kelly FP (1996). Modelling communication networks, present and future. *Phil. Trans. R. Soc. Lond. A*, **354**, 437–463.

[6] Hernandez-Valencia EJ, Benmohamed L, Nagarajan R and Chong S (1997). Rate control algorithms for the ATM ABR service. *European Transactions on Telecommunications*, **8**, 7–20.

[7] Chiu D-M and Jain R (1989). Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN Systems*, **17**, 1–14.

[8] Bolot J-C and Shankar AU (1990). Dynamic behavior of rate-based flow control mechanisms. *ACM SIGCOMM Computer Commun. Review*, **20**, 35–49.

[9] Fendick KW, Rodrigues MA and Weiss A (1992). Analysis of a rate-based feedback control strategy for long-haul data transport. *Performance Evaluation*, **16**, 67–84.

[10] Bonomi F, Mitra D and Seery JB (1995). Adaptive algorithms for feedback-based flow control in high-speed wide-area networks. *IEEE J. Selected Areas Communications*, **13**, 1267–1283.

[11] Courcoubetis C and Weber RR (1996). Buffer overflow asymptotics for a switch handling many traffic sources. *Journal Applied Probability*, **33**, 886-903.

[12] Kelly FP (1996). Notes on effective bandwidths. In *Stochastic Networks: Theory and Applications* (ed. F.P. Kelly, S. Zachary and I. Ziedins). Volume 4 of *Royal Statistical Society Lecture Notes Series*, pp. 141–168. Oxford University Press, Oxford.

[13] Charny A, Ramakrishnan KK and Lauck A (1996). Time scale analysis and scalability issues for explicit rate allocation in ATM networks. *IEEE/ACM Transactions on Networking*, **4**, 569–581.

[14] Mazumdar R, Mason LG and Douligeris C (1991). Fairness in network optimal flow control: optimality of product forms. *IEEE Transactions on Communications*, **39**, 775–782.

[15] Kelly FP (1994). On tariffs, policing and admission control for multiservice networks. *Operations Research Letters*, **15**, 1–9.

[16] MacKie-Mason JK and Varian HR (1994). Pricing the Internet. In *Public Access to the Internet* (ed. B. Kahin and J. Keller). Prentice-Hall, Englewood Cliffs, New Jersey.

[17] McKnight LW and Bailey JP (1997). *Internet Economics*. MIT Press, Cambridge, Massachussetts.

[18] Kelly FP (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, **8**, 33–37. An updated version is available at http://www.statslab.cam.ac.uk/ frank/elastic.html .

[19] Arrow KJ and Hahn FH (1971). *General Competitive Analysis*. Oliver and Boyd, Edinburgh.

[20] Hahn F (1982). Stability. In *Handbook of Mathematical Economics*, Volume II (ed. K.J. Arrow and M.D. Intriligator). pp. 745–793. North-Holland, Amsterdam.

[21] Varian HR (1992). *Microeconomic Analysis*. Third edition. Norton, New York.

[22] Gardner R(1995). *Games for Business and Economics*. Wiley, New York.

[23] Whittle P (1971). *Optimization Under Constraints*. Wiley, Chichester.

[24] Courcoubetis C, Siris VA and Stamoulis GD (1996). Integration of pricing and flow control for Available Bit Rate services in ATM networks. *Proc. IEEE Globecom '96*, London.

[25] Miller RK and Michell AN (1982). *Ordinary Differential Equations*. Academic Press, New York.

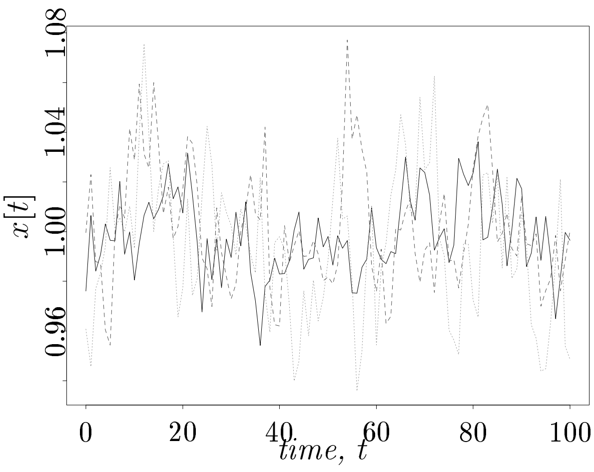
[26] Courcoubetis C, Kesidis G, Ridder A, Walrand J and Weber RR (1995). Admission control and routing in ATM networks using inferences from measured buffer occupancy. *IEEE Transactions on Communications*, **43**, 1778–1784.

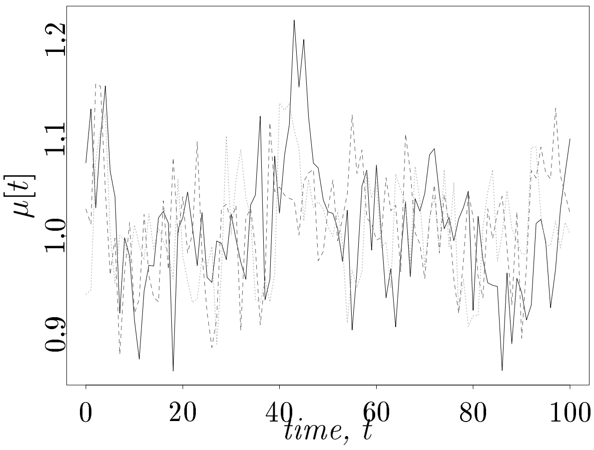
Figure 1: A four node network.

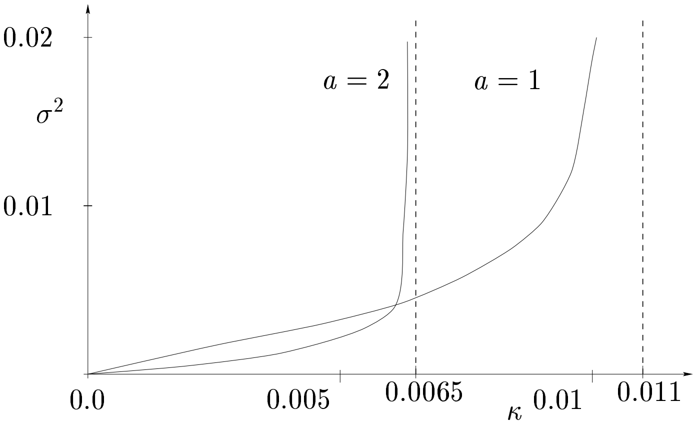
Figure 2: Rates on three randomly chosen routes.

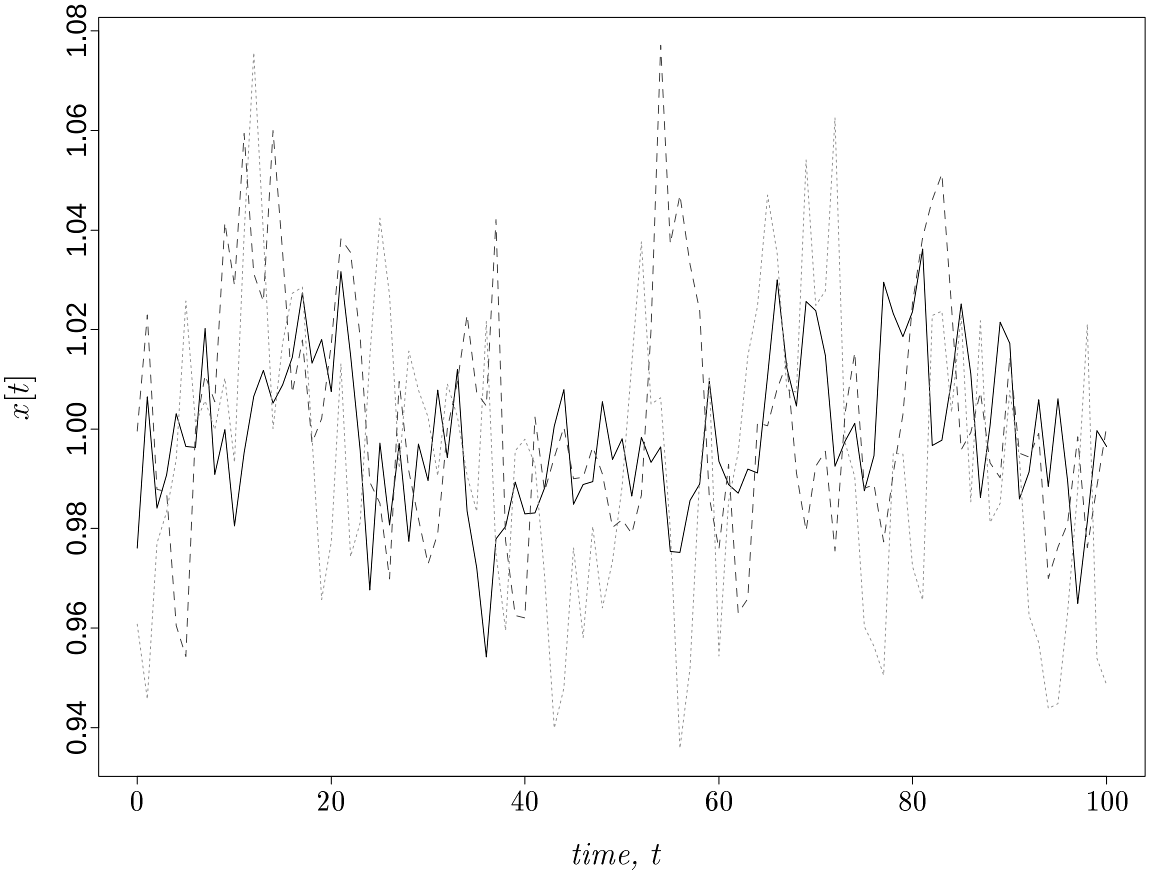
Figure 3: Shadow prices for three randomly chosen resources.

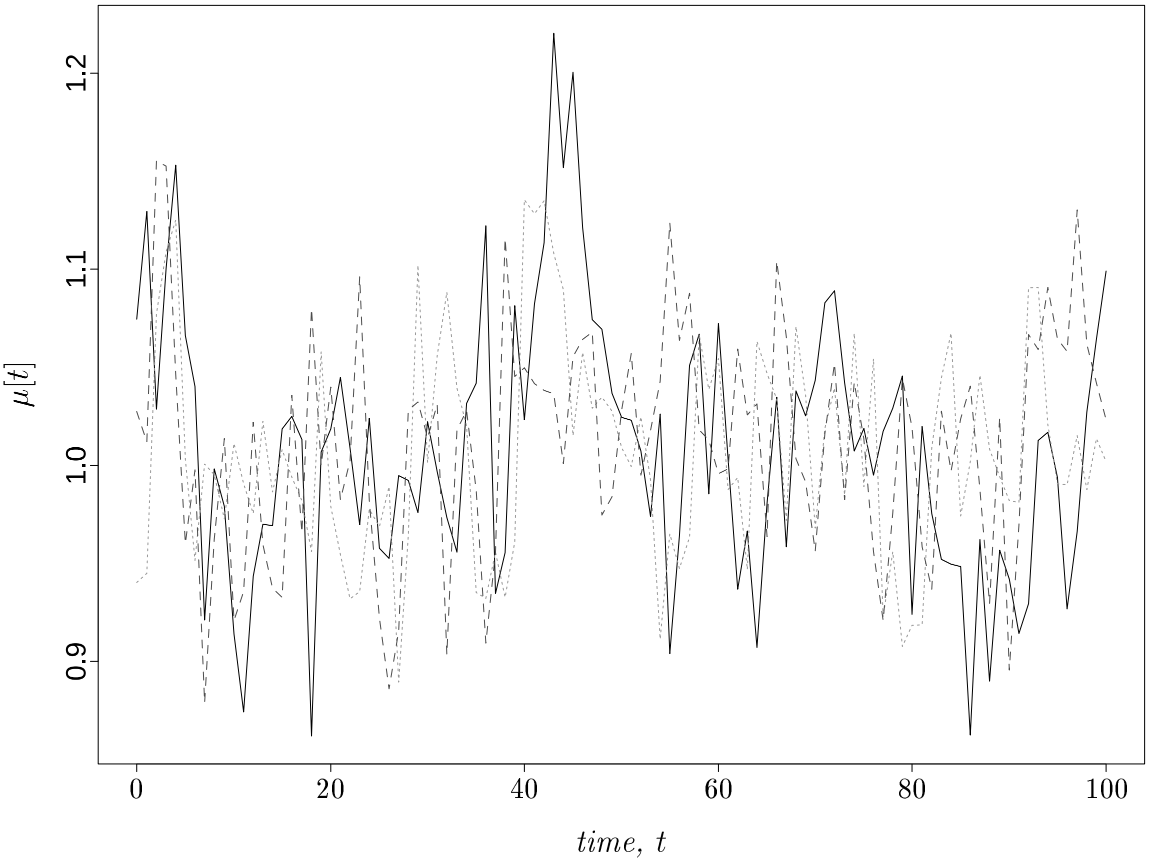
Figure 4: Relation between the gain parameter κ and the mean square deviation of shadow prices, σ^2, for the resources of the random network. The parameter a labels the sensitivity of the relationship between flow rates and shadow prices at resources.

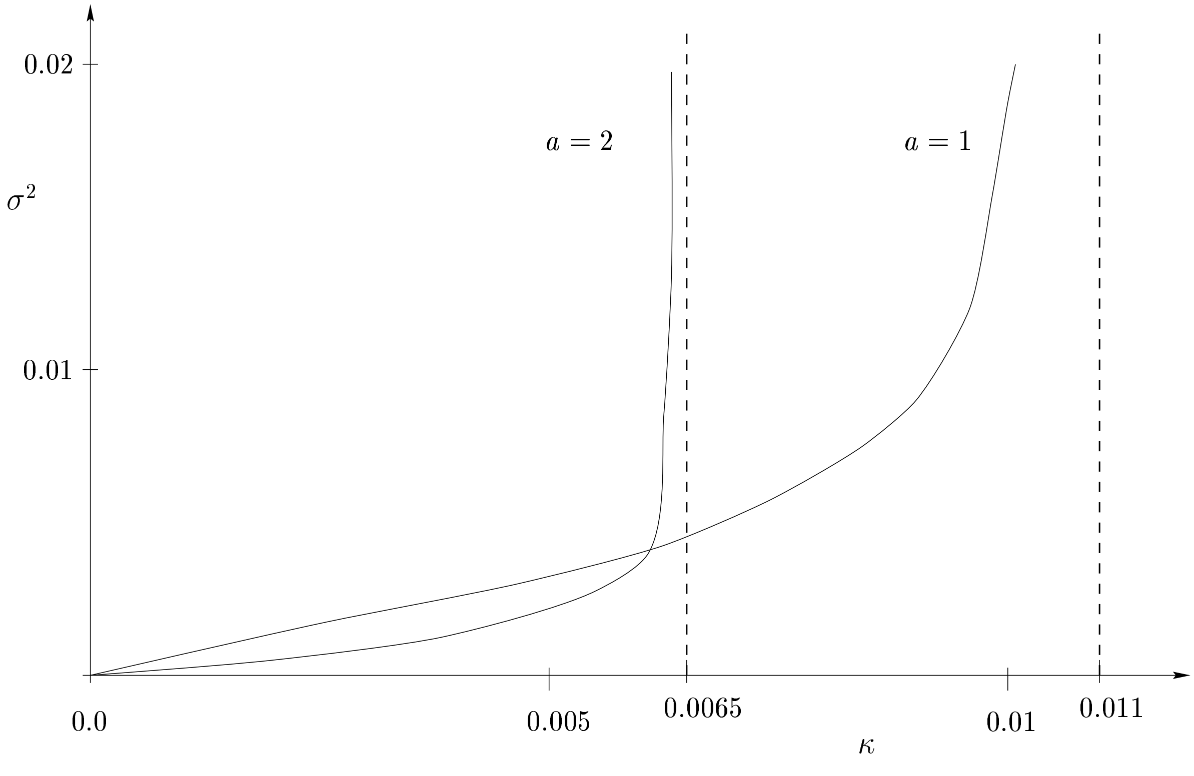

1.08
1.04
1.00
0.96
x[t]
0
20
40
60
80
100
time, t


1.2
1.1
1.0
0.9
$\mu[t]$
0
20
40
60
80
100
time, t


0.02
0.01
σ^2
$a = 2$
$a = 1$
0.0
0.005
0.0065
κ
0.01
0.011


$x[t]$
1.08
1.06
1.04
1.02
1.00
0.98
0.96
0.94
0
20
40
60
80
100
$time,\ t$


1.2
1.1
1.0
0.9
$\mu[t]$
0
20
40
60
80
100
time, t


0.02
0.01
σ^2
$a = 2$
$a = 1$
0.0
0.005
0.0065
0.01
0.011
κ