

Equilibrium Bandwidth and Buffer Allocations for Elastic Traffics

Steven H. Low, *Senior Member, IEEE*

Abstract—Consider a set of users sharing a network node under an allocation scheme that provides each user with a fixed minimum and a random extra amount of bandwidth and buffer. Allocations and prices are adjusted to adapt to resource availability and user demands. Equilibrium is achieved when all users optimize their utility and demand equals supply for nonfree resources. We analyze two models of user behavior. We show that at equilibrium expected return on purchasing variable resources can be higher than that on fixed resources. Thus users must balance the marginal increase in utility due to higher return on variable resources and the marginal decrease in utility due to their variability. For the first user model we further show that at equilibrium where such tradeoff is optimized *all* users hold strictly positive amounts of variable bandwidth and buffer. For the second model we show that if both variable bandwidth and buffer are scarce then at equilibrium every user either holds both variable resources or none.

Index Terms—Bandwidth and buffer allocation, elastic traffic, equilibrium allocation, equilibrium pricing.

I. INTRODUCTION

THIS PAPER is motivated by two emerging trends in packet-switched networks in the past decade. The first is the increasing popularity of resource reservation as a means to guarantee quality of service (QoS), as exemplified by the recent standards on the mechanism (but not the algorithm) of resource reservation, e.g., [31], [22]. However, unlike circuit-switched or leased-line networks where reservation takes the form of a fixed amount of dedicated bandwidth, in packet-switched networks, it may also include a variable component, as in the available bit rate (ABR) service of an ATM network where a user can receive a minimum cell rate (MCR) plus some random extra bandwidth. In response to the first phenomenon is the design of a large number of packet-scheduling policies to provide the reserved bandwidth to competing users; some recent examples include, e.g., generalized processor sharing of [19], a family of fair-queueing algorithms [8], [27], and the virtual partitioning policy of [16], etc; also see [30] for a survey of several earlier

schemes. The common theme across these algorithms is that a user is guaranteed a minimum share of resources and gets random extra amounts depending on network condition. For elastic traffics [25] that can tolerate some degree of delay or loss, buffer is also a scarce resource to be traded-off in network resource allocation, e.g., [11], [7], [20], [6]. Again, buffer allocation can be implemented by schemes ranging from complete partitioning, in which all users are guaranteed fixed amounts of buffer, to complete sharing, in which no user is guaranteed any fixed amounts of buffer, e.g., [4].

In this paper we describe a novel model for such bandwidth and buffer allocation schemes and study the equilibrium allocation that would result when users interact under such schemes.

We consider the idealized situation where users can freely choose their shares of fixed and variable bandwidth or buffer to maximize their benefit. The network coordinates their choices through resource pricing. This approach seems more desirable than one in which the network decides the allocation to all users without regard to their individual valuation of resources. A critical issue in such an approach concerns equilibrium: whether an equilibrium exists, how it might be approached and what properties it possesses. This paper presents a first step in answering these questions. Equilibrium is desirable as it is Pareto optimal when the utility functions are concave increasing, i.e., it is not possible to increase the utility of any user without decreasing that of others [28]. Moreover, equilibrium is critical in understanding and controlling a large variety of natural, social, and manmade systems. For instance, flow control has been formulated as a welfare maximization in [13], [1] where the objective is to maximize aggregate user utility over bandwidth allocation, subject to capacity constraints at network resources. There the flow control mechanism is derived as a distributed gradient projection algorithm to solve the dual problem, in which network links adjust prices according to the rule of supply and demand and users adjust their rates to maximize their own benefit, in much the same way as the models and algorithms presented below. The unique equilibrium is the stable point to which all trajectories of source rates converge. Moreover, it determines the fairness of the allocation. The key difference here is the inclusion of buffer resource and the differentiation of fixed and variable resources. This greatly complicates the equilibrium situation.

In Section II, we describe two models of allocating variable bandwidth and buffer. In Sections III and IV, we present our equilibrium analysis for each of the models. Some implications of the results are explained in Section V.

Manuscript received June 12, 1997; revised February 23, 1998; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor G. Sasaki. This work was supported in part by the Australian Research Council under Grant S49813050 and Grant A49930405. Partial and preliminary results have been presented in INFORMS Applied Probability Conference, Boston, MA, June/July 1997 and in Infocom'98, San Francisco, CA, March 1998.

The author was with the Department of Electrical & Electronic Engineering, University of Melbourne, Vic. 3052 Australia. He is now with the Computer Science Department and also with the Electrical Engineering Department, California Institute of Technology, Pasadena, CA 91125 USA.

Publisher Item Identifier S 1063-6692(00)04994-3.

II. MODEL, ASSUMPTIONS, AND NOTATIONS

A. Model

For our purposes a network provides two types of resources, bandwidth and buffer, in two flavors, fixed and variable. A bandwidth allocation is specified by a pair $x = (x_0, x_1) \geq 0$, with the interpretation that a fixed amount $x_0 R_0$ and a random amount $x_1 R_1$ of bandwidth will be made available to a user that is granted the allocation x . Similarly a buffer allocation is specified by a pair $y = (y_0, y_1) \geq 0$, with the interpretation that, if granted, a fixed amount $y_0 B_0$ and a random amount $y_1 B_1$ of buffer will be made available. Here, $R_0 > 0$, $B_0 > 0$ are real numbers and $R_1 \geq 0$, $B_1 \geq 0$ are (almost surely) nonnegative random variables. R_1 and B_1 are generally statistically dependent. They model the random fluctuation of resource availability in a real network.

The network also sets prices on the resources. Fixed bandwidth is priced at p_0 per R_0 amount and variable bandwidth is priced at p_1 per R_1 (random) amount. Fixed buffer is priced at q_0 per B_0 amount and variable buffer is priced at q_1 per B_1 amount. An allocation $(x, y) = (x_0, x_1, y_0, y_1)$ thus costs $px + qy = p_0 x_0 + p_1 x_1 + q_0 y_0 + q_1 y_1$. Hence a user is priced not on the actual amounts of resources that are made available to it (which is random depending on the realization of R_1 and B_1), but on its reservation (x, y) .

The network is shared by a set of users $1, \dots, N$. User n is free to choose a bandwidth and buffer allocation (x_n, y_n) it desires and pays $px_n + qy_n$ for it, if granted. User n then receives a random amount $r = xR$ of bandwidth and a random amount $b = yB$ of buffer, where $xR = x_0 R_0 + x_1 R_1$ and $yB = y_0 B_0 + y_1 B_1$. This provides user n a utility of $u_n(E(r+b), \text{var}(r+b))$, where u_n is a deterministic function from $\mathbb{R} \times \mathbb{R}_+$ to \mathbb{R} . We will assume below that the utility function $u_n(\mu, v)$ is strictly increasing in μ and strictly decreasing in v . This assumes that user n 's preference depends on the (random) resources (r, b) only through their first two moments. Moreover, increasing the mean allocation increases user n 's utility, while increasing the variance of the allocation decreases its utility. This is motivated by the fact that the performance of applications typically improves as they get more network resources and degrades as the availability of these resources fluctuates.¹

Example 1: First, note that R_0 and R_1 (and similarly for B_0 and B_1) are just the units of accounting, and not the total capacity to be shared by competing users. The total capacities are $\sum_n x_{n0} R_0$ and $\sum_n x_{n1} R_1$. In multilayer coded video transmission, a user (sender) transmits a base layer and an additional layer to enhance the quality. Suppose each user n requests, and is guaranteed, a rate x_{n0} for its base layer. The user also requests an allocation of x_{n1} for its enhancement layer but will receive a random amount depending on network condition. Buffer is needed to prevent loss when the user's instantaneous rate exceeds the available bandwidth. Similarly, buffer can be allocated to user n according to its reservation specified by (y_{n0}, y_{n1}) . \square

¹The value of bandwidth and buffer to an application may be different. Often this effect can be modeled by including in the utility function appropriate weights α, β , the utility being $u_n(E(\alpha r + \beta b), \text{var}(\alpha r + \beta b))$ instead. All results in the paper generalize in a straightforward manner to the case of nonunity weights.

In this paper, we consider two models of user behavior. In the first model **M1**, we assume that user n is endowed with an initial allocation $\bar{x}_n = (\bar{x}_{n0}, \bar{x}_{n1}) > 0$ and $\bar{y}_n = (\bar{y}_{n0}, \bar{y}_{n1}) > 0$. User n 's goal is to purchase a possibly new allocation (x_n, y_n) by selling some of its initial holdings so as to maximize its benefit, i.e., it solves the following maximization problem:

M1:

$$\begin{aligned} \max_{(x,y) \geq 0} \quad & u_n(E(xR + yB), \text{var}(xR + yB)) - (px + qy) \\ \text{subject to} \quad & px + qy = p\bar{x}_n + q\bar{y}_n. \end{aligned}$$

The constraint says that the new resource cost of user n , the left-hand side, remains the same as user n 's initial cost, the right-hand side. This is desirable because (if the prices here are not just control signals, but indeed form part of the monetary tariff paid by the users) users will then face the same connection charge throughout their service period even though their allocations may fluctuate in response to changes in resource availability or user demands.

Model **M1** excludes the possibility that a user may wish to purchase more resources than their current holdings and enjoy a higher quality, or less to save. This is permitted in the next model where the budget constraint is removed and user n chooses its allocation to solve

M2:

$$\max_{(x,y) \geq 0} \quad u_n(E(xR + yB), \text{var}(xR + yB)) - (px + qy).$$

In both **M1** and **M2** the allocations (x_n, y_n) are restricted to be nonnegative. A variant of **M1** where the nonnegativity constraint on (x_n, y_n) is removed is treated in [12]. It models users (resellers) who can both buy and sell bandwidth and buffers among themselves through the network. The nonnegativity constraint here turns out to destroy the simple structure of the process and greatly complicates the analysis.

Let $X_i = \sum_n x_{ni}$ (respectively, $\bar{X}_i = \sum_n \bar{x}_{ni}$) and $Y_i = \sum_n y_{ni}$ (respectively, $\bar{Y}_i = \sum_n \bar{y}_{ni}$) denote the total demand for (respectively, supply of) bandwidth and buffer i , respectively. For model **M2** \bar{X} and \bar{Y} represent the capacity in the network. We focus on the equilibrium situation.

Definition 1: A set of allocation vectors $(x_n, y_n, n = 1, \dots, N) \geq 0$ and a price vector $(p, q) \geq 0$ form a *competitive equilibrium* (or just *equilibrium*) if

- 1) (x_n, y_n) solves the maximization problem for all users n ; and
- 2) for $i = 0, 1$, $X_i \leq \bar{X}_i$, $Y_i \leq \bar{Y}_i$, and $p_i(X_i - \bar{X}_i) = 0$, $q_i(Y_i - \bar{Y}_i) = 0$.

We call (p, q) an *equilibrium price* and $(x_n, y_n, n = 1, \dots, N)$ an *equilibrium allocation*.

Hence at equilibrium all users n achieve their optimality given the price vector (p, q) , and for all *nonfree* resources, demand (X_i, Y_i) equals supply (\bar{X}_i, \bar{Y}_i) .

B. Assumptions and Notation

We make the following assumptions:

- A1:** $ER_1 > 0, EB_1 > 0, \text{cov}(R_1, B_1) \neq 0$, and that R_1 and B_1 are linearly independent (i.e., there exists no constant a such that $R_1 = aB_1$ almost surely).
- A2:** $u_n(\mu, v) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is jointly concave in its arguments and satisfies

$$\frac{\partial u_n}{\partial \mu}(\mu, v) > 0 \quad \text{and} \quad \frac{\partial u_n}{\partial v}(\mu, v) < 0.$$

- A3:** For all $(\mu, v) > 0$,

$$\frac{\frac{\partial u_n}{\partial \mu}(\mu, v)}{-\frac{\partial u_n}{\partial v}(\mu, v)} < 2 \frac{v}{\mu}.$$

- A4:** For model **M1**, the current allocations are strictly positive, i.e., $\bar{x}_{ni} > 0$ and $\bar{y}_{ni} > 0$ for $i = 0, 1, n = 1, \dots, N$.

Assumption **A1** implies that the variabilities of bandwidth and buffer are correlated. This is certainly so if they model the left-over resources from serving higher priority users. As discussed in the last section, by **A2** we assume that user n 's utility increases with the mean allocation and decreases with its variance. In practice u_n can be restricted to a finite region. Assumption **A3** is technical. Its sole purpose is to rule out the unnatural case where some users may desire at optimality only variable bandwidth without fixed bandwidth nor any buffer, fixed or variable, or only variable buffer without fixed buffer nor any bandwidth, fixed or variable (Lemma 4 in Section III-D). Though reasonable, **A3** is admittedly quite restrictive, but as the example in Section III-B shows, it is not necessary for the main results to hold.

Let $r = ER_1$, $b = EB_1$ be the means of R_1 and B_1 respectively. Let $\sigma_R^2 = \text{var}(R_1)$, $\sigma_B^2 = \text{var}(B_1)$, and $c = \text{cov}(R_1, B_1)$. As in the last subsection $(x_n, y_n) = (x_{n0}, x_{n1}, y_{n0}, y_{n1})$ denotes user n 's allocation, $X_i = \sum_n x_{ni}$ (respectively, $\bar{X}_i = \sum_n \bar{x}_{ni}$) and $Y_i = \sum_n y_{ni}$ (respectively, $\bar{Y}_i = \sum_n \bar{y}_{ni}$) denote the total demand for (respectively, supply of) bandwidth and buffer i , respectively.

Define for $(x, y) = (x_0, x_1, y_0, y_1)$

$$f_n(x, y) = u_n(\mu(x, y), v(x, y)) \quad (1)$$

where

$$\begin{aligned} \mu(x, y) &= xER + yEB \\ &= x_0R_0 + y_0B_0 + x_1r + y_1b \end{aligned} \quad (2)$$

$$\begin{aligned} v(x, y) &= \text{var}(xR + yB) \\ &= \sigma_R^2 x^2 + \sigma_B^2 y^2 + 2cx_1y_1. \end{aligned} \quad (3)$$

For model **M1**, since $px_n + qy_n$ is constrained to be a constant given price $(p, q) > 0$, user n 's maximization problem is equivalent to

M1:

$$\begin{aligned} &\max_{(x, y) \geq 0} f_n(x, y) \\ &\text{subject to} \quad px + qy = p\bar{x}_n + q\bar{y}_n. \end{aligned}$$

For model **M2**, it is

M2:

$$\max_{(x, y) \geq 0} f_n(x, y) - (px + qy).$$

In the following we will use $(x_n(p, q), y_n(p, q))$ to denote user n 's solution at prices (p, q) .

III. EQUILIBRIUM OF **M1**

In this section we first discuss how an equilibrium might be approached through simple price adjustment schemes. Then we present some interesting properties on equilibrium prices and allocations. Finally we prove these results. As can be seen the proof on equilibrium allocations can be quite intricate but tedious.

A. Convergence

Using standard equilibrium analysis (see, e.g., [18]) we can show that a competitive equilibrium exists for model **M1**. The basic idea of the proof in [18] is to apply Kukatani's fixed point theorem to establish that an equilibrium is a solution of an appropriate welfare maximization whose maximizer obviously exists.

Proposition 1: There exists a competitive equilibrium.

Suppose the network dynamically adjusts prices (p, q) based on user demands (X, Y) and supply (\bar{X}, \bar{Y}) , and users dynamically make requests (x_n, y_n) , $n = 1, \dots, N$, based on current prices. The system dynamic depends on how the resources are rationed and their prices adjusted before equilibrium is reached, i.e., when the total requested allocation (X, Y) is not equal to the total supply (\bar{X}, \bar{Y}) . Once equilibrium is reached, *if ever*, no source has any incentive to unilaterally deviate from it (a Nash equilibrium). The prices and the requests can be described by a continuous time dynamical system where prices $(p(t), q(t))$ at time t are solutions to

$$(\dot{p}(t), \dot{q}(t)) = (g_1(x(t), y(t)), g_2(x(t), y(t))). \quad (4)$$

Here $g_i : \mathbb{R}^{4N} \rightarrow \mathbb{R}$ are the price adjustment rules, based on user requests, and $(x(t), y(t)) = (x_n(t), y_n(t))$, $n = 1, \dots, N$ is the vector of user requests at time t that solves the user maximization problem **M1** using current prices $(p(t), q(t))$. Will the system converge to an equilibrium under simple rules (g_1, g_2) that do not require the knowledge of user utility functions on the part of the network?

Before reviewing three classes of such algorithms, we remark that (4) tacitly assumes that users n are memoryless in that their decisions at time t are independent of decisions at all other times, future or past, given the current prices. For model **M1**, the constraint also implies that, before an equilibrium is reached, resources are rationed in such a way that every user n 's resource cost is always equal to the cost of its initial allocation *at current prices* $(p(t), q(t))$ so that user n faces the same cost throughout. A simple way to ensure this is to always allocate to user n its initial allocation before an equilibrium is reached. Alternatively we can interpret the dynamical system as modeling a request phase (a *tatonnement* process [28]) where, *iteratively*, users make requests and the network adjusts prices

until the market (approximately) clears and an equilibrium is (approximately) reached. Only then are the users' allocations actually changed.

To simplify notation let $Z = (X - \bar{X}, Y - \bar{Y})$ be the vector of excess demands for the resources and $s = (p, q)$ be their prices. $Z(t)$ denotes the excess demands at current prices $s(t)$.

The simplest scheme is to adjust the price according to the law of supply and demand: if there is excess demand for a resource, raise its price; otherwise, reduce it. That is

$$\dot{s}(t) = \gamma Z(t). \quad (5)$$

This scheme does not converge in general [23], but it does converge at least locally if the aggregate demand Z satisfies a *gross substitute* condition [21, pp. 167, 171, 173] $\partial Z_i / \partial s_i < 0$ and $\partial Z_i / \partial s_j > 0$ for $i \neq j$, $i, j = 1, \dots, 4$. The condition means that raising the price of a resource decreases its demand and increases the demand for other resources. It imposes certain conditions on the user utility functions u_n [5].

When the gross substitute condition does not hold, a global Newton scheme is described in [26] that essentially scales the vector of excess demand vector Z by the inverse (if exists) of its gradient matrix $\nabla Z = [\partial Z_i / \partial s_j]$:

$$\dot{s}(t) = [\nabla Z]^{-1}(t) Z(t).$$

It is shown there that, starting from a boundary point, the prices and allocations converge to an equilibrium when the excess demand Z is merely continuous and satisfies a mild boundary condition.

A path-following algorithm has been proposed, and its convergence proved, in [29] where the network adjusts prices by solving a simple minimization problem based on the excess demand. Essentially the price of the resource with the largest excess is raised until there is no excess.

Note that all of the above schemes use only the excess demand and do not require the knowledge of individual utilities on the part of the network. Equilibrium can be approached under suitable conditions. It is hence important to examine the properties of an equilibrium, to which we now turn.

B. Equilibrium Prices and Allocations

It is well known that an equilibrium price (p, q) only gives relative prices of the resources in that if $(x_n, y_n, n = 1, \dots, N)$ is an equilibrium allocation at price (p, q) , then it is also an equilibrium allocation at price $(\lambda p, \lambda q)$ for all $\lambda > 0$. The first part of the following theorem then shows that for an equilibrium price (p, q) , $p_0 > 0$ and hence we can assume without loss of generality (after possibly a change of units) that $p_0 = R_0 = q_0 = B_0 = 1$.

Theorem 2: If (p, q) is an equilibrium price then

- i) $p_0 > 0, q_0 > 0$ and hence $X_0 = \bar{X}_0, Y_0 = \bar{Y}_0$. Moreover $R_0/p_0 = B_0/q_0$;
- ii) $p_1 \leq r$ or $q_1 \leq b$ after scaling so that $p_0 = R_0 = q_0 = B_0 = 1$.

Let $P = \{(p, q) \geq 0 \mid p_0 = q_0 = 1, p_1 \leq r, \text{ or } q_1 \leq b\}$. The theorem says that we can restrict equilibrium prices to the set P . Indeed if $c > 0$ then we can strengthen the result to " $p_1 \leq r$

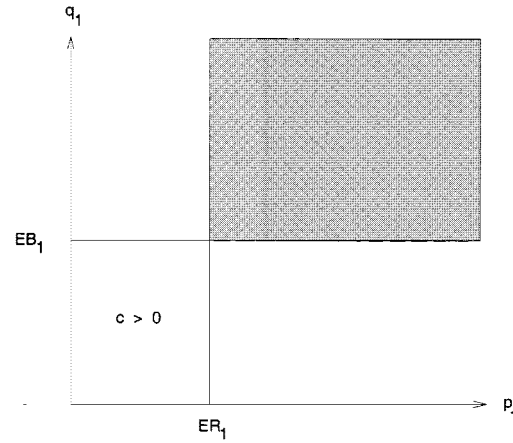


Fig. 1. Region of feasible equilibrium prices. Shaded region is not feasible. If $c > 0$ the feasible region is the lower left rectangle at the origin.

and $q_1 \leq b$," as shown in the next subsection.² The theorem is illustrated in Fig. 1.

To interpret the theorem we think of $E\bar{R}_i/p_i$ and $E\bar{B}_i/q_i$ as the *total return* on investing in resource i . Theorem 2 then implies that at equilibrium the total return on fixed resources is equal, and the total expected return on variable resources is generally higher. From assumption **A2** that variable resources cause a reduction in utility because of their variance. At equilibrium since all users achieve optimality the marginal increase in utility by exchanging one dollar worth of fixed resources for one dollar worth of variable resources must be balanced by the marginal decrease in utility due to increased variability.

It is then interesting to find that in equilibrium *every* user holds strictly positive amounts of variable bandwidth and buffer, as the next theorem shows.

Theorem 3: If $(p, q, x_n, y_n, n = 1, \dots, N)$ is an equilibrium with $p_1 > 0$ or $q_1 > 0$, then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

From Definition 1 equilibrium price $p_1 = 0$ or $q_1 = 0$ if there are excess variable resources. The qualification that " $p_1 > 0$ or $q_1 > 0$ " restricts the theorem to the situation where not both variable bandwidth and buffer are in excess. Indeed if $c < 0$ then the qualification is not necessary (Proposition 6).

The theorem has a similar flavor to the well known fact in the context of investment where it is optimal for every investor to diversify [24], [10], [17], [9]. The security models there however have an important difference: investors are allowed to hold short positions, i.e., (x_n, y_n) can be negative as well as positive. The case that allows negative allocations is treated in [12]. The nonnegativity constraint in our model complicates greatly the equilibrium analysis.

We now illustrate the results with a numerical example.

Example 2—Dynamics of Model M1: In this example, there are $N = 4$ users whose utility functions are

$$u_n(\mu, v) = -\log \frac{\mu^2}{v + \nu_n}$$

²The covariance c of R_1 and B_1 may be positive if, for example, R_1 and B_1 model the leftover resources not consumed by higher-priority traffics where a surge in higher-priority traffics leads to reduction in *both* R_1 and B_1 .

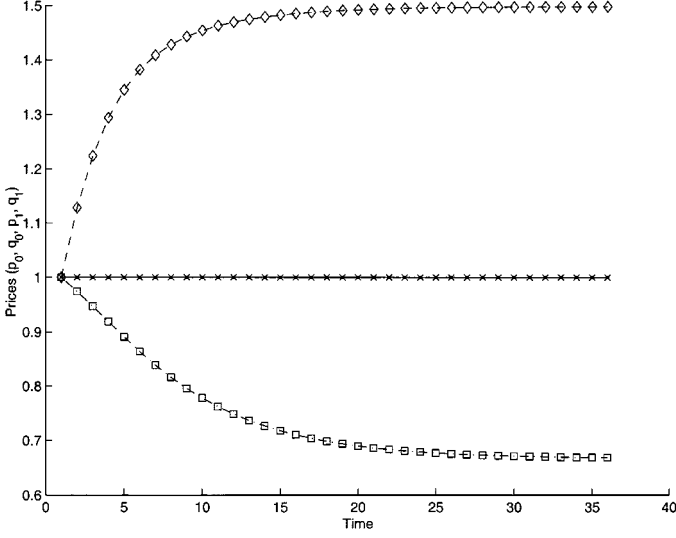


Fig. 2. Prices. Prices are normalized with $(p_0(t), q_0(t)) = (1, 1)$. Dash line with square is the price $p_1(t)$ for variable bandwidth and that with diamond is $q_1(t)$ for variable buffer.

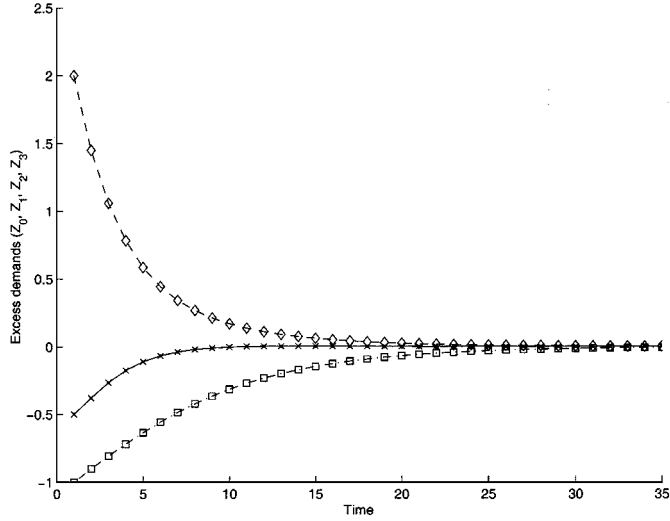


Fig. 3. Excess demands. Dash line with square is the excess demand for variable bandwidth and that with diamond is the excess demand for variable buffer. The excess demands for the fixed bandwidth and fixed buffer follow identical trajectory represented by the solid line.

where $\nu_n = n$. Every user has the same initial allocation $(\bar{x}_n, \bar{y}_n) = (1, 1, 1, 1)$, i.e., all resources are evenly allocated to the users initially. We will see that the resources will not be evenly distributed across users at equilibrium because of the difference in their utilities. We take $R_0 = B_0 = 1$ and $r = b = 4$, i.e., the variable resources have an expected value four times that of fixed resources. The variability is described by $\sigma_R^2 = 1.5$, $\sigma_B^2 = 1$, and $c = 0.5$. Hence bandwidth has a higher variability than buffer and they are positively correlated. Prices are adjusted according to (5) in proportion to excess demand with a constant of proportionality $\gamma = 0.05$.

Starting from an initial price of $(p(0), q(0)) = (1, 1, 1, 1)$, the system dynamic is illustrated in Figs. 2–5. Fig. 2 depicts the normalized prices with $p_0(t)$ and $q_0(t)$ set to 1 in each iteration. The price for variable buffer is higher than that for variable bandwidth because variable buffer has a less variance and hence

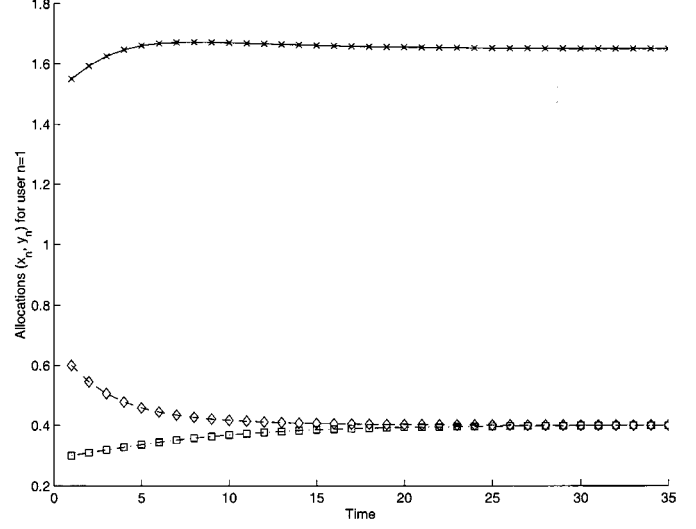


Fig. 4. Allocations of user $n = 1$. Dash line with square is the variable bandwidth allocation and that with diamond is the variable buffer allocation. The allocation for the fixed bandwidth and fixed buffer follow identical trajectory represented by the solid line.

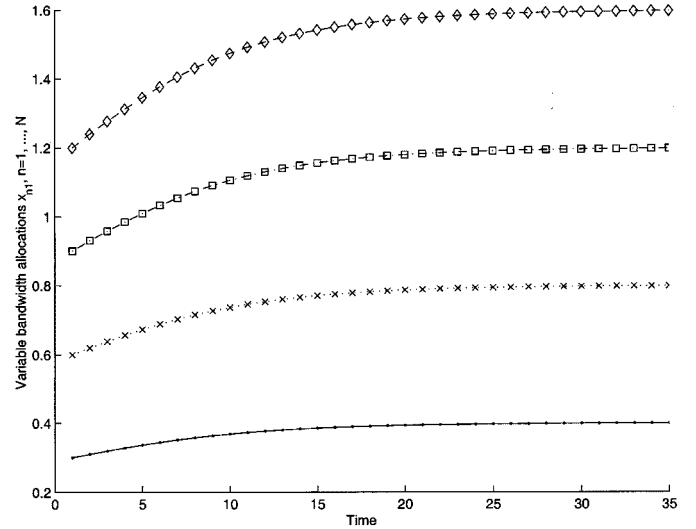


Fig. 5. Variable bandwidth allocations. The allocation for user 1 is the lowest line, that for user 2 is the second lowest line, ..., and that for user 4 is the highest line.

a higher demand. Fig. 3 shows that the excess demands all converge to zero as all prices are strictly positive. Fig. 4 describes the allocations for user $n = 1$. It is more interesting to compare allocations across users. User n has a strictly larger *marginal* decrease $|\partial u_n / \partial v| = 1/(v + \nu_n)$ in utility due to resource variability than user $n + 1$, and hence it favors less variable resources than user $n + 1$, as shown in Fig. 5.

Note that both Theorems 2 and 3 hold even though the utility functions do not satisfy assumption **A3**. The stability of the simple price adjustment scheme however can be sensitive to problem parameters. \square

C. Proof of Theorem 2

First, note that $f_n(x_n, y_n)$ is *strictly* increasing in (x_0, y_0) (but not necessarily in x_1, y_1). Hence if $p_0 = 0$ or $q_0 = 0$

then $X_0 > \bar{X}_0$ or $Y_0 > \bar{Y}_0$ respectively, violating Definition 1. Hence $p_0 > 0$, $q_0 > 0$ and $X_0 = \bar{X}_0$, $Y_0 = \bar{Y}_0$. If $p_0/R_0 > q_0/B_0$ then $x_{n0} = 0$ for all n , violating the definition of equilibrium. To see this, suppose a user n has $x_{n0} > 0$. It can change its allocation of fixed resources from (x_{n0}, y_{n0}) to $(0, y_{n0} + x_{n0}p_0/q_0)$ which increases μ without affecting v . This increases its utility f_n , contradicting that (x_n, y_n) is optimal for n . Hence $p_0/R_0 \leq q_0/B_0$. A similar argument shows that $p_0/R_0 \geq q_0/B_0$ and hence $p_0/R_0 = q_0/B_0$.

To prove part ii, scale the equilibrium price such that $p_0 = R_0 = q_0 = B_0 = 1$. If $p_1 > r$ and $q_1 > b$ then we claim that $x_{n1} = 0$ and $y_{n1} = 0$ for all n , violating the condition for equilibrium. Otherwise, if $x_{n1} > 0$ or $y_{n1} > 0$ for some n , then user n can sell all its variable resources and buy an additional amount $(x_{n1}p_1 + y_{n1}q_1)/p_0$ of fixed bandwidth. This strictly increases μ and decreases v , and hence increases the utility, contradicting the optimality of (x_n, y_n) .

Indeed if $c > 0$ then both $p_1 \leq r$ and $q_1 \leq b$. Otherwise suppose $p_1 > r$. Then a user with positive amount of x_{n1} can sell it and purchase an additional amount $x_{n1}p_1/p_0$ of fixed bandwidth. This increases the μ and decreases v , a contradiction.

D. Proof of Theorem 3

First note that it follows directly from assumption **A2** and (2)–(3) that the objective function $f_n(x, y)$ is concave in (x, y) , and hence the first order optimality condition is both necessary and sufficient.

Define the Lagrangian function

$$\begin{aligned} L(x_n, y_n, \gamma_n, \hat{\eta}_n, \hat{\lambda}_n) \\ = f_n(x_n, y_n) - \gamma_n(px_n + qy_n - p\bar{x}_n - q\bar{y}_n) \\ + \hat{\eta}_n x_n + \hat{\lambda}_n y_n. \end{aligned}$$

Then $(x_n, y_n) \geq 0$ is optimal for user n if and only if there exist $\gamma_n, \hat{\eta}_n \geq 0$, and $\hat{\lambda}_n \geq 0$ such that [2]

$$\begin{aligned} \nabla_x f_n - \gamma_n p + \hat{\eta}_n &= 0 \\ \nabla_y f_n - \gamma_n q + \hat{\lambda}_n &= 0 \\ px_n + qy_n &= p\bar{x}_n + q\bar{y}_n \\ \hat{\eta}_n x_n &= 0, \quad \hat{\lambda}_n y_n = 0, \quad i = 0, 1. \end{aligned}$$

In view of Theorem 2 we will fix for the rest of this paper that $p_0 = R_0$ and $q_0 = B_0$ for an equilibrium price (p, q) . Define

$$\beta_n(x, y) = \frac{\frac{\partial u_n}{\partial \mu}(\mu(x, y), v(x, y))}{-2\frac{\partial u_n}{\partial v}(\mu(x, y), v(x, y))} > 0$$

$$\begin{aligned} a_1 &= r - p_1 \\ a_2 &= b - q_1 \\ A &= \begin{bmatrix} \sigma_R^2 & c \\ c & \sigma_B^2 \end{bmatrix}. \end{aligned}$$

Then expanding ∇f_n in the Karush–Kuhn–Tucker condition, eliminating γ_n , and noting $p_0\hat{\lambda}_{n0} = q_0\hat{\eta}_{n0}$, we have that $(x_n, y_n) \geq 0$ is optimal for user n if and only if

$$A \begin{bmatrix} x_{n1} \\ y_{n1} \end{bmatrix} = \beta_n(x_n, y_n) \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (6)$$

$$- \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} \eta_{n0} + \begin{bmatrix} \eta_{n1} \\ \lambda_{n1} \end{bmatrix} \quad (7)$$

$$px_n + qy_n = p\bar{x}_n + q\bar{y}_n \quad (8)$$

$$\eta_{n0}x_{n0} = 0, \quad \eta_{n0}y_{n0} = 0 \quad (9)$$

$$\eta_{n1}x_{n1} = 0, \quad \lambda_{n1}y_{n1} = 0 \quad (10)$$

for some $\eta_{n0} \geq 0$, $\eta_{n1} \geq 0$, and $\lambda_{n1} \geq 0$. Moreover, by assumption **A1** the matrix A in (7) is invertible and its determinant $|A| = \sigma_R^2\sigma_B^2 - c^2 > 0$ (Schwarz inequality). Finally define ζ_i by

$$\begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} = A^{-1} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} \sigma_B^2 a_1 - ca_2 \\ \sigma_R^2 a_2 - ca_1 \end{bmatrix}. \quad (11)$$

The next lemma is a consequence of assumption **A3**. It says that it is never optimal for a user n to reserve only variable bandwidth without fixed bandwidth nor any buffer, fixed or variable, nor is it optimal to reserve only variable buffer without fixed buffer nor any bandwidth, fixed or variable.

Lemma 4: Under assumption **A3** there exists no price $(p, q) \geq 0$ with $(p, q) \neq 0$ for which $(x_n, y_n) = (0, 0, x_{n1}, 0)$ or $(x_n, y_n) = (0, 0, 0, y_{n1})$ is optimal for user n .

Proof: Suppose $(x_n, y_n) = (0, 0, x_{n1}, 0)$ is optimal for a user n under price $(p, q) \geq 0$. We will show that this contradicts assumption **A3**. The proof that $(x_n, y_n) = (0, 0, 0, y_{n1})$ cannot be optimal follows a similar argument.

Since $(p, q) \neq 0$ and $\bar{x}_n > 0$ and $\bar{y}_n > 0$, we have from (8)

$$x_{n1} = \frac{p\bar{x}_n + q\bar{y}_n}{p_1} \geq \bar{x}_{n1} > 0. \quad (12)$$

Since $x_{n1} > 0$, $\eta_{n1} = 0$ by complementary slackness, and hence from (7)

$$\sigma_R^2 x_{n1} = \beta_n(x_n, y_n) a_1 - p_1 \eta_{n0}$$

or

$$\beta_n(x_n, y_n) a_1 \geq \sigma_R^2 x_{n1}.$$

But $0 < a_1 = r - p_1 \leq r$. Hence

$$\beta_n(x_n, y_n) \geq x_{n1} \frac{\sigma_R^2}{r}.$$

Substituting $(x_n, y_n) = (0, 0, x_{n1}, 0)$ into β_n and noting that

$$\mu(x_n, y_n) = x_{n1}r, \quad \text{and} \quad v(x_n, y_n) = x_{n1}^2 \sigma_R^2$$

we have

$$\frac{\frac{\partial u_n}{\partial \mu}(\mu(x_n, y_n), v(x_n, y_n))}{-\frac{\partial u_n}{\partial v}(\mu(x_n, y_n), v(x_n, y_n))} \geq 2 \frac{v(x_n, y_n)}{\mu(x_n, y_n)}.$$

But $v(x_n, y_n) = x_{n1}^2 \sigma_R^2 \geq \bar{x}_{n1}^2 \sigma_R^2$ by (12), contradicting assumption **A3**. This completes the proof of the lemma. \square

The next lemma is used in the proof of Proposition 6. It says in particular that if $c < 0$ then an equilibrium allocation (x_n, y_n) cannot have $x_{n1} = 0$ and $y_{n1} = 0$. This is a weaker form of Proposition 6 which asserts that indeed *both* x_{n1} and y_{n1} are strictly positive, for all n .

Lemma 5: Suppose $c < 0$. If $(x_n, y_n, n = 1, \dots, N)$ is an equilibrium allocation, then

- i) $x_{n1} > 0$ or $y_{n1} > 0$;
- ii) $\sigma_R^2 x_{n1} + cy_{n1} > 0$ or $cx_{n1} + \sigma_B^2 y_{n1} > 0$.

Proof: i) Suppose for the contrary that $x_{n1} = 0$ and $y_{n1} = 0$. Since $p_0 > 0$ and $q_0 > 0$ at equilibrium, (8) implies that $x_{n0} > 0$ or $y_{n0} > 0$, which then implies that $\eta_{n0} = 0$ by complementary slackness. From (7) we have $a_1 = a_2 = 0$, or $p_1 = r > 0$ and $q_1 = b > 0$. Then for *any* user m , (7) implies

$$\begin{bmatrix} x_{m1} \\ y_{m1} \end{bmatrix} = A^{-1} \left(- \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} \eta_{m0} + \begin{bmatrix} \eta_{m1} \\ \lambda_{m1} \end{bmatrix} \right) \quad (13)$$

since $c < 0$ entries of A^{-1} are all positive. (13) then implies that $\eta_{m0} = 0$. Otherwise, $x_{m0} = y_{m0} = 0$, and hence neither “ $x_{m1} = y_{m2} = 0$ ” (contradicting the budget constraint) nor “ $x_{m1} > 0$ and $y_{m2} > 0$ ” [contradicting (13)] nor “exactly one of x_{m1} and y_{m1} is positive” (contradicting Lemma 4). But $\eta_{m0} = 0$ in (13) can only be satisfied by $x_{m1} = y_{m1} = 0$ because of complementary slackness. Thus we have proved that if there exists n such that $x_{n1} = y_{n1} = 0$, then $X_1 = Y_1 = 0$, contradicting that $p_1 > 0$ and $q_1 > 0$. Hence for all n , $x_{n1} > 0$ or $y_{n1} > 0$.

ii) Suppose not. Then

$$-cy_{n1} \geq \sigma_R^2 x_{n1} \quad (14)$$

$$-cx_{n1} \geq \sigma_B^2 y_{n1}. \quad (15)$$

If $x_{n1} = 0$ then $y_{n1} > 0$ by i), contradicting (15). Similarly $x_{n1} > 0$ and $y_{n1} = 0$ contradicts (14). If $x_{n1} > 0$ and $y_{n1} > 0$, then $c^2 \geq \sigma_R^2 \sigma_B^2$, a contradiction. This completes the proof. \square

Theorem 3, which says that at equilibrium every user desires strictly positive amounts of random resources, is proved by the next two propositions, one for the case $c < 0$ and the other for $c > 0$. The main idea in the proof of the next proposition is that if x_{n1} and y_{n1} are not both positive then x_{n0} and y_{n0} are both zero, contradicting Lemma 4.

Proposition 6: Suppose $c < 0$. If $(x_n, y_n, n = 1, \dots, N)$ is an equilibrium allocation, then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

Proof: We start with a simple lemma that says that when R_1 and B_1 are negatively correlated then there is *some* user that holds strictly positive amounts of variable bandwidth and buffer.

Lemma 7: Suppose $c < 0$. If $(x_n, y_n, n = 1, \dots, N)$ is an equilibrium allocation, then $X_1 > 0$ and $Y_1 > 0$.

Proof: Fix an equilibrium $(p, q, x_n, y_n, n = 1, \dots, N)$. Suppose for the sake of contradiction that $X_1 = 0$, i.e., $x_{n1} = 0$ for all n . By Lemma 4, $x_{n0} > 0$ or $y_{n0} > 0$, implying that $\eta_{n0} = 0$. Moreover $p_1 = 0$ since $\bar{X}_1 > 0$ and $p_1(X_1 - \bar{X}_1) = 0$. Hence $a_1 = r - p_1 = r > 0$. Substituting $x_{n1} = 0$, $\eta_{n0} = 0$, $p_1 = 0$ and $a_1 = r$ into (7) we have

$$cy_{n1} \geq \beta_n(x_n, y_n)r > 0$$

which is impossible since $c < 0$. Hence $X_1 > 0$. That $Y_1 > 0$ is proved by a similar argument. \square

Fix an equilibrium $(p, q, x_n, y_n, n = 1, \dots, N)$. From Theorem 2 at least one of

$$a_1 = r - p_1 \quad \text{and} \quad a_2 = b - q_1$$

must be nonnegative. We will consider three alternatives separately: $a_i \geq 0$, $i = 1, 2$ (proved in Lemma 8 below); $a_1 \leq 0$ (proved in Lemma 9 below); and $a_2 \leq 0$ (symmetric argument omitted).

From (7) we have

$$\sigma_R^2 x_{n1} + cy_{n1} = \beta_n(x_n, y_n)a_1 - p_1\eta_{n0} + \eta_{n1} \quad (16)$$

$$cx_{n1} + \sigma_B^2 y_{n1} = \beta_n(x_n, y_n)a_2 - q_1\eta_{n0} + \lambda_{n1} \quad (17)$$

for some $\eta_{n0} \geq 0$, $\eta_{n1} \geq 0$, $\lambda_{n1} \geq 0$.

Lemma 8: Suppose $c < 0$. If $a_i \geq 0$ for $i = 1, 2$ then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

Proof: By Lemma 5-ii, (16) > 0 or (17) > 0 . Suppose (16) > 0 . Then $x_{n1} > 0$. If in addition (17) ≥ 0 then $y_{n1} > 0$ since $cx_{n1} < 0$. Otherwise if (17) < 0 then $\eta_{n0} > 0$ since $a_2 \geq 0$ and $\lambda_{n1} \geq 0$ in the right-hand side of (17). Complementary slackness then implies that $x_{n0} = y_{n0} = 0$. By Lemma 4, $y_{n1} > 0$. Hence if (16) > 0 then $x_{n1} > 0$ and $y_{n1} > 0$.

A similar argument shows that, in the second case (17) > 0 , we also have $x_{n1} > 0$ and $y_{n1} > 0$. \square

Lemma 9: Suppose $c < 0$. If $a_1 \leq 0$ then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

Proof: We first claim that $\sigma_R^2 x_{n1} + cy_{n1} \leq 0$. Suppose not. Then both sides of (16) are strictly positive. Since $c < 0$ we must have $x_{n1} > 0$. Since $a_1 \leq 0$ and $\eta_{n0} \geq 0$ we must have $\eta_{n1} > 0$, contradicting the complementary slackness condition $\eta_{n1}x_{n1} = 0$.

Hence $\sigma_R^2 x_{n1} + cy_{n1} \leq 0$. But this and Lemma 5-ii imply that $cx_{n1} + \sigma_B^2 y_{n1} > 0$, which in turn implies that $y_{n1} > 0$ since $cx_{n1} \leq 0$. Moreover this is true for all n . We are thus left to prove that $x_{n1} > 0$ for all n .

By Lemma 7 there exists m such that both y_{m1} and x_{m1} are strictly positive. Then $\eta_{m1} = \lambda_{m1} = 0$ and hence from (7)

$$\begin{bmatrix} x_{m1} \\ y_{m1} \end{bmatrix} = \beta_m(x_m, y_m) \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} - A^{-1} \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} \eta_{m0} > 0.$$

The last term is nonpositive since $c < 0$. Hence $\zeta_i > 0$ for $i = 1, 2$.

Indeed we claim that for any n if $x_{n1} = 0$ then $x_{n0} = y_{n0} = 0$, contradicting Lemma 4. To see this, note that

$$\begin{bmatrix} x_{n1} \\ y_{n1} \end{bmatrix} = \beta_n(x_n, y_n) \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} + A^{-1} \left(- \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} \eta_{n0} + \begin{bmatrix} \eta_{n1} \\ 0 \end{bmatrix} \right)$$

since $c < 0$, $A^{-1} > 0$. Hence $x_{n0} = 0$ and $\zeta_1 > 0$ imply that $\eta_{n0} > 0$. This means that $x_{n0} = y_{n0} = 0$ by the complementary slackness condition (9), contradicting Lemma 4.

Hence $x_{n1} > 0$ and $y_{n1} > 0$ for all users n . \square

The case where $a_2 \leq 0$ follows a similar argument as in Lemma 9. Hence the proof of Proposition 6 is complete. \square

From Definition 1, equilibrium price $p_1 = 0$ or $q_1 = 0$ if there are excess variable resources. When not both variable bandwidth and buffer are in excess, we have the following result. The main idea in the proof is that if x_{n1} and y_{n1} are not both positive then either the equilibrium condition that demand equals supply for nonfree resources is violated or Lemma 4 or assumption **A4** is violated.

Proposition 3: Suppose $c > 0$. If $(p, q, x_n, y_n, n = 1, \dots, N)$ is an equilibrium such that not both $p_1 = 0$ and $q_1 = 0$, then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

Proof: For every user n , there are three mutually exclusive cases: 1) $x_{n1} = y_{n1} = 0$; 2) either $x_{n1} = 0$ or $y_{n1} = 0$ but not both; and 3) $x_{n1} > 0$ and $y_{n1} > 0$. We will show that cases 1) and 2) are impossible, and hence $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

Case 1: $\exists n$ such that $x_{n1} = y_{n1} = 0$. From Theorem 2 we know that $a_1 = r - p_1 \geq 0$. If $a_1 > 0$ then since the left-hand side of (16) equals zero we have $\eta_{n0} > 0$. Complementary slackness then implies $x_{n0} = y_{n0} = 0$, contradicting Theorem 2-i. Hence $a_1 = 0$. But then for every m we have from (16)

$$\sigma_R^2 x_{m1} + c y_{m1} = -p_1 \eta_{m0} + \eta_{m1}$$

imposing that $x_{m1} = 0$ for every m . Hence $X_1 = \sum_m x_{m1} = 0$. But since $a_1 = 0$, $p_1 = r > 0$ contradicting that $p_1(X_1 - \bar{X}_1) = 0$ at equilibrium. Hence for no n does $x_{n1} = y_{n1} = 0$.

Case 2: $\exists n$ such that either $x_{n1} = 0$ or $y_{n1} = 0$ but not both. We will consider $x_{n1} > 0$ and $y_{n1} = 0$; the other case is symmetric.

Since $y_{n1} = 0$, by Lemma 4, $x_{n0} > 0$ or $y_{n0} > 0$, implying that $\eta_{n0} = 0$. Hence from (7)

$$\begin{bmatrix} \sigma_R^2 x_{n1} \\ c x_{n1} \end{bmatrix} = \beta_n(x_n, y_n) \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda_{n1} \end{bmatrix}$$

yielding

$$x_{n1} = \beta_n(x_n, y_n) a_1 / \sigma_R^2 = \beta_n(x_n, y_n) a_2 / c + \lambda_{n1} / c. \quad (18)$$

Since $\lambda_{n1} \geq 0$ we have

$$\sigma_R^2 a_2 - c a_1 \leq 0. \quad (19)$$

Substituting $a_1 = r - p_1$ and $a_2 = b - q_1$ we have

$$\sigma_R^2 b - c r \leq \sigma_R^2 q_1 - c p_1. \quad (20)$$

Now we claim that $y_{m1} = 0$ for all users m . Suppose not. Then for some m we have $y_{m1} > 0$ and hence $\lambda_{m1} = 0$. From (7) and (11) we get

$$\begin{aligned} y_{m1} &= \beta_m(x_m, y_m) \zeta_2 - (\sigma_R^2 q_1 - c p_1) \eta_{m0} / |A| \\ &= \beta_m(x_m, y_m) \left[(\sigma_R^2 b - c r) - \left(1 + \frac{\eta_{m0}}{|A| \beta_m(x_m, y_m)} \right) \cdot (\sigma_R^2 q_1 - c p_1) \right] \leq 0 \end{aligned}$$

where the last inequality follows from (20) and the fact that $\eta_{n0} / |A| \beta_n(x_n, y_n) \geq 0$. Hence $y_{m1} = 0$ for all m and $Y_1 = \sum_m y_{m1} = 0$.

This implies $q_1 = 0$ since at equilibrium $q_1(Y_1 - \bar{Y}_1) = -q_1 \bar{Y}_1 = 0$. By assumption of the theorem $p_1 > 0$ and $X_1 = \bar{X}_1$. This implies the existence of m such that

$$x_{m1} \geq \bar{x}_{m1}.$$

For this m we have from (18), noting that $a_1 = r - p_1 > 0$ from (19) and that $a_2 = b > 0$,

$$\beta_m(x_m, y_m) = x_{m1} \frac{\sigma_R^2}{a_1} \geq \bar{x}_{m1} \frac{\sigma_R^2}{a_1} \geq \bar{x}_{m1} \frac{\sigma_R^2}{R_1}. \quad (21)$$

Now since $y_{m1} = 0$,

$$\begin{aligned} \mu(x_m, y_m) &= x_{m1} r + x_{m0} R_0 + y_{m0} B_0 \\ v(x_m, y_m) &= x_{m1}^2 \sigma_R^2 \end{aligned}$$

and hence

$$\frac{v(x_m, y_m)}{\mu(x_m, y_m)} \leq x_{m1} \frac{\sigma_R^2}{R_1} \leq \bar{x}_{m1} \frac{\sigma_R^2}{R_1} \leq \beta_m(x_m, y_m) \quad (22)$$

from (21). But $v(x_m, y_m) = \bar{x}_{m1}^2 \sigma_R^2$ and hence (22) contradicts assumption **A4**. This proves that Case 2 is impossible.

Hence $x_{n1} > 0$ and $y_{n1} > 0$ for all n and the proposition is proved. \square

Propositions 6 and 10 thus prove Theorem 3.

IV. EQUILIBRIUM OF **M2**

In this section we first discuss how an equilibrium might be approached, and then present and prove some equilibrium properties.

A. Convergence

The equilibrium situation with model **M2** where user optimization has no budget constraint is simpler. We will relate the noncooperative individual optimizations to a single-objective welfare maximization. This not only proves the existence of an equilibrium, but also provides a nice interpretation for the law of supply and demand as a price adjustment scheme to approach an equilibrium.

Consider the welfare maximization

$$\max_{(x_n, y_n, n=1, \dots, N) \geq 0} \sum_n f_n(x_n, y_n) \quad (23)$$

$$\text{subject to } X_i \leq \bar{X}_i, Y_i \leq \bar{Y}_i. \quad (24)$$

The dual problem is (see e.g., [3, Section 3.4.2], [14])

$$\min_{(p, q) \geq 0} D(p, q) = \sum_s B_n(p, q) + (p \bar{X} + q \bar{Y}) \quad (25)$$

where

$$B_n(p, q) = \max_{(x_n, y_n) \geq 0} f_n(x_n, y_n) - (p x_n + q y_n). \quad (26)$$

Proposition 11: The set of equilibrium allocations is nonempty and convex. Moreover, $(p^*, q^*, x_n^*, y_n^*, n =$

$1, \dots, N$) is an equilibrium if and only if $(x_n^*, y_n^*, n = 1, \dots, N)$ solves the welfare maximization (23)–(24) and (p^*, q^*) solves its dual problem (25)–(26).

Proof: Since the objective function of the welfare maximization is continuous and the feasible set is compact, a maximizer exists. Moreover, since it is concave, $(x_n^*, y_n^*, n = 1, \dots, N)$ is a maximizer of (23)–(24) if and only if it satisfies the first-order optimality condition with a corresponding Lagrange multiplier (p^*, q^*) . But it can be verified that this condition is equivalent to Definition 1, and hence $(p^*, q^*, x_n^*, y_n^*, n = 1, \dots, N)$ is an equilibrium. The convexity of the set of equilibria follows from the concavity of the welfare maximization. \square

The reformulation of user problems in terms of a welfare maximization also suggests a natural price adjustment scheme to approach an equilibrium. Note that (26) is equivalent to the user maximization in **M2**, and that at price (p^*, q^*) that is dual optimal, the price-allocation pair $(p^*, q^*, x_n(p^*, q^*), y_n(p^*, q^*), n = 1, \dots, N)$ is primal-dual optimal by the duality theory, where $(x_n(p^*, q^*), y_n(p^*, q^*))$ is a maximizer of user n 's problem. As noted in the proof of the proposition, it is hence also an equilibrium in the sense of Definition 1. Hence we will approach an equilibrium by having the network iteratively compute a dual-optimal price vector and users respond to it according to **M2**.

Indeed we will compute a dual-optimal price vector using the gradient projection algorithm (e.g., [14], [3]) where the prices are adjusted in opposite direction of the gradient $\nabla D(p, q)$:

$$(p(t+1), q(t+1)) = [(p(t), q(t)) - \gamma \nabla D(p, q)]^+.$$

Here $\gamma > 0$ is a step size, and $(z)^+ = \max\{z, 0\}$. But the gradient turns out to be just the negative of excess demand [3, pp. 669] $\nabla D(p, q) = -Z(t)$ where

$$Z(t) = ((X(p(t), q(t)), Y(p(t), q(t))) - (\bar{X}, \bar{Y}))$$

and

$$\begin{aligned} & (X(p(t), q(t)), Y(p(t), q(t))) \\ &= \left(\sum_n x_n(p(t), q(t)), \sum_n y_n(p(t), q(t)) \right) \end{aligned}$$

are the aggregate demands at prices $(p(t), q(t))$. Hence the price adjustment rule becomes

$$(p(t+1), q(t+1)) = [(p(t), q(t)) + \gamma Z(t)]^+.$$

This is just the law of supply and demand: if the excess demand for a resource is positive, raise its price; otherwise reduce it. A dual optimal price vector is obtained when excess demands are reduced to zero for nonfree resources. Hence this scheme to approach an equilibrium (in the sense of Definition 1) can be thought of as simply the gradient projection algorithm to solve the dual (and hence the primal welfare maximization) problem. The algorithm converges provided that the gradient $\nabla D(p, q)$ of the dual objective function $D(p, q)$ is Lipschitz continuous [3, pp. 214].

B. Equilibrium Prices and Allocations

Unlike in model **M1**, here the equilibrium prices are not relative. But $(p_0, q_0) > 0$ as before and hence for fixed resources supply equals demand. Moreover the total return on fixed resources are the same.

Theorem 12: Let $(p, q, x_n, y_n, n = 1, \dots, N)$. Then $p_0 > 0, q_0 > 0$, and hence $X_0 = \bar{X}_0$ and $Y_0 = \bar{Y}_0$. Moreover $R_0/p_0 = B_0/q_0$.

Under Assumption **A3** the positive equilibrium allocation of variable resources holds.

Theorem 13: If $(x_n, y_n, n = 1, \dots, N)$ is an equilibrium allocation then $x_{n1} > 0$ and $y_{n1} > 0$ for all n .

If we replace the strict inequality in assumption **A3** by inequality,

A3': For all $(\mu, v) > 0$,

$$\frac{\frac{\partial u_n}{\partial \mu}(\mu, v)}{-\frac{\partial u_n}{\partial v}(\mu, v)} \leq 2 \frac{v}{\mu}.$$

then the equilibrium situation becomes more complicated. There are still users that hold strictly positive amounts of both variable bandwidth and variable buffer, but there may be users that hold zero variable resources. Indeed some users can be “priced out” of the market at equilibrium, i.e., $(x_n, y_n) = (0, 0)$. To express precisely the equilibrium situation on the allocation of variable resources, let us define (by slight abuse of notation) $N = \{1, \dots, N\}$ to be the set of all users, and partition N into four disjoint subsets according to their equilibrium allocation of *variable* bandwidth and buffer:

$$N = N_0 \cup N_1 \cup N_{11}$$

where N_0 is the set of users n that hold zero variable bandwidth and zero variable buffer $x_{n1} = y_{n1} = 0$, N_1 is the set of users n whose allocations are such that exactly one of x_{n1} and y_{n1} is strictly positive, and N_{11} is the set of users n with $x_{n1} > 0$ and $y_{n1} > 0$.

Theorem 14: Suppose $(p, q, x_n, y_n, n \in N)$ is an equilibrium allocation. If Assumption **A3** is replaced by **A3'** then

- i) $N_{11} \neq \emptyset$;
- ii) at most one of N_0 and N_1 can be nonempty;
- iii) if N_0 is nonempty then $p_1 > 0$ and $q_1 > 0$;
- iv) if N_1 is nonempty then $p_1 = 0$ or $q_1 = 0$, or both.

Hence there will always be users who desire both variable resources as in model **M1**. For the other users either *none* holds any variable resources or *every* user holds exactly one variable resource. In the former case, supply equals demand for all resources; in the latter case, there will be free variable resource. In particular if both variable resources are scarce ($p_1 > 0$ and $q_1 > 0$) then $N = N_0 \cup N_{11}$, i.e., a user either desires both variable resources or none.

C. Proof of Theorem 14

The proof of Theorems 12 and 13 is simple and omitted. We now prove Theorem 14. The optimality condition for user n is:

$(x_n, y_n) \geq 0$ is optimal if and only if there exists $(\eta_n, \lambda_n) \geq 0$ such that

$$R_0 \frac{\partial u_n}{\partial \mu} - p_0 + \eta_{n0} = 0 \quad (27)$$

$$B_0 \frac{\partial u_n}{\partial \mu} - q_0 + \lambda_{n0} = 0 \quad (28)$$

$$r \frac{\partial u_n}{\partial \mu} + 2(\sigma_R^2 x_{n1} + c y_{n1}) \frac{\partial u_n}{\partial v} - p_1 + \eta_{n1} = 0 \quad (29)$$

$$b \frac{\partial u_n}{\partial \mu} + 2(c x_{n1} + \sigma_B^2 y_{n1}) \frac{\partial u_n}{\partial v} - q_1 + \lambda_{n1} = 0 \quad (30)$$

$$\eta_{ni} x_{ni} = 0, \quad \lambda_{ni} y_{ni} = 0, \quad i = 0, 1 \quad (31)$$

The basic observation is:

Lemma 15: Suppose (x_n, y_n) is optimal for user n at prices (p, q) .

- i) If $x_{n1} > 0$ then either $y_{n1} > 0$ or $p_1 = 0, x_{n0} = y_{n0} = 0$;
- ii) If $y_{n1} > 0$ then either $x_{n1} > 0$ or $q_1 = 0, x_{n0} = y_{n0} = 0$;
- iii) $(p_0, q_0) > 0$. If $x_{n1} = y_{n1} = 0$ then $(p_1, q_1) > 0$.

Proof: For part i, suppose $x_{n1} > 0$ and $y_{n1} = 0$. From (29) and (31) we have

$$\beta_n(x_n, y_n) = \frac{\sigma_R^2 x_{n1}}{r} + \frac{p_1}{r} \frac{1}{-2 \frac{\partial u_n}{\partial v}}.$$

Now, $v(x_n, y_n) = \sigma_R^2 x_{n1}^2$ and $\mu(x_n, y_n) = x_{n1}r + x_{n0}R_0 + y_{n0}B_0$. Condition A3' requires that $\beta_n \leq v/\mu$ and hence

$$\frac{\sigma_R^2 x_{n1}}{r} + \frac{p_1}{r} \frac{1}{-2 \frac{\partial u_n}{\partial v}} \leq \frac{\sigma_R^2 x_{n1}^2}{x_{n1}r + x_{n0}R_0 + y_{n0}B_0}$$

which can be satisfied if and only if $p_1 = 0, x_{n0} = y_{n0} = 0$. Part ii can be proved in a similar manner. Part iii follows from (27)–(30) and A2. \square

We can now prove Theorem 14.

Proof (Theorem 14): We will prove in the reverse order. Part iv follows from Lemma 15-i and ii. Part iii follows from Lemma 15-iii. For part ii, suppose $N_0 \neq \phi$ and hence $(p_1, q_1) > 0$. But then Lemma 15-i and ii imply that, for all users n , $x_{n1} > 0 \Leftrightarrow y_{n1} > 0$, i.e., $n \in N_0 \cup N_{11}$. Hence $N_1 = \phi$.

To prove part i, note that at equilibrium since $(p_0, q_0) > 0$ there exists a user n with $x_{n0} > 0$ or $y_{n0} > 0$. Such an n must satisfy $x_{n1} > 0 \Leftrightarrow y_{n1} > 0$ by Lemma 15-i and ii, i.e., if $x_{n0} > 0$ or $y_{n0} > 0$ then $n \in N_0 \cup N_{11}$. If some of such users n are in N_{11} then the theorem is proved. If all such users n are in N_0 , then we claim that there exists a user m with $x_{m0} = y_{m0} = 0$ that is in N_{11} .

To see this, note that $N_0 \neq \phi$ and hence $(p_1, q_1) > 0$ by Lemma 15-iii and $N_1 = \phi$ by part ii of the theorem. But these imply that $X_1 > 0$ and $Y_1 > 0$ and moreover $x_{m1} > 0 \Leftrightarrow y_{m1} > 0$. Hence $N_{11} \neq \phi$. \square

V. CONCLUSION

We have described a model for resource allocation schemes that provide a user with a fixed and a variable share of bandwidth and buffer. When these users interact by exchanging their resource allocations in order to maximize their benefit, an equilibrium results at which the marginal increase in utility due to the higher expected return of variable resources is balanced by the marginal decrease in utility due to their variability. We have studied two models of user behavior. For the first model we have shown that at equilibrium all users desire strictly positive amounts of variable bandwidth and buffer. This implies that ABR or variable bit rate (VBR) may be more suitable than continuous bit rate (CBR) services for elastic traffics. The characterization of equilibrium prices may also be useful in the design of price adjustment scheme. For example prices can first be computed according to the law of supply and demand and then projected into the feasible domain of equilibrium prices. An example in [12] shows that this can significantly speed up convergence toward an equilibrium. For the second model we show that there are still users who desire both variable resources, but there will also be users who hold none or one of the variable resources depending on the scarcity of the variable resources. When both variable bandwidth and buffer are scarce every user either holds both or none. Moreover, some users can be squeezed out of service completely.

This paper presents a first step toward understanding the equilibrium situation of a network shared by a large number of users under such resource allocation schemes. As we have seen, some of the properties are not intuitive and their analysis can be rather intricate but tedious. Whether equilibrium analysis can be fruitfully exploited in the design and management of networks, as it is in economics [15], remains to be seen.

ACKNOWLEDGMENT

The present much simpler proof for Theorem 2 is due to an anonymous reviewer. The author would like to thank A. Greenberg for his hospitality while a portion of this work was being performed at AT&T Laboratories.

REFERENCES

- [1] S. Athuraliya, D. Lapsley, and S. Low, "An enhanced random early marking algorithm for Internet flow control," in *Proc. IEEE Infocom*, Tel Aviv, Israel, Mar. 2000.
- [2] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
- [3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [4] A. K. Choudhury and E. L. Hahne, "Dynamic queue length thresholds for shared-memory packet switches," *IEEE/ACM Trans. Networking*, vol. 6, pp. 130–140, Apr. 1998.
- [5] G. Debreu, "Excess demand functions," *J. Math. Econ.*, vol. 1, pp. 15–22, 1974.
- [6] N. Duffield and S. Low, "The cost of quality in networks of aggregate traffic," in *Proc. IEEE Infocom'98*, San Francisco, CA, Mar. 1998, pp. 525–532.
- [7] A. Elwalid, D. Mitra, and R. Wentworth, "A new approach for allocating buffer and bandwidth to heterogeneous, regulated traffic in an ATM node," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1115–1127, Aug. 1995.
- [8] S. J. Golestani, "A self-clocked fair queueing scheme in high speed applications," in *Proc. Infocom'94*, 1994, pp. 636–646.

- [9] R. A. Jarrow, *Finance Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [10] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Rev. Econ. and Statist.*, vol. 47, pp. 13–37, 1965.
- [11] S. H. Low and P. P. Varaiya, "Burst reducing servers in ATM networks," *Queueing Syst.*, vol. 20, pp. 61–84, 1995.
- [12] S. H. Low, "Equilibrium allocation and pricing of variable resources among user-suppliers," *Perform. Eval.*, vol. 34, no. 4, pp. 207–225, Dec. 1998.
- [13] S. H. Low and D. E. Lapsley, "Optimization flow control—I: Basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, pp. 861–874, Dec. 1999.
- [14] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison-Wesley, 1984.
- [15] A. S. Manne, Ed., *Economic Equilibrium: Model Formulation and Solution*: Math. Prog. Soc., 1985, vol. 23, Math. Prog. Study.
- [16] D. Mitra and I. Ziedins, *Virtual Partitioning by Dynamic Priorities: Fair and Efficient Resource-Sharing by Several Services*. ser. Lecture Notes in Comp. Sci. (Proc. Int. Zurich Sem. Digital Commun.), B. Plattner, Ed. New York, NY: Springer, 1996.
- [17] J. Mossin, "Equilibrium in a capital asset market," *Econometrica*, vol. 34, pp. 768–783, 1965.
- [18] T. Negishi, "Welfare economics and existence of an equilibrium for a competitive economy," *Metroeconomica*, vol. XII, pp. 92–97, 1960.
- [19] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks—The single node case," *IEEE/ACM Trans. Networking*, vol. 1, pp. 344–357, June 1993.
- [20] F. L. Presti, Z. Zhang, J. Kurose, and D. Towsley, "Source time scale and optimal buffer/bandwidth trade-off for regulated traffic in an ATM node," in *Proc. Infocom'97*, Apr. 1997.
- [21] J. Quirk and R. Saposnik, *Introduction to General Equilibrium Theory and Welfare Economics*. New York, NY: McGraw-Hill, 1968.
- [22] S. Sathaye, "Traffic Management Specification v 4.0," ATM Forum Traffic Management Group, Oct. 1996.
- [23] H. Scarf, "Some examples of global instability of the competitive equilibrium," *Int. Econ. Rev.*, vol. 1, pp. 157–172, 1960.
- [24] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, pp. 425–442, 1964.
- [25] S. Shenker, "Fundamental design issues for the future Internet," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1176–1188, July 1995.
- [26] S. Smale, "A convergent process of price adjustment and global Newton methods," *J. Math. Econ.*, vol. 3, pp. 107–120, 1976.
- [27] D. Stiliadis and A. Varma, "Rate-proportional servers: A design methodology for fair queueing algorithms," *IEEE/ACM Trans. Networking*, vol. 6, pp. 164–174, Apr. 1998.
- [28] H. R. Varian, *Microeconomic Analysis*, 3rd ed. New York, NY: Norton, 1992.
- [29] W. I. Zangwill and C. B. Garcia, "Equilibrium programming: The path following approach and dynamics," *Math. Prog.*, vol. 21, pp. 262–289, 1981.
- [30] H. Zhang, "Service disciplines for guaranteed performance service in packet-switching networks," *Proc. IEEE*, vol. 83, Oct. 1995.
- [31] L. Zhang, S. E. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A new resource reSerVation protocol," *IEEE Network*, pp. 8–18, Sept. 1993.



Steven H. Low (M'92–SM '99) received the B.S. degree from Cornell University, Ithaca, NY, and the Ph.D. degree from the University of California, Berkeley, in 1992, both in electrical engineering.

He was a consultant to NEC in the U.S. in 1991 and was with AT&T Bell Laboratories, Murray Hill, NJ, from 1992 to 1996. He was with the University of Melbourne, Australia, since 1996. He is currently with the California Institute of Technology, Pasadena, CA, as an Associate Professor. He has held visiting academic positions in the U.S. and

Hong Kong, and has consulted with companies in Australia and the U.S. His research interests are in the control and optimization of communications networks and protocols, and network security and privacy.

Dr. Low was a co-recipient of the IEEE William R. Bennett Prize Paper Award in 1997 and the 1996 R&D 100 Award. He is on the editorial board of IEEE/ACM TRANSACTIONS ON NETWORKING and has been a guest editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.