TECHNICAL RESEARCH REPORT

Stability of rate control system with time-varying communication delays

by Richard J. La, Priya Ranjan

CSHCN TR 2004-16
(ISR TR 2004-31)

Stability of Rate Control System with Time-Varying Communication Delays

Richard J. La and Priya Ranjan

***Abstract*— We adopt the optimization framework for the rate allocation problem proposed by Kelly and investigate the stability of the system with *arbitrary* communication delays between network elements with time-varying queue dynamics. We first present the conditions for the existence of a solution of the system. Second, we establish the conditions for the system stability with arbitrary delays for a family of popular utility and price functions, and then extend the results to more general utility and price functions. We demonstrate that the stability of such a system can be studied by considering a discrete time system derived from a simpler homogeneous delay case where all users have the same fixed delay. Numerical examples are provided to validate our analyses.**

***Index Terms*— Control theory, economics.**

I. Introduction

Kelly [10] has proposed an optimization framework for the rate allocation for elastic traffic where the objective of the system is to maximize the aggregate utility of the users. Here the utility of a user could either represent the true utility or preferences of the user or a utility function that is assigned to the user by the end user rate control algorithms, *e.g.,* Transmission Control Protocol (TCP) and Proportional-Fair Congestion Controller (PFCC). In the latter case the selection of the utility function determines the end user algorithms and the trade-off between the fairness among the users and system efficiency [1], [11], [15], [19], *e.g.,* PFCC. Using the proposed framework he has shown that the system optimum is achieved at the equilibrium between the end users and resources. Based on this observation researchers have proposed various rate-based algorithms, in conjunction with a variety of active queue management (AQM) mechanisms, that solve the system optimization problem or its relaxation [10], [14], [15], [16].

The convergence of most of these algorithms, however, has been established only in the absence of feedback delay initially. Modeling the communication delay is especially important when the delay is non-negligible and/or the delay could be widely varying, *e.g.,* multi-hop mobile wireless network. Tan and Johari [9] have studied the case with homogeneous users, *i.e.,* same round-trip delays and same log utility functions, and provided local stability conditions in term of users' gain parameters and communication delays. In general their results state that the product of gain parameter and communication delays should be no larger than some constant. Similar results have been obtained in [4], [18] in the context of single flow and single resource with more general utility functions and in [2] in the context of single bottleneck with multiple heterogeneous users. The given stability conditions are similar to those in [9] and state that the product of the delay and gain parameter of end user algorithms needs to be smaller than some constant. These results focus on characterizing *sufficient* conditions on the communication delay and gain parameter for stability.

Ranjan *et al.* [20] and Ying *et al.* [25] have studied the stability of the rate control system in the presence of arbitrary *fixed* delays between network elements, *e.g.,* network resources and end users, and arbitrary gain parameters of the end users. These approaches are consistent with the philosophy that network protocols must be *simple* and *robust* given the complexity and scale of the Internet, and may prove to be more suitable for wireless ad hoc networks where delays are often expected to be unpredictable and widely varying and satellite networks in which the delays are significant. In particular, the stability conditions given in [2], [4] may be sufficient to ensure smooth operation of the network when the network is operating normally. However, a network can occasionally experience high congestion and behave unpredictably due to the presence of a large amount of nonresponsive traffic, *e.g.,* broadcast of a concert on-line, and/or a collapse of a part of network as a result of, for example, a link failure or routing instability. In such a scenario the system may temporarily deviate from the stable regime characterized by the conditions in [2], [4] because of the increased queueing delay and/or a larger number of flows, and the unstable behavior of the rate control mechanism can aggravate

R. J. La and P. Ranjan are with the Department of ECE and ISR, University of Maryland, College, MD 20742. E-mail: {hyongla,priya}@isr.umd.edu.

the congestion level, potentially leading to a congestion collapse. A more detailed discussion of these results is presented in Section III.

In this paper we extend the results in [20] and consider the case of state-dependent time-varying delays between network resources and end users and establish the conditions on utility and resource price functions that ensure global stability with arbitrary delays with queue dynamics. Modeling time-varying delays is important when queueing delays are comparable to the fixed delays and/or the delays are expected to vary widely due to a large variation in queueing delays (*e.g.*, multi-hop wireless networks). Our analysis is based on the invariance-based global stability results for nonlinear delay differential equations [7], [8], [17]. This kind of global stability results are different from those based on Lyapunov or Razumikhin theorems for delay differential equations used in [2], [4], [18], [23], [25] or from passivity approach [24]. A simple case of single-resource has been studied in our earlier work [21].

The main results of this paper can be summarized as follows:

1) The stability conditions for a rate control system with a popular family of users' utility and resource price functions are derived (Section V). These stability results are extended to more general utility and price functions under a set of assumptions (Section VI).
2) The rate control system with arbitrary time-varying delays is stable *if* the same system with a *fixed* homogeneous delay [20] is stable with appropriate initial conditions (Section VII). In other words, the detailed delay structure/dynamics between network elements and end users are not critical to the stability of the system. Combined with the results in [20] this implies that the stability of the system can be studied by looking at a simple discrete time system that arises from the underlying market structure of the rate control problem.

This paper is organized as follows. Section II describes the optimization problem for rate control. An overview of the previous work on characterizing stability conditions in the presence of a communication delay is provided in Section III. Section IV describes the system model with time-varying communication delays, and provides conditions for the existence of a solution. The stability conditions with a popular family of utility and resource price functions are derived in Section V. The stability conditions in Section V are extended to the cases with more general utility and price functions in Section VI. A relationship between the stability conditions for the time-varying delay case studied in this paper and those of a simple fixed homogeneous delay case is discussed in Section VII. Numerical examples are presented in Section VIII. We conclude in Section IX.

II. Background

In this section we briefly describe the rate control problem in the proposed optimization framework. Consider a network with a set $\mathcal{L}$ of resources or links and a set $\mathcal{I}$ of users. Let C_l denote the finite capacity of link $l \in \mathcal{L}$. Each user has a fixed route r_i, which is a non-empty subset of $\mathcal{L}$. We define a zero-one matrix A, where $A_{i,l} = 1$ if link l is in user i's route r_i and $A_{i,l} = 0$ otherwise. When the throughput of user i is x_i, user i receives utility $U_i(x_i)$. As mentioned earlier, this utility function could represent either the user's true utility or some function assigned to the user through the selected end user algorithms. We take the latter view and assume that the utility functions of the users are used to select the desired rate allocation among the users (*i.e.*, the desired operating point of the system), which also determines the end user algorithms as will be shown shortly. The utility $U_i(x_i)$ is an increasing, strictly concave and continuously differentiable function of x_i over the range $x_i \geq 0$.[1] Furthermore, the utilities are additive so that the aggregate utility of rate allocation $x = (x_i, i \in \mathcal{I})$ is $\sum_{i\in\mathcal{I}} U_i(x_i)$. Let $U = (U_i(\cdot), i \in \mathcal{I})$ and $C = (C_l, l \in \mathcal{L})$.[2] The rate control problem can be formulated as the following optimization problem [10]:
SYSTEM(U,A,C):

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in\mathcal{I}} U_i(x_i) \\ \text{subject to} \quad & A^T x \leq C \ , \ x \geq 0 \end{aligned} \tag{1}$$

The first constraint in the problem says that the total rate through a resource cannot be larger than the capacity of the resource. Instead of solving (1) directly, which is difficult for any large network, Kelly in [10] has proposed to consider the following two simpler problems.

Suppose that each user i is given the price per unit flow λ_i. Given λ_i, user i selects an amount to pay per unit time, w_i, and receives a rate $x_i = \frac{w_i}{\lambda_i}$.[3] Then, the user's optimization problem becomes the following [10].
$USER_i(U_i; \lambda_i)$:

$$\begin{aligned} \text{maximize} \quad & U_i\left(\frac{w_i}{\lambda_i}\right) - w_i \\ \text{over} \quad & w_i \geq 0 \end{aligned} \tag{2}$$

[1] Such a user is said to have elastic traffic.
[2] All vectors are assumed to be column vectors.
[3] This is equivalent to selecting its rate x_i and agreeing to pay $w_i = x_i \cdot \lambda_i$.

The network, on the other hand, given the amounts the users are willing to pay, $w = (w_i, i \in \mathcal{I})$, attempts to maximize the sum of weighted log functions $\sum_{i\in\mathcal{I}} w_i \log(x_i)$. Then the network's optimization problem can be written as follows [10].

NETWORK(A,C;w) :

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in\mathcal{I}} w_i \log(x_i) \\ \text{subject to} \quad & A^T x \le C \ , \ x \ge 0 \end{aligned} \tag{3}$$

Note that the network does not require the true utility functions $(U_i(\cdot), i \in \mathcal{I})$, and pretends that user i's utility function is $w_i \cdot \log(x_i)$ to carry out the computation. It is shown in [10] that one can always find vectors $\lambda^\star = (\lambda_i^\star, i \in \mathcal{I}), w^\star = (w_i^\star, i \in \mathcal{I})$, and $x^\star = (x_i^\star, i \in \mathcal{I})$ such that $w_i^\star$ solves $USER_i(U_i; \lambda_i^\star)$ for all $i \in \mathcal{I}$, $x^\star$ solves $NETWORK(A, C; w^\star)$, and $w_i^\star = x_i^\star \cdot \lambda_i^\star$ for all $i \in \mathcal{I}$. Furthermore, the rate allocation $x^\star$ is also the unique solution to $SYSTEM(U, A, C)$.

Assume that every user adopts rate-based flow control. Let $w_i(t)$ and $x_i(t)$ denote user i's willingness to pay per unit time and rate at time t, respectively.[4] Now suppose that at time t each resource $l \in \mathcal{L}$ charges a price per unit flow of $\mu_l(t) = p_l(\sum_{i:l\in r_i} x_i(t))$, where $p_l(\cdot)$ is an increasing function of the total rate going through it. Consider the system of differential equations

$$\frac{d}{dt} x_i(t) = \kappa_i \left(w_i(t) - x_i(t) \sum_{l\in r_i} \mu_l(t) \right) \ . \tag{4}$$

These equations can be motivated as follows. Each user first computes a price per unit time it is willing to pay, namely $w_i(t)$. Then, it adjusts its rate based on the feedback provided by the resources in the network to equalize its willingness to pay and the total price. Following [11] we assume $w_i(t)$ is set to $x_i(t) \cdot U_i'(x_i(t))$. With this choice of willingness to pay $w_i(t)$ one can see that user i always tries to reach a point where $U_i'(x_i) = \sum_{l\in r_i} p_l(\sum_{j:l\in r_j} x_j)$, *i.e.*, the marginal utility from additional rate, equals the price per unit flow, which maximizes the net utility which is the utility minus the total price. This is in fact the solution to the user optimization problem in (2) [22]. The feedback signal from a resource $l \in \mathcal{L}$ can also be interpreted as a congestion indicator, requiring a reduction in the flow rates going through the resource. For more detailed explanation of (4), refer to [11]. Since we assume that the utility functions of the users are selected to decide the rate allocation amongst the users, under (4) one can see that, in fact, both the users' utility functions and resource price functions can be utilized to decide the operating point of the system. Therefore, the design of rate control algorithms is equivalent to selecting the users' utility functions and the price functions of the resources in the network.

[4]Throughout the rest of the paper we refer to the willingness to pay per unit time as simply willingness to pay.

Kelly *et al.* [11] have shown that under some conditions on $p_l(\cdot), l \in \mathcal{L}$, the above system of differential equations converges to a point that maximizes the following expression

$$\mathcal{U}(x) = \sum_i U_i(x_i) - \sum_l \int_0^{\sum_{i:l\in r_i} x_i} p_l(y)\, dy. \tag{5}$$

Note that the first term in (5) is the objective function in our *SYSTEM(U, A, C)* problem. Thus, the algorithms proposed by Kelly *et al.* solves a relaxation of the $SYSTEM(U, A, C)$ problem.

III. Previous Work

The analysis in [11] of the convergence of the rate control algorithms given by (4), however, does not model the communication delay that is present between the resources and the end users. There has been some previous work on studying the stability of the system in the presence of communication delay. Here we only briefly summarize some of work that is most relevant to our work presented in this paper. Tan and Johari [9] have analyzed the case where every user has the same round-trip delay and utility function given by $w \log(\cdot)$, *i.e.*, $w(t) = w$. They have characterized the conditions on local stability in terms of the gain parameter κ and and communication delay d. Their results state that there exists some constant D such that the product of the gain parameter κ and communication delay d should be smaller than D. In addition, they have shown the convergence rate of the system in the case of single-user single-resource.

A similar sufficient condition is also obtained in [18] in the context of single-flow single-resource. Suppose that the end user algorithms is given by

$$\dot{x}(t) = \kappa(w - x(t-d)p(x(t-d))) \ ,$$

where $w, \kappa > 0$ and d is the communication delay. This models the end user algorithms with $U(x) = w \cdot \log(x)$ with a feedback delay of d. The authors show that, if $p(\cdot)$ is a function of class $\mathcal{C}^1$ that is nonnegative, nondecreasing, and bounded in norm by 1 such that $p'(\cdot)$ is nonincreasing and $\lim_{x\to\infty} p(x) = 1$, then the system is stable provided that $0 \le d \le \frac{1}{4\kappa}$.

Recently Deb and Srikant [4] have investigated the stability of the system in the context of single flow and single resource with more general utility functions, and have provided a *sufficient* condition for stability. Assume that the rate x is constrained to $[l, M]$. Let d denote the feedback delay from the network resource to the single user. The resource price function is denoted by $p(\cdot)$, and $x^\star$ is the unique solution to (5). The proposed end user algorithms is given by

$$\dot{x}(t) = \kappa \left(w - \frac{1}{x(t)U'(x(t))} x(t-d)p(x(t-d)) \right) .$$

Define

$$A(l, M) = 1 + \frac{w \min_{l \le x \le M}(-xU''(x) - U'(x))}{\max_{l \le x \le M} h(x)} \quad (6)$$

$$B(l, M) \quad (7)$$
$$= \frac{w \max_{l \le x \le M} \left| xU''(x) + U'(x) \right| + \max_{l \le x \le M} h(x)}{\min_{l \le x \le M} xU'(x)}$$

where $h(x(t-d)) = \lim_{y \to x(t-d)} \frac{yp(y) - x^\star p(x^\star)}{y - x^\star}$. Their main results state that if there exists some constant $q > 1$ such that $\sqrt{q}\kappa d < \frac{A(l,M)}{B(l,M)}$, then the system is globally exponentially stable.

Alpcan and Basar [2] have studied the stability of a system with a single resource and multiple flows, using a delay based algorithms, and provided a *sufficient* condition for stability. Although the algorithms uses the estimated queueing delay as the feedback information, the authors assume that feedback delay is fixed. Denote the feedback delay of flow i by r_i. Assume $x^\star$ is the solution to (5) and $0 \le x_i \le x_{i,max}$, where $x_{i,max}$ is assumed not to exceed the minimum capacity of the links on the user's route. Let $\tilde{x} := x - x^\star$ and $g_i(\tilde{x}) := \frac{dU_i(x)}{dx} - \frac{dU_i(x^\star)}{dx}$. Define

$$k_{min} := \min_i \inf_{-x_i^\star \le \tilde{x}_i \le x_{i,max} - x_i^\star} \left| \frac{g(\tilde{x}_i)}{\tilde{x}_i} \right| . \quad (8)$$

The end user's rate evolves according to

$$\dot{x}_i(t) = \frac{dU_i(x_i(t))}{dx_i} - \alpha_i q(\underline{x}(t - r_i)) \ , \ i = 1, \ldots, I \ ,$$
$$\dot{q}(t) = \frac{\sum_{i=1}^{I} x_i(t - r_i)}{C} - 1 \ ,$$

where q is the queueing delay and $\alpha_i > 0$. They show that if $r_{max} := \max_i r_i < \frac{k_{min} C}{2I\alpha_{max}}$, then the system is asymptotically stable.

Ranjan *et al.* [20] studied the global stability of the system with a class of utility and resource price functions. These utility and resource prices functions are also considered in this paper and are described in more details in Section V. Suppose that the rate update rules of the users are given by

$$\frac{d}{dt} x_i(t) = \kappa_i \left(\frac{1}{x_i^{a_i}(t)} - x_i(t - T_i)\mu^i(t) \right) \quad (9)$$

where

$$\mu^i(t) = \sum_{l \in r_i} \left(\frac{\sum_{j \in I_l} x_j(t - Z_{j,l} - T_{i,l})}{C_l} \right)^{b_l} ,$$

$Z_{j,l}$ is the forward delay from sender j to resource l, $T_{j,l}$ is the reverse delay from resource l to sender j, T_j is the round-trip delay of user j, *i.e.*, $T_j = Z_{j,l} + T_{j,l}$, and I_l is the set of users traversing resource l. Here the parameters $a_i, i \in \mathcal{I}$, and $b_l, l \in \mathcal{L}$, determine the responsiveness of the users and resource price functions, respectively. They have shown that if $a_i > 1 + \max_{l \in r_i} b_l$ for all $i \in \mathcal{I}$, then the system is globally asymptotically stable regardless of $T_{i,l}, Z_{i,l}$, and T_i. A similar result with homogeneous users and resource price functions has been obtained by Ying *et al.* [25].

IV. Network Model with Delays

Some of previous studies [2], [4], [18] described in Section III have modeled the delays between the network resources and end users. However, in all of these studies the delays are assumed to be fixed. In a network, however, the delays experienced by the packets depend on the queue sizes at the bottlenecks at the time of arrival. Modeling queueing delays may be more important in wireless networks where the capacities are limited. In this section we first describe the network model that captures the delays between the network resources and end users with queue dynamics. A simple case of single-flow, single-link with state-dependent time-varying delay is considered in [21].

Consider a set $\mathcal{I} = \{1, \ldots, N\}$ of users sharing a network consisting of a set $\mathcal{L} = \{1, \ldots, L\}$ of resources as described in Section II. The route of user i is given by r_i. In this paper we focus only on the forward path. In other words, the route r_i consists only of the resources in the forward path. Although the reverse path can be modeled in a similar manner as the forward path, for notational simplicity, we model the reverse path as a single link with a fixed delay given by $T_i^r, i \in \mathcal{I}$. Let I_l be the set of users traversing resource $l \in \mathcal{L}$, *i.e.*, $I_l = \{i \in \mathcal{I} \mid l \in r_i\}$. The feedback information from the resources to user i, which is typically carried by acknowledgments (ACKs), is delayed due to link propagation and transmission delays as well as the queueing delays at the bottlenecks.

Let D_l, $l \in \mathcal{L}$, denote the transmission and propagation delay of resource l. We assume that the links in $r_i = \{l_{(i,1)}, \ldots, l_{(i,R_i)}\}$ are arranged in the order user i packets visit, where $R_i = |r_i|$, and let $l_{(i,k)}$ denote the k-th link along r_i. When used, $(R_i + 1)$-th resource in user i's route refers to receiver i. We first define the following:

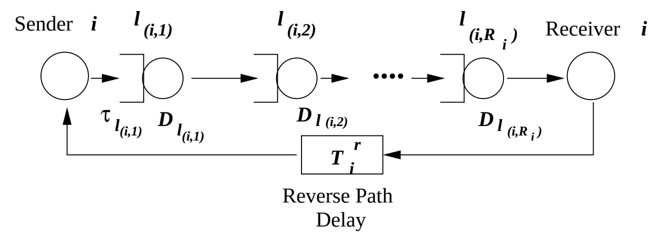


Fig. 1. Delay model.

- $T_i(t)$, $i \in \mathcal{I}$ – round-trip delay of a user i's packet whose acknowledgment (ACK) arrives at the sender at time t.
- $\tau_l(t)$, $l \in \mathcal{L}$ – queueing delay of a packet that gets to the head of queue at resource l at time t, *i.e.*, $\tau_l(t) = \frac{q_l(t-\tau_l(t))}{C_l}$, where $q_l(t)$ is the queue size at resource l at time t.

For each $i \in \mathcal{I}$ and $t \geq 0$ let $\Gamma^i_{R_i+1}(t) = T^r_i$, and for each $k \in \{1, \ldots, R_i\}$, define (recursively starting with $k = R_i$)

$$\Gamma^i_k(t) = \Gamma^i_{k+1}(t) + D_{l_{(i,k)}} + \tau_{l_{(i,k)}}(t - D_{l_{(i,k)}} - \Gamma^i_{k+1}(t))$$

The variable $\Gamma^i_k(t)$ denotes the delay experienced by the feedback signal from resource $l_{(i,k)}$ that arrives at sender i at time t. In other words, $t - \Gamma^i_k(t)$ is the time at which the packet whose ACK is received at time t arrived at the resource $l_{(i,k)}$. Similarly, for each $i \in \mathcal{I}$ and $m = 1, \ldots, R_i$, we define $\Gamma^{i,m}_m(t) = 0$ and

$$\begin{aligned}\Gamma^{i,m}_k(t) = \Gamma^{i,m}_{k+1}(t) &+ D_{l_{(i,k)}} \\ &+ \tau_{l_{(i,k)}}(t - D_{l_{(i,k)}} - \Gamma^{i,m}_{k+1}(t)) \ , \\ &\quad k = 1, \ldots, m-1 \ .\end{aligned}$$

The variable $\Gamma^{i,m}_k(t)$ gives us the delay experienced from resource $l_{(i,k)}$ to resource $l_{(i,m)}$ by user i packet that arrives at resource $l_{(i,m)}$ at time t. We define $\Gamma^{i,0}_k(t) = 0$.

Under this general model, the end user dynamics are given by

$$\frac{d}{dt}x_i(t) = \kappa_i \left(x_i(t)U_i^{'}(x_i(t)) - x_i(t - T_i(t))\mu^i(t) \right) \quad (10)$$

where

$$\mu^i(t) = \sum_{k=1}^{R_i} \mu_{l_{(i,k)}}(t - \Gamma^i_k(t))$$

and

$$\mu_l(t) = p_l\Big(\sum_{j \in I_l} x_j(t - \Gamma^{j,K(j,l)}_1(t)) \Big) \ .$$

Here $K(j,l)$ is the order of link l in user j's route if $l \in r_j$ and 0 otherwise. Under this model the price of resource l at time t depends on the rates of the users some time back due to the delay from the senders to the resource. The feedback signal generated by the resource price functions is then further delayed.

Now we describe the evolution of queue sizes $q_l(t), l \in \mathcal{L}$. Let $B_l, l \in \mathcal{L}$, denote the *finite* buffer size at resource l. Then, the queue dynamics can be captured by the following differential equations:

$$\frac{d}{dt}q_l(t) = \begin{cases} \sum_{j \in I_l} x_j(t - \Gamma^{j,K(j,l)}_1(t)) - C_l \ , \\ \qquad \text{if } 0 < q_l(t) < B_l \\ \left[\sum_{j \in I_l} x_j(t - \Gamma^{j,K(j,l)}_1(t)) - C_l\right]^+ , \\ \qquad \text{if } q_l(t) = 0 \\ \left[\sum_{j \in I_l} x_j(t - \Gamma^{j,K(j,l)}_1(t)) - C_l\right]^- , \\ \qquad \text{if } q_l(t) = B_l \end{cases} \quad (11)$$

where $[a]^+ = \max(0, a)$ and $[b]^- = \min(0, b)$.

The system given by (10) and (11) is a straightforward extension of a model used in the literature [11], [14], [20]. This type of model is, however, an approximation to a real system, and a few aspects of a real system are not modeled explicitly. For instance, when the aggregate rate at a resource is larger than its capacity, although the queue size and hence the queueing delay at the resource increase the total departure rate of the users is allowed to be larger than the capacity, Also, in a real network, the rate of a user decreases after traversing a bottleneck experiencing packet losses due to a finite buffer size. This thinning effect of user rates is not modeled in (10) and (11). However, when packet losses are not high, these shortcomings of the model do not cause a significant discrepancy in system behavior. These effects can also be mitigated by maintaining a small memory of transmission rate at the source and providing accurate prices in the packet header using a larger number of bits for congestion notification.

In this paper we are interested in studying the stability of the system given by the set of delay differential equations in (10) and (11). The goal of this paper is to study how the time-varying nature of the delays affects the stability of the system. To be more precise, we are interested in finding necessary and/or sufficient conditions on the utility and resource price functions that will ensure the convergence of $x_i(t), i \in \mathcal{I}$, to the solution of (5) regardless of the delays D_l and queue dynamics.

We begin with the existence and uniqueness of a solution of the system given by (10) and (11) in the next subsection. The issue of existence and in particular uniqueness is nontrivial for delay differential equations with state-dependent delays. There are practical examples where state-dependent delays may lead to unbounded state [3].

A. Existence and Uniqueness of Solutions

In this subsection we establish the conditions for the existence and uniqueness of a solution of (10) and (11) only for a single resource case with no forward delay to the resource. The existence of a solution for multiple resource cases is left open. However, we suspect that the approach taken in this subsection can be extended to these cases. To this end we use the framework developed by Hartung and Turi in [5], [6]. They consider a general setup for delay differential equations with distributed and state-dependent delays:

$$\dot{z}(t) = \xi(t, z(t), \Lambda(t, z_t)) \tag{12}$$

where $\Lambda(\cdot)$ describes the role played by delayed state variable and can be written as

$$\Lambda(t, z_t) = \int_{-r}^{0} d_s \nu(s, t, z_t) z(t+s) \ , \tag{13}$$

where r is the maximum possible delay, $z(t) \in \mathbb{R}^n$ for $n > 0$, z_t denotes the segment $z_t(s) \equiv z(t+s)$ for $s \in [-r, 0]$, $\nu(\cdot, \cdot, \psi)$ is an $n \times n$ matrix valued function of bounded variation on $[-r, 0]$, $\psi \in C([-r, 0], \mathbb{R}^n)$, and the integral is the Stieltjes-integral of $z(t + \cdot)$ with respect to $\nu(\cdot, t, z_t)$. The set $C([-r, 0], \mathbb{R}^n)$ denotes the set of continuous functions on $[-r, 0]$ with domain $\mathbb{R}^n$. Note that their general setup can handle the case where the function $\xi(\cdot)$ depends on time t, although our model does not require this.

Since there is only one resource, we remove the subscript and denote the queue size at the resource at time t by $q(t)$. Let $z(t) = [\overline{x}(t); q(t)]$ and $\dot{z}(t) = [\dot{\overline{x}}(t); \dot{q}(t)]$, where $\overline{x}(t) = (x_i(t), i \in \mathcal{I})$. Our model can now be viewed as a special case of (12) and be obtained by extending Example 1.3 in [5, pp. 2] as follows. Let $\nu(\cdot)$ be a diagonal matrix with

$$\nu_{ii}(s, t, \psi) \equiv \chi_{[-T_i(t,\psi),0]}(s) \ , \ s \in [-r, 0], \ i = 1, \ldots, N$$

where $T_i(t, \psi)$ gives the round-trip delay of user i at time t given some continuous function ψ in $\mathbb{R}_+^{N+1}$ (in place of z_t),[5] and $\chi_{[-T,0]}(s)$ is the characteristic function of the interval $[-T, 0]$, *i.e.,*

$$\chi_{[-T,0]}(s) = \begin{cases} 1 & \text{if } -T \le s \le 0 \\ 0 & \text{if } s < -T \text{ or } s > 0 \end{cases} \ ,$$

and $\nu_{N+1,N+1}(s, t, \psi) \equiv 0$. It is clear that $\nu(\cdot, t, \psi)$ is of bounded variation on $[-r, 0]$ for all $t \in \mathbb{R}_+$. Then, we have

$$\Lambda(t, z_t) = ((x_i(t - T_i(t)), i \in \mathcal{I}); 0)$$

[5] $T_i(t, \psi)$ is the round-delay that satisfies $T_i(t, \psi) = T_j^r + \frac{\psi_{N+1}(t - T_i(t,\psi))}{C}$, where C is the capacity of the resource.

The results developed in [5] tell us that a solution exists if (i) the function $\xi(\cdot)$ belongs to the Banach-space of bounded continuous functions on an appropriate domain of definition, and (ii) the initial function belongs to the space of continuous functions [5, pp. 15]. In our system the function $\xi(\cdot)$ is given by the right-hand side of (10) and (11), and the first condition can be easily verified. The second condition is a reasonable assumption on the initial conditions considered in this work as user rates and queue sizes must be continuous in time. Finally, the uniqueness of a solution can be guaranteed if $\xi(\cdot)$ of (12) is locally Lipschitz in both second and third argument, which can be verified in our case. These conditions provide us the basis for continuation of solutions and studying their stability.

V. Stability of General Networks

In this section we first establish of the stability of the delay differential system of (10) and (11) using a popular family of user utility and resource price functions [1], [25] for notational simplicity. The convergence results established in this section will be extended in Section VI to more general functions.

A. Utility and Resource Price Functions

The class of users' utility functions that we consider is of the form

$$U_a(x) = -\frac{1}{a}\frac{1}{x^a} \ , \ a > 0 \ . \tag{14}$$

In particular, $a = 1$ has been found useful for modeling the utility function of TCP algorithms [12]. This class of utility functions in (14) has been used extensively in engineering literature [10], [12]. Also, it has been used to carry out a trade-off between system throughput and fairness among the users [1]. With the utility functions of the form in (14) one can easily show that the price elasticity of demand decreases with a as follows. Given a price per unit flow p, the optimal rate $x^*(p)$ of the user that maximizes the net utility $U_a(x) - p \cdot x$ is given by $p^{-\frac{1}{1+a}}$. The price elasticity of demand, which measures how responsive the demand is to a change in price, is defined to be the percent change in demand divided by the percent change in price [22]. In our case the price elasticity of demand is given by

$$\frac{p}{x^*(p)}\frac{dx^*(p)}{dp} = \frac{p}{p^{-\frac{1}{1+a}}} \cdot \frac{-1}{1+a} p^{-\frac{1}{1+a}-1} = \frac{-1}{1+a} \ . \tag{15}$$

Therefore, one can see that the price elasticity of demand decreases with a,[6] *i.e.,* the larger a is, the less responsive the demand is.

[6] When comparing the price elasticity, typically the absolute value of (15) is used.

The class of resource price functions that we are interested in is of the form

$$p_l(y) = c_l \cdot \left(\frac{y}{C_l} \right)^b \ , \qquad (16)$$

where $b > 0$, c_l is some positive constant, and C_l is the capacity of resource $l \in \mathcal{L}$. However, C_l can be replaced with any positive constant, *e.g.*, virtual capacity in AVQ, so that the price function can be dynamically adjusted based on the current load using the virtual capacity as the control variable [13]. Throughout this paper we assume that $c_l = 1$ unless stated otherwise. This kind of marking function arises if the resource is modeled as an $M/M/1$ queue with a service rate of C_l packets per unit time and a packet receives a mark with a congestion indication signal if it arrives at the queue to find at least b packets in the queue. The parameter b is used to change the shape of the price function. The larger b is, the more convex and responsive the price function is.

With the utility and resource prices defined in this subsection, eq. (10) can be rewritten as

$$\frac{d}{dt} x_i(t) = \kappa_i \left(x_i^{-a_i}(t) - x_i(t - T_i(t)) \mu^i(t) \right) \qquad (17)$$

where

$$\mu_i(t) = \sum_{l \in r_i} c_l \cdot \left(\frac{\tilde{x}^l(t)}{C_l} \right)^{b_l} \qquad (18)$$

and

$$\tilde{x}^l(t) = \sum_{j \in I_l} x_j(t - \Gamma^i_{K(i,l)}(t) - \Gamma_1^{j,K(j,l)}(t - \Gamma^i_{K(i,l)}(t))) \ .$$

B. Stability Results

This subsection presents the conditions for the stability of the system given by (17) and (18). First, we introduce an assumption on the rates of the users. Since the rate of a user is limited in practice due to the link capacity and receiver buffer size, we assume that the rate of each user is upper bounded by some constant X_{max}. Similarly user rates are bounded away from zero from the fact that there is a lower bound on the transmission rate. For instance, in the case of TCP the transmission rate of a connection cannot be smaller than one packet size divided by the round-trip time of the connection. We denote this lower bound on user rates by $X_{min} > 0$. This lower bound X_{min} can be arbitrarily close to 0.

Assumption 1: A user rate belongs to a compact set $[X_{min}, X_{max}]$.

We introduce a sufficient condition on the user and resource price parameters that ensures the stability of the differential equation system given by (17) and (18).

Assumption 2: Assume that

$$a_i > 1 + \max_{l \in r_i} b_l \quad \text{for all } i \in \mathcal{I} \ . \qquad (19)$$

The above assumption tells us that for a fixed utility functions of the users, there is a limit on how responsive the resource price functions can be to changes in the arrival rates, and vice versa. Hence, this assumption captures the trade-off in the responsiveness between the end users and the resource prices.

The following main result states that if a solution of the differential equations exists, then the rates of the end users converge to the optimal rates that maximize (5) if the end users satisfy Assumption 2.

Theorem 1: Suppose that the utility functions of the end users are of the form (14), and that the price functions of the resources are given by (16). Under Assumptions 1 and 2, if a solution of the system given by (17) and (18) exists (see subsection IV-A), then $\overline{x}(t)$ converges to $\overline{x}^\star$ as $t \to \infty$ for every solution $\overline{x}(t)$, where $\overline{x}^\star$ is the solution to (5).

VI. General Utility and Resource Price Functions

In this section we extend the results in Section V to the case with more general utility and resource price functions and establish the convergence of (10) under a set of assumptions. Throughout this section we implicitly assume the existence of a solution of (10) and (11). In order to facilitate our analysis, for each user $i \in \mathcal{I}$ we define a function which gives us its willingness to pay as a function of its current rate:

$$y_i = x_i \cdot U_i'(x_i) := g_i(x_i) \ . \qquad (20)$$

Using (20) we rewrite (10) as

$$\begin{aligned} &\dot{y}_i(t) \qquad (21) \\ &= \kappa_i g_i'(g_i^{-1}(y_i(t)))[y_i(t) \\ &\quad - g_i^{-1}(y_i(t - T_i(t))) \sum_{k=1}^{R_i} p_{l_{(i,k)}}(\sum_{j \in I_{l_{(i,k)}}} g_j^{-1}(y_j(t - \\ &\quad \Gamma^i_k(t) - \Gamma_1^{j,K(j,l_{(i,k)})}(t - \Gamma^i_k(t))))] \\ &= \kappa_i g_i'(g_i^{-1}(y_i(t))) \Big[y_i(t) - \check{f}_i(\overline{g}^{-1}(\overline{y}(t - T_i(t))), \\ &\qquad (\overline{g}^{-1}(\tilde{y}_{(i,l)}(t)), l \in r_i)) \Big] \qquad (22) \end{aligned}$$

where

$$\tilde{y}_{(i,l)}(t) = \Big(y_j(t-\Gamma^i_{K(i,l)}(t)-\Gamma_1^{j,K(j,l)}(t-\Gamma^i_{K(i,l)}(t))),\ j\in\mathcal{I}\Big),\quad l\in r_i$$

and

$$\begin{aligned}&\check{f}_i\left(\overline{g}^{-1}(\overline{y}(t-T_i(t))),(\overline{g}^{-1}(\tilde{y}_{(i,l)}(t)),l\in r_i)\right)\\ &=\check{f}_i(\overline{g}^{-1}(\overline{y}(t-T_i(t))),\overline{g}^{-1}(\tilde{y}_{(i,l_{(i,1)})}(t)),\ldots,\\ &\qquad \overline{g}^{-1}(\tilde{y}_{(i,l_{(i,R_i)})}(t)))\\ &=g_i^{-1}(y_i(t-T_i(t)))\\ &\quad\times\Big[\sum_{k=1}^{R_i}p_{l_{(i,k)}}\Big(\sum_{j\in I_{l_{(i,k)}}}g_j^{-1}(y_j(t-\Gamma^i_k(t)-\\ &\qquad\Gamma_1^{j,K(j,l_{(i,k)})}(t-\Gamma^i_k(t))))\Big)\Big] \end{aligned}\tag{23}$$

Let us define

$$\begin{aligned}Y^i(t)&=\Big(\overline{y}(t-T_i(t)),\tilde{y}_{(i,l_{(i,1)})}(t),\ldots,\tilde{y}_{(i,l_{(i,R_i)})}(t)\Big)\\ Y(t)&=\Big(Y^1(t),\ldots,Y^N(t)\Big)\ ,\end{aligned}\tag{24}$$

and $G_i(Y^i(t)) = (\overline{g}^{-1}(\overline{y}(t-T_i(t))),\ \overline{g}^{-1}(\tilde{y}_{(i,l_{(i,1)})}(t)),\ \ldots,\ \overline{g}^{-1}(\tilde{y}_{(i,l_{(i,R_i)})}(t)))$. Now we can write (22) in the following matrix form:

$$\dot{\overline{y}}(t)=\overline{\kappa}(\overline{y}(t))\left[\check{F}(Y(t))-\overline{y}(t)\right]\tag{25}$$

where $\check{F}_i(Y(t)) = \check{f}_i(G_i(Y^i(t)))$, and $\overline{\kappa}(\cdot)$ is the state dependent diagonal gain matrix with

$$\overline{\kappa}_{ii}(\overline{y}(t))=-\kappa_i g_i^{'}(g_i^{-1}(y_i(t)))\ .\tag{26}$$

We make the following assumption on $g_i(\cdot), i\in\mathcal{I}$, and the resource price functions $p_l(\cdot), l\in\mathcal{L}$.

Assumption 3: (i) The function $g_i(x_i)$ is strictly decreasing with $g_i^{'}(x_i) < 0$ for all $x_i > 0$, (ii) the price functions $p_l(x)$ are strictly increasing in x for all $l\in\mathcal{L}$, and (iii) $g_i(x_i)$ is Lipschitz continuous on $[X_{min}, X_{max}]$ and $p_l(x)$ is Lipschitz continuous on $[X_{min}, |I_l|\cdot X_{max}]$, where X_{min} and X_{max} are the assumed lower and upper bound on $x_i, i\in\mathcal{I}$, respectively.

This assumption ensures that the state dependent gain matrix $\overline{\kappa}(\cdot)$ is a positive definite matrix. Assumption (i) implies that the marginal utility decreases faster than x_i^{-1} from the definition of $g_i(x_i)$. Hence, a change in price per unit flow does not cause a large change in user demand, *i.e.,* the rate at which the marginal utility equals the price per unit flow, and a user demand is relatively insensitive to price changes, *i.e.,* user demands are inelastic. It also guarantees the existence of $g_i^{-1}, i\in\mathcal{I}$. Therefore, the convergence of $\overline{y}(t)$ to $\overline{y}$ implies the convergence of $\overline{x}(t)=\overline{g}^{-1}(\overline{y}(t))$ to $\overline{x}=\overline{g}^{-1}(\overline{y})$.

It is easy to verify that the utility functions given in (14) satisfy the above assumption. From the definition of $g_i(x_i)$, we have

$$g_i(x_i)=x_i\cdot U_i^{'}(x_i)=x_i\frac{1}{x_i^{a_i+1}}=\frac{1}{x_i^{a_i}}\ .\tag{27}$$

From (27) one can easily see that Assumption 3 holds.

We define the invariance and a fixed point of the map $\check{F}(\cdot)$. Let $\Xi := \sum_{i\in\mathcal{I}}(R_i+1)$.

Definition 1: A set $D\subset\mathbb{R}^N_+$ is said to be invariant under the map $\check{F}(\cdot)$ defined in (25) if $\check{F}(Y)\in D$ whenever $Y\in D^{\Xi}$, *i.e.,* $Y=(Y^1,\ldots,Y^N)$ and $Y^i\in D^{R_i+1}$ for all $i\in\mathcal{I}$. A vector $\overline{y}^\star\in\mathbb{R}^N_+$ is said to be a fixed point of $\check{F}(\cdot)$ if $\check{F}(\overline{y}^\star,\ldots,\overline{y}^\star)=\overline{y}^\star$.

One can verify that if $\overline{y}^\star$ is a fixed point of $\check{F}(\cdot)$, then $\overline{g}^{-1}(\overline{y}^\star) = (g_1^{-1}(y_1^\star), g_2^{-1}(y_2^\star),\ldots,g_N^{-1}(y_N^\star))$ is a solution to (5) from (22) and (23), *i.e.,* $\overline{y}^\star=\overline{g}(x^\star)$, where $x^\star$ is a solution to (5).

Let $T^i_{max} = T^r_i + \sum_{l\in r_i}(D_l+\frac{B_l}{C_l})$ and $T_{max} = \max_{i\in\mathcal{I}} T^i_{max}$. We denote by $C_n([-T_{max},0],A)$ the set of n-dimensional functions that are continuous over $[-T_{max},\ 0]$ with the range A.

We now state the assumption under which the asymptotic stability of (25) is established.

Assumption 4: Multidimensional map $\check{F}:\mathbb{R}_+^{N\cdot\Xi}\rightarrow\mathbb{R}^N_+$ has a fixed point $\overline{y}^\star\in\mathbb{R}^N_+$, where $\overline{g}^{-1}(\overline{y}^\star)$ is the solution to (5). Also, assume that there is a sequence of closed, convex product spaces $D_k, k\geq 0$, such that $\check{F}(D_k^{\Xi})\subset int(D_{k+1})\subset D_{k+1}\subset int(D_k)$ and $\cap_{k\geq 0}D_k=\{\overline{y}^\star\}$, where $int(A)$ denotes the interior of the set A.

Clearly, the existence of sequence $D_k, k\geq 0$, that satisfies the above assumption depends on the selected utility and resource price functions, but not on the delays $D_l, l\in\mathcal{L}$, and $T^r_i, i\in\mathcal{I}$. One can verify that Assumption 4 holds with utility and price functions of (14) and (16), respectively, under Assumption 2.

Let $Y_{D_0} = C_N([-T_{max},0],D_0)$ and $Q = C_L([-T_{max},0],\mathcal{B} := \prod_{l\in\mathcal{L}}\mathcal{B}_l)$, where $\mathcal{B}_l=[0,B_l]$, be subsets of initial functions of $\overline{y}(s)$ and $\overline{q}(s)=(q_l(s), l\in\mathcal{L})$, $s\in[-T_{max},0]$, respectively. The solution of (25) constructed using initial functions $\phi\in Y_{D_0}$ and $\psi\in Q$ is denoted by $\overline{y}_{\phi,\psi}(t)$. When there is no confusion, we omit the subscript and use $\overline{y}(t)$. Given the initial function $\phi\in Y_{D_0}$, the initial function

ψ of queue sizes is assumed to satisfy (11).

Theorem 2: (Asymptotic Stability) All solutions $\overline{y}(t)$ starting with initial functions $\phi \in Y_{D_0}$ and $\psi \in Q$ that satisfies (11) given ϕ converge to $\overline{y}^\star$ as $t \to \infty$ for all $D_l, B_l \in \mathbb{R}_+$ and $T_i^r \in \mathbb{R}_+$.

Proof: The proof is provided in Appendix I. ∎

Our stability results consider the case where *arbitrary* delays $D_l, l \in \mathcal{L}$, and $T_i^r, i \in \mathcal{I}$, are allowed. However, if the delays are finite and upper bounded by some constant, then the system may be stable without satisfying our conditions stated in this subsection, and less stringent stability conditions may be sufficient to ensure stability *if* the upper bounds on the delays are known a priori.

One can verify that if a system is stable with a *homogeneous, constant* feedback delay as described in [20], then the system with more general (state-dependent) time-varying delays described in this section is also stable with appropriate initial conditions. Therefore, our results demonstrate that the *delay-independent* stability of the system studied in this paper depends critically on the selection of users' utility functions and resource price functions, but not on the detailed delays between network elements, *e.g.,* end users and resources. Hence, the stability of the system with a given set of users and network resources can be studied by considering a simple fixed homogeneous delay system and a natural underlying discrete time map with suitable initial conditions [20]. This result offers a tool with a potential to significantly simplify the stability study of a delay differential system given by (10) by reducing it to that of a simple system, which is described in the next section.

VII. Comparison with Deterministic Delays Case

In this section we discuss the relationship between a homogeneous delay system, *i.e.,* where every user has the same *fixed* feedback delay, and a time-varying delay system discussed in Section VI.

We first describe the system where all users have the same fixed feedback delay $T > 0$. We assume that there is no forward path delay and all of the delay lies in the reverse path, which is the same for all users and is given by T. Then, (10) can be written in a much simpler form

$$\begin{aligned} &\frac{d}{dt} x_i(t) \\ &= \kappa_i \Big(x_i(t) U_i^{'}(x_i(t)) - x_i(t-T) (\sum_{l \in r_i} \mu_l(t-T)) \Big) \end{aligned} \tag{28}$$

where $\mu_l(t-T) = p_l(\sum_{j \in I_l} x_j(t-T))$. Using the same definition $y_i(t) = x_i(t) \cdot U_i^{'}(x_i(t))$, we can write (28) in the following simple matrix form:

$$\dot{\overline{y}}(t) = \overline{\kappa}(\overline{y}(t)) \left[F(\overline{y}(t-T)) - \overline{y}(t) \right] \tag{29}$$

where $\overline{\kappa}(\overline{y}(t))$ is defined in (26) and

$$\begin{aligned} F_i(\overline{y}) &= f_i(\overline{g}^{-1}(\overline{y})) \\ &= g_i^{-1}(y_i) \sum_{l \in r_i} p_l (\sum_{j \in I_l} g_j^{-1}(y_j)) \ , \ \overline{y} \in \mathbb{R}_+^N \ . \end{aligned}$$

Note that since every user has the same delay, the multidimensional nonlinear map $F(\cdot)$ has the domain $\mathbb{R}_+^N$ (as opposed to $\mathbb{R}_+^{N \cdot \Xi}$ in section VI).

One can define a natural discrete time map from (29) that highlights the underlying market mechanism at work. Moreover, one can show that there is a close relationship between the stability of the discrete time system and that of (10) in the presence of arbitrary delays. Consider the following:

$$\overline{y}_{n+1} = F(\overline{y}_n) \ , \ n \in Z_+ := \{0, 1, 2, \ldots\} \tag{30}$$

where $\overline{y}_n \in \mathbb{R}_+^N$, and $F_i(\overline{y}) = f_i\left(g_1^{-1}(y_1), \ldots, g_N^{-1}(y_N)\right)$. Here one can easily see the market structure where in each iteration every user updates its willingness to pay to its total price based on the previous values of willingness to pay of all users, which determine the market price of resources. Therefore, one can interpret this as a simple economic model where users are probing the market and adjusting their bids based on the delayed market price and their utility functions.

Suppose that the multidimensional map $F : \mathbb{R}_+^N \to \mathbb{R}_+^N$ has some fixed point $\overline{y}^\star$, *i.e.,* $\overline{y}^\star = F(\overline{y}^\star)$, and that there is a sequence of closed, convex product spaces $\check{D}_k, k \geq 0$, such that $F(\check{D}_k) \subset int(\check{D}_{k+1}) \subset \check{D}_{k+1} \subset int(\check{D}_k)$ and $\cap_{k \geq 0} \check{D}_k = \{\overline{y}^\star\}$. Then, it is shown that if the initial functions $\phi \in \check{D}_0$, *i.e.,* $\overline{y}(s) \in \check{D}_0$ for all $-T \leq s \leq 0$, then $\lim_{t \to \infty} \overline{y}(t) = \overline{y}^\star$ for all $T > 0$ [20].

A key observation to be made here is the following: Consider a system consisting of a set of users and resources that satisfy Assumption 3. First, suppose that the users have the same delay T and the assumption in the previous paragraph holds for some sequence of $\check{D}_k, k \geq 0$, and hence the system is stable provided that the initial functions lie in $\check{D}_0$. Then, one can easily verify that even when the users have heterogeneous time-varying delays as described in Section IV, the same sequence $D_k = \check{D}_k, k \geq 0$, satisfies Assumption 4 from the monotonicity properties of the maps by Assumption 3 and, hence, the system is stable if the initial functions lie in $D_0 = \check{D}_0$.

This observation can be explained to some extent as follows. Suppose that in the homogeneous delay case, the delay of the flows is T, and in the heterogeneous delays case, the delays of the flows satisfy $T^i_{max} \le T$ for all $i \in \mathcal{I}$. Then, since the communication delays of the users in the heterogeneous delay case are no larger than T of the homogeneous delay case, one may expect the system with heterogeneous delays to be stable if the system with homogeneous delay is stable. However, since our stability results for homogeneous delay case hold for any arbitrary T [20], one should expect the system with heterogeneous time-varying delays to be stable irrespective of $D_l, l \in \mathcal{L}$, and $T^r_i, i \in \mathcal{I}$, as well.

VIII. Numerical Examples

In this section we present a few numerical examples to illustrate the results in the previous sections. We demonstrate that under the stability condition given in Theorem 1 indeed the user rates converge to the solution of (5), while when the condition does not hold, for sufficiently large delays, the system becomes unstable.

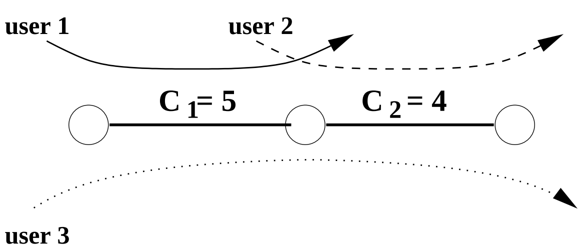
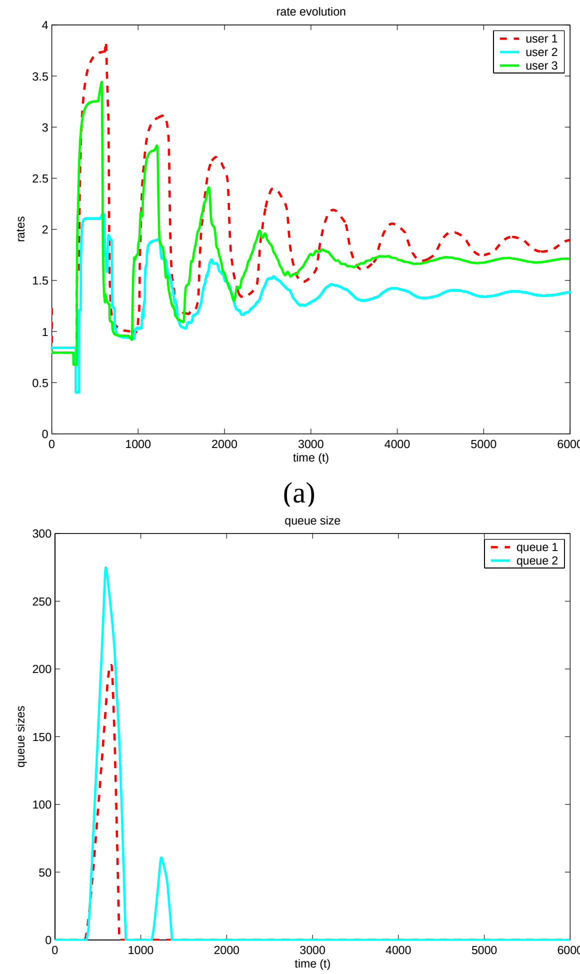


Fig. 2. Topology of the example network.

We consider a simple network consisting of two links. The first link is shared by users 1 and 3, while the second link is share by users 2 and 3. This is shown in Fig. 2. The capacities of the links are set to $C = [C_1 \ C_2]^T = [5 \ 4]^T$. The delays are set to $[D_1 \ D_2] = [50 \ 40]$, and $[T^r_1 \ T^r_2 \ T^r_3] = [289 \ 250 \ 277]$. The buffer sizes are set to $[B_1 \ B_2] = [400 \ 1200]$. The utility functions are of the form in (14) with $a_1 = a_2 = 3$ and $a_3 = 4$, and the resource price functions are of (16) with $c_l = 0.2$, $l = 1, 2$. We select two different sets of parameters $b_l, l = 1, 2$, to create both a stable system and an unstable system. The initial function is given by $\phi_x(s) = [x_1(s) \ x_2(s) \ x_3(s)]^T = [3 \ 1 \ 2]^T$ for all $s \in [-T_{max}, \ 0]$, and the initial queue sizes were set to zero at $t = -T_{max}$. The gain parameters are set to $\kappa_i = 1$ for all users.

A. Stable System

The resource price parameters for the first case are set to $b_1 = b_2 = 1.9$. Since $a_i > b^i_{max} + 1$ for all users, the system is stable. One can solve the optimization problem in (5) and show that the solution is given by $x^\star = [1.844 \ 1.696 \ 1.370]^T$. Fig. 3 plots the evolution of user rates and queue sizes according to (10) and (11), respectively. As one can see the user rates converge to the optimal rates.



(a)

(b)

Fig. 3. User rate and queue size evolution of a stable system.

B. Unstable System

In the second case the resource price parameters are increased to $b_1 = b_2 = 3.5$. One can easily see that the stability conditions in Theorem 1 do not hold. Fig. 4 shows the unstable behavior of the system as the user rates show no sign of settling down, resulting in large oscillations in the rates as well as in queue sizes.

IX. Conclusions

In this paper we have studied the problem of designing a robust congestion control mechanism in the presence of arbitrary delays between the end users and network resources with time-varying queue sizes. We have provided a condition for the stability of the system, which

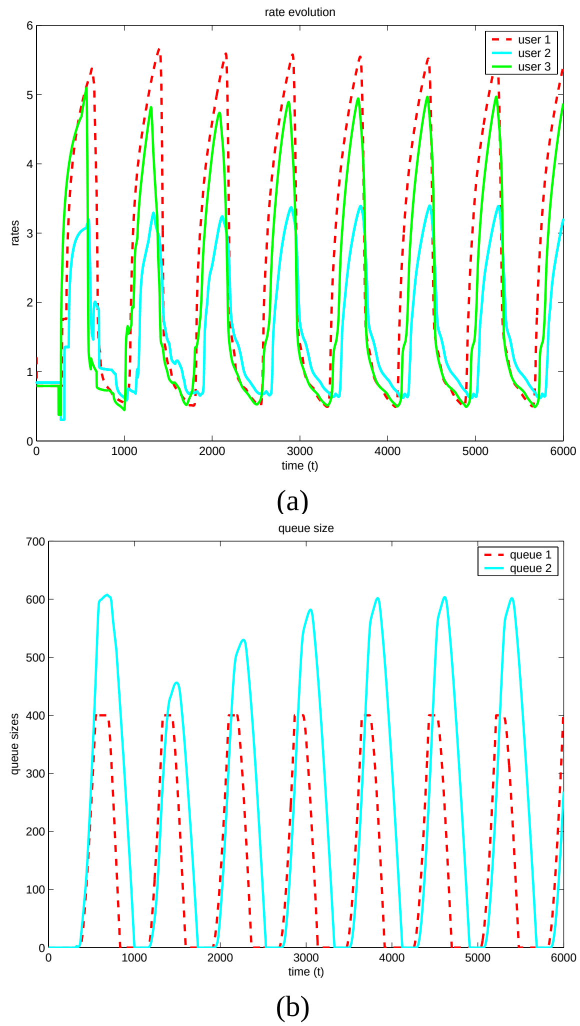


Fig. 4. User rate and queue size evolution of an unstable system.

is applied to establish the stability of the system with a family of popular utility and resource price functions. We have demonstrated that the stability of such a system can be also studied by considering a much simpler discrete time system where all users have the same fixed homogeneous delay.

Modeling the time-varying queue dynamics may be important in multi-hop wireless networks where, due to a limited capacity, queueing delays may be comparable to transmission and propagation delays. Our results provide a efficient tool for studying the stability of a rate control system with time-varying delays by allowing us to look at a much simpler discrete time system instead, which can be analyzed more easily.

References

[1] R. Agrawal, A. Bedekar, R. J. La, and V. G. Subramanian. C3wpf scheduler. In *Proc. ITC*, 2001.

[2] T. Alpcan and T. Basar. A utility-based congestion control scheme for Internet-style networks with delay. In *Proc. of INFOCOM*, 2003.

[3] M. Büger and M. W. Martin. The escaping disaster: A problem based on state-dependent delays. *to appear in Z.A.M.P.*, 2002.

[4] S. Deb and R. Srikant. Global stability of congestion controllers for the internet. In *Proc. of IEEE CDC*, 2002.

[5] F. Hartung. *On classes of functional differential equations with state-dependent delays*. PhD thesis, University of Texas at Dallas, Dallas, Texas, 1995.

[6] F. Hartung and J. Turi. Stability in a class of functional differential equations with state dependent delays. *Qualitative Problems for Differential Equations and Control Theory*, pages 15–31, 1995.

[7] A. F. Ivanov, M. A. Pinto, and S. I. Trofimchuk. Global behavior in nonlinear systems with delayed feedback. *Proc. Conference of Decison and Control, Sydney*, 2000.

[8] A. F. Ivanov and A. N. Sharkovsky. Oscillations in singularly perturbed delay equations. *Dynamics Reported*, 1:165–224, 1991.

[9] R. Johari and D. Tan. End-to-end congestion control for the internet: Delays and stability. *IEEE Transactions on Networking*, 2000.

[10] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–7, January 1997.

[11] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, March 1998.

[12] S. Kunniyur and R. Srikant. End-to-end congestion control schemes: Utility functions, Random losses and ECN marks. In *Proc. of IEEE INFOCOM*, Tel-Aviv, Israel, 2000.

[13] S. Kunniyur and R. Srikant. Analysis and design of an adaptive virtual queue algorithm for active queue management. In *Proc. of ACM SIGCOMM*, 2001.

[14] S. Kunniyur and R. Srikant. A time scale decomposition approach to adaptive ecn marking. In *Proc. of IEEE INFOCOM*, Anchorage, AK, 2001.

[15] R. J. La and V. Anantharam. Utility-based rate control in the Internet for elastic traffic. *IEEE/ACM Transactions on Networking*, 10(2):272–86, April 2002.

[16] S. Low and D. Lapsley. Optimization flow control. *IEEE/ACM Transactions on Networking*, 7(6):861–874, June 1997.

[17] J. Mallet-Paret and R. D. Nussbaum. Global continuation and asymptotic behaviour for periodic solutions of a differential-delay equation. *Annali di Matematica Pura ed Applicata*, 145(4):33–128, 1986.

[18] F. Mazenc and S.-I. Niculescu. Remarks on the stability of a class of tcp-like congestion control models. In *Proc. of IEEE CDC*, 2003.

[19] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.

[20] P. Ranjan, E. H. Abed, and R. J. La. Delay, elasticity, and stability trade-offs in rate control. Technical report, submitted to IEEE/ACM Trans. on Networking, 2003.

[21] P. Ranjan, R. J. La, and E. H. Abed. Global stability in the presence of state-dependent delay for rate control. In *Proc. of CISS*, Oahu, HI, 2004.

[22] Hal R. Varian. *Intermediate Microeconomics*. Norton, New York, 1996.

[23] Z. Wang and F. Paganini. Global stability with time-delay in network congestion control. In *Proc. of IEEE CDC 2002*, Las Vegas, 2002.

[24] J. Wen and M. Arcak. A unifying passivity framework for network flow control. In *Proc. of IEEE INFOCOM*, 2003.

[25] L. Ying, G. E. Dullerud, and R. Srikant. Global stability of Internet congestion controllers with heterogeneous delays. In *Proc. of ACC*, 2004.

APPENDIX I
PROOF OF THEOREM 2

In order to prove the theorem, We first show that the invariance property of the map $\check{F}(\cdot)$ also implies the invariance property of (25). Then, we prove that Assumption 4 implies the stability of (25) as well. To this end, we use the following three lemmas.

Lemma 1: (Invariance) Suppose that $D \subset \mathbb{R}_+^N$ is a closed, convex, invariant product space under $\check{F}(\cdot)$. Then, for any initial function $\phi \in C_N([-T_{max}, 0], D) := X_D$ and $\psi \in Q$ that satisfies (11) the resulting $\overline{y}(t)$ from (25) belongs to the set D for all $t \geq 0$.

Proof: We prove the lemma by contradiction. Suppose that the lemma is false. Then, there exist some initial function $\phi \in X_D$, $\psi \in Q$, and $t,\ t \geq 0$, such that $y(t) \notin D$. Define

$$t_0 = \inf\{t \geq 0 \mid \text{ for every interval } [t,\ t^{'}),\ \text{where } t^{'} > t, \\ \exists\ t_1,\ t < t_1 < t^{'},\ \text{such that } \overline{y}(t_1) \notin D\}\ .$$

Then, there is $i \in \mathcal{I}$ such that for all $(t_0, t^{'})$, where $t^{'} > t_0$, there exists $\tilde{t}_0, t_0 < \tilde{t}_0 < t^{'}$, such that $y_i(\tilde{t}) \notin D_i := \text{proj}_i(D)$, where $\text{proj}_i(D)$ denotes the projection of i-th component of D. We assume that $y_i(t)$ leaves the interval D_i through the right end, *i.e.*, $y_i(t_0) = \sup\ D_i$. Then, for all $(t_0, t^{'})$ there exists $\tilde{t}_0, t_0 < \tilde{t}_0 < t^{'}$, such that $y_i(\tilde{t}_0) > \sup\ D_i$ and $\dot{y}_i(\tilde{t}_0) > 0$. This, however, leads to a contradiction as follows.

From (22) we have

$$\begin{aligned} &\dot{y}_i(\tilde{t}_0) \\ &= \kappa_i g_i^{'}(g_i^{-1}(y_i(\tilde{t}_0)))\Big[y_i(\tilde{t}_0) - \check{f}_i(\overline{g}^{-1}(\overline{y}(\tilde{t}_0 - T_i(\tilde{t}_0))), \\ &\qquad (\overline{g}^{-1}(\tilde{y}_{(i,l)}(\tilde{t}_0)), l \in r_i))\Big] < 0 \end{aligned}$$

because $\kappa_i g_i^{'}(g_i^{-1}(y_i(\tilde{t}_0))) < 0$ and $\check{f}_i(\overline{g}^{-1}(\overline{y}(\tilde{t}_0 - T_i(\tilde{t}_0))), (\overline{g}^{-1}(\tilde{y}_{(i,l)}(\tilde{t}_0)), l \in r_i)) \in D_i$ from the assumption and, hence, is less than or equal to $\sup D_i$ ($< y_i(\tilde{t}_0)$). This contradicts the earlier assumption that $\dot{y}_i(\tilde{t}_0) > 0$. The other case that $y_i(t)$ leaves D_i through the left end, *i.e.*, $y_i(t_0) = \inf\ D_i$, can be shown to lead to a similar contradiction. Therefore, the lemma follows. ■

Lemma 2: Fix $k, k \geq 0$. Let $\tilde{D}$ be an open product space that contains $\check{F}(D_k^{\Xi})$ and whose closure is contained in $int(D_k)$, *i.e.*, $cl(\tilde{D}) \subset int(D_k)$. Suppose that the initial function $\phi \in C_N([-T_{max}, 0], D_k)$ and $\psi \in Q$. Then, there exists a finite $\tilde{t},\ \tilde{t} \geq 0$, such that, for all $t \geq \tilde{t}$, $y(t) \in \tilde{D}$.

In order to prove the lemma, we first prove the following coordinate-wise invariance.

Lemma 3: (Coordinate-wise Invariance) If $y_i(\tilde{t}) \in \tilde{D}_i = \text{proj}_i(\tilde{D})$ for some $\tilde{t} \geq 0$, then $y_i(t) \in \tilde{D}_i$ for all $t \geq \tilde{t}$.

Proof: Suppose that the lemma is not true, and there exists $\bar{t} > \tilde{t}$ at which $y_i(\bar{t}) = \inf \tilde{D}_i$ or $y_i(\bar{t}) = \sup \tilde{D}_i$. We let

$$\bar{t} = \inf\{t \geq \tilde{t} \mid y_i(t) \in \partial\tilde{D}_i\}\ ,$$

where $\partial\tilde{D}_i$ is the boundary of the set $\tilde{D}_i$, and assume $y_i(\bar{t}) = \sup \tilde{D}_i > \sup \text{proj}_i(\check{F}(D_k^{\Xi}))$. Then, we can find $\bar{t}_1 < \bar{t}$ such that for all $t \in (\bar{t}_1, \bar{t})$, $y_i(t) \in \tilde{D}_i \setminus \text{proj}_i(\check{F}(D_k^{\Xi}))$. This implies that $\dot{y}_i(t) < 0$ for all $t \in (\bar{t}_1, \bar{t})$ from (22) because $\check{F}_i(Y(t)) \leq \sup \text{proj}_i(\check{F}(D_k^{\Xi}))$ and, thus, $y_i(\bar{t}) < \sup\ \tilde{D}_i$, leading to a contradiction. A similar argument can be used for the case $y_i(\bar{t}) = \inf\ \tilde{D}_i$. ■

Now let us proceed with the proof of Lemma 2.

Proof: (Lemma 2) Suppose that the lemma is false. Then, from Lemma 3 there exists $i \in \mathcal{I}$ such that for all $t \geq 0$, $y_i(t) \notin \tilde{D}_i$. We show that this leads to a contradiction. Suppose that $y_i(t) \geq \sup\ \tilde{D}_i$ for all $t \geq 0$. Since $\sup\ \tilde{D}_i > \sup\ \text{proj}_i(\check{F}(D^{\Xi}))$ with $\delta := \sup\ \tilde{D}_i - \sup\ \text{proj}_i(\check{F}(D^{\Xi})) > 0$, there exists some positive constant ε such that $\dot{y}_i(t) \leq -\varepsilon \cdot \delta < 0$ from (22). For example, if we let $\varepsilon = \inf_{\overline{y} \in D_k}\ \min_{i \in \mathcal{I}} \overline{\kappa}_{ii}(\overline{y})$, then $\dot{y}_i(t) \leq -\varepsilon \cdot \delta$ for all $t \geq 0$. This, however, implies that $y_i(t) \downarrow -\infty$ as $t \uparrow \infty$, contradicting the assumption that $y_i(t) \geq \sup\ \tilde{D}_i$ for all $t \geq 0$. A similar contradiction can be shown when we assume $y_i(t) \leq \inf\ \tilde{D}_i$ for all $t \geq 0$. This completes the proof of the lemma. ■

Lemma 4: Let D be a closed, convex, invariant product space and $\tilde{D}$ an open product space that contains $\check{F}(D^{\Xi})$ and whose closure is contained in $int(D)$, *i.e.*, $cl(\tilde{D}) \subset int(D)$. Suppose that the initial function $\phi \in C_N([-T_{max}, 0], D)$, $\psi \in Q$, and $y(t_1) \in \tilde{D}$ for some $t_1 \geq 0$. Then, $y(t) \in \tilde{D}$ for all $t \in [t_1,\ t_1 + T_{max}]$.

Proof: The lemma follows directly from Lemma 3. ■

The proof of the theorem can now be completed as follows. By repeatedly applying Lemmas 1, 2, and 4, one can find a sequence of finite $t_k, k = 1, 2, \ldots$, such that $y(t) \in D_k$ for all $t \geq t_k$. The theorem now follows from Assumption 4 that $\cap_{k=1}^{\infty} D_k = \{\overline{y}^{\star}\}$.