A Pricing Mechanism for Intertemporal Bandwidth Sharing with Random Utilities and Resources

Alberto Pompermaier
Department of Mathematics
London School of Economics
Houghton St.
London WC2A 2AE
England
email: a.pompermaier@lse.ac.uk

July 9, 2002

CDAM Research Report LSE-CDAM-2002-06

Abstract. This paper presents a pricing mechanism for the allocation of bandwidth within telecommunications networks. At the beginning of each period bandwidth is allocated among different users over a fixed finite number of time intervals. Since users are uncertain about their future bandwidth needs, their utility functions are random. An efficient allocation is achieved by applying standard economic principles (tatonnement process). At the beginning of each period the allocation changes, as uncertainty is resolved and new bandwidth becomes available. Different ways to deal with demand shocks (new users joining the network) and supply shocks (faults in the network) are proposed. Finally, a trade-off between economic and engineering efficiency is highlighted and one simple way to reduce the time needed to obtain a given level of economic efficiency is suggested.

1 Introduction

Over the past ten-fifteen years there has been an increasing interest in the development of pricing mechanisms for the allocation of telecommunications network resources. The main aim of these mechanisms is to allocate resources in such a way that those users who derive greater utility from using the network are not denied access by those who place a lower value on it. In other words, price mechanisms are designed to guarantee economic efficiency. At the same time, though, they should not impose an excessive overhead on the network, in the sense that the time and information needed by the system to compute the allocation prices should be kept to a minimum (trade off between engineering and economic efficiency).

An approach that achieves economic efficiency is the smart-market approach of MacKie-Mason and Varian [10], whereby the price to send a packet varies minute-by-minute to reflect the current degree of network congestion. Each packet has a bid in its header to indicate how much its sender is willing to pay to send it. However, users do not pay their bid-price, rather the lower market-clearing price. Packet marking is used also in Gibbens and Kelly [3]. They show that appropriate marking of packets at overloaded resources and charging the users for each mark leads to an efficient use of the network. In Kelly [7] users are charged a tariff based on declared and measured characteristics of traffic. The design of the tariff ensures that users make truthful declarations (no incentive to 'cheat'). Another pricing scheme is presented in Jiang and Jordan [5]. Their model incorporates statistical multiplexing and is based on effective bandwidth pricing. Lazar and Semret [8] introduce a new kind of auction, the progressive second price auction, in which users submit price and quantity bids to a central auctioneer. Fulp [1] develops an approach in which bandwidth is bought and sold in two types of markets: the reservation market and the spot market. In the reservation market, bandwidth is traded in amounts for a duration of time. In the spot market, users can immediately use whatever amount of bandwidth they find affordable, with no reservation overhead. A review of other approaches can be found in Jiang and Jordan [4].

In this paper, we propose a pricing mechanism for the allocation of a generic arbitrarily divisible resource among multiple users. The model is dynamic in the sense that the resource, which we shall refer to as bandwidth, is allocated over multiple periods. Our approach is therefore multi-market, as the one adopted by Fulp [1]. However, while in Fulp [1] in each period users can only bid for the next time interval, we allow users to bid for an arbitrarily large finite number of periods. The advantage of a multi-market approach is that it guarantees future bandwidth availability. This is particularly important for

inelastic applications as changing bandwidth amounts may result in sudden reductions of Quality of Service levels.
The allocation is driven by the users' requirements, which are formally derived by assigning to each of them a utility function. The system administrator, who is in charge of allocating the resource, has limited information about the utility functions and in any case less than the users themselves. This informational asymmetry gives rise to the need for a pricing mechanism, which would be otherwise unnecessary[1].
As it is natural in an intertemporal framework, we allow for uncertainty. That is, we consider the case of users' needs unexpectedly changing over time. The consequence is that the intertemporal allocation of bandwidth becomes dynamic. Demand shocks (new users joining the network) and supply shocks (faults in the network) are also accounted for in our model.
The remainder of the paper is structured as follows. In Section 2 we outline the basic model; in Section 3 we consider demand and supply shocks; in Section 4 we discuss the trade-off between engineering and economic efficiency while concluding remarks can be found in Section 5.

2 The model

In this section we outline the basic building blocks of the model, the users, the system administrator, and the way in which they interact.

2.1 Bandwidth demand

Demand for bandwidth is determined by the aggregation of the individual demands of all users having access to the resource. We assume that each user derives utility from bandwidth. Specifically, user k utility function in period 0 is as follows

$$U_{k,0} := U_{k,0}(x_{k,0}, x_{k,1}, ..., x_{k,T}). \tag{2.1}$$

The subscript k indicates that the utility function refers to user k, with $k = 1, 2, ..., K$. $x_{k,s}$ is the amount of period s bandwidth allocated to user k, with $s = 0, 1, ...T$. $U_{k,0}$ is continuous, twice differentiable, increasing in each of its arguments, concave, and random.
The randomness of $U_{k,0}$ reflects the fact that, when entering the network in

[1]A deterministic multi-period model in which users' utility functions are known to the system administrator is developed in Fulp [2].

period 0, user k does not know exactly how many jobs she will have to carry out in each future period. Since the execution of jobs requires bandwidth, it follows that she does not know how useful it will be to hold bandwidth. Hence, $U_{k,0}$ is random. The user's best guess in period 0 is given by $\overline{U}_{k,0}$, which denotes k's anticipation of $U_{k,0}$ in period 0.
Each user is given an initial amount of income, $I_{k,0}$, that they can use to buy bandwidth. We assume that this income takes on the form of 'funny money', $m_{k,0}$[2]. Hence

$$I_{k,0} = m_{k,0} \tag{2.2}$$

So, the objective of user k in period 0 is to maximise the anticipation of (2.1) given the information available in period 0 subject to the following intertemporal budget constraint

$$\beta(m_{k,t} - p_{t,0}x_{k,t}) = m_{k,t+1}. \tag{2.3}$$

$m_{k,t}$ is the amount of money that user k owns at the beginning of period t, $p_{t,0}$ is the price of period t bandwidth established in period 0, and $\beta \equiv (1 + r)$, where $r \geq 0$ is the interest rate[3]. (2.3) states that in period $t + 1$ user k can spend an amount of money equal to the one saved the previous period, increased by the interest rate, r.
Through repeated substitution (2.3) can be rewritten as follows[4]

$$m_{k,0} = \sum_{s=0}^{T} \frac{p_{s,0}x_{s,0}}{\beta^s}. \tag{2.4}$$

The solution to the problem of maximising (2.1) subject to (2.4) yields a set of $T + 1$ demand functions, $x^D_{k,t,0}$, for each user. Specifically

$$x^D_{k,t,0} := x^D_{k,t,0}(P_0, m_{k,0}) \qquad t = 0, 1, ..., T \qquad k = 1, 2, ..., K, \tag{2.5}$$

[2]The type of network we have in mind is rather private than commercial. Hence, the use of 'funny money' instead of real one.

[3]Note that $p_{t,0}$ is a price agreed in period 0 but to be paid in period t.

[4](2.4) is derived by setting $m_{k,T+1} = 0$. That is, we assume that each user spends all their money over the period $(0, T)$. This assumption stems from the fact that money does not enter the utility function. Hence, there is no reason for them to keep any of it beyond period T.

where $P_0 = (p_{0,0}, p_{1,0}, ..., p_{T,0})$. Concavity of (2.1) ensures that there is a unique solution to user k's maximisation problem, so that, given P_0 and $m_{k,0}$, $x^D_{k,t,0}$ is uniquely determined. Moreover, (2.5) is homogenous to the degree 0 in prices and income. This can be easily seen from the budget constraint (2.4). Any proportional increment in both prices and money leaves the budget constraint unchanged. That is, it has no impact on k's maximisation problem. As a consequence k's demand (2.5) is unaffected.

2.2 Bandwidth supply

The total amount of bandwidth available in each period, S_t, is assumed to be random. We use $\overline{S}_{t,0}$ to denote the amount of bandwidth that it is anticipated to be available for period t, whereby the anticipation is taken in period 0. Bandwidth supply is completely price inelastic as its amount does not depend on the price at which it is allocated.

The task of allocating bandwidth among the users is devoted to a 'system administrator'. Its objective is to allocate bandwidth in such a way that a weighted sum of the users' utilities is maximised. In other words, in period 0, the administrator wants to maximise

$$\overline{W}_0 := \sum_{k=1}^{K} \sigma_{k,0} \overline{U}_{k,0}, \tag{2.6}$$

where $\sigma_{k,0}$ denotes the weight that the administrator gives to k's utility at the beginning of period 0 and is therefore a function of the preferences of the administrator[5]. If there were symmetric information between the administrator and each user, that is, if the administrator knew about the characteristics of $U_{k,0}$ as much as user k does, then a pricing mechanism would be unnecessary. In fact, the administrator would simply work out the optimal allocation by maximising (2.6) subject to the constraint given by bandwidth availability. Hence, in what follows, we shall assume that there is asymmetric information between the administrator and each user, in the sense that each user has an informational advantage over the administrator about her own utility function. In this case the implementation of a pricing mechanism becomes necessary if a function like (2.6) is to be maximised.

[5] In the case of a corporate intranet each user's utility can be reinterpreted as that user's contribution to corporate profits as a function of the amount of resources she consumes. (2.6) would then measure the contribution of the network services to corporate profits (see MacKie-Mason *et al.* [11]).

2.3 Intertemporal equilibrium in period 0

The equilibrium in period 0 is obtained by equating total demand and supply for each period's bandwidth.
Total demand for period t bandwidth, $D_{t,0}(P_0)$, is derived by aggregating (2.5) over all users. That is

$$D_{t,0}(P_0) := \sum_{k=1}^{K} x_{k,t,0}^{D} \qquad t = 0, 1, ..., T. \tag{2.7}$$

Supply is simply the anticipated bandwidth availability, $\overline{S}_{t,0}$. An equilibrium price vector $P_0^* = (p_{0,0}^*, p_{1,0}^*, ..., p_{T,0}^*)$ is one equating demand and supply in each period. That is, one that solves

$$Z_{t,0}(P_0) := D_{t,0}(P_0) - \overline{S}_{t,0} = 0 \qquad t = 0, 1, ..., T. \tag{2.8}$$

At least one such price vector exists and leads to an efficient allocation. In fact

Proposition 1 *Given (2.1) and an initial allocation of money*

$$M_0^* = (m_{1,0}^*, m_{2,0}^*, ..., m_{K,0}^*)$$

(a) there exists at least one price vector $P_0^* = (p_{0,0}^*, p_{1,0}^*, ..., p_{T,0}^*)$ satisfying (2.8);
(b) the corresponding allocation

$$X_0^* = (x_{1,0,0}^*, ..., x_{1,T,0}^*; x_{2,0,0}^*, ..., x_{2,T,0}^*; x_{K,0,0}^*, ..., x_{K,T,0}^*)$$

is Pareto efficient;
(c) for some choice of weights $\sigma_{k,0}$, X_0^* maximises $\overline{W}_0$.

Proof: see Varian [14], especially Chapters 17 and 21[6]. □

The allocation X_0^* is found by substituting P_0^* into (2.5). Pareto efficiency (point (b)) means that bandwidth is allocated in such a way that no reallocation can yield a higher utility for any of the users without lowering the utility of at least one of them. As for point (c), as intuition suggests, the weight

[6] A more in-depth treatment of competitive equilibria can be found in Mas-Colell *et al.* [12].

given to user k's utility, $\sigma_{k,0}$, is positively correlated to her initial amount of money, $m_{k,0}$ (see Varian [14]).
Since $U_{k,0}$ is to a certain extent k's private information, the administrator can not use (2.8) to derive the equilibrium price vector as it can not work out aggregate demand. An auction mechanism is therefore needed. Specifically, we assume that the administrator sets up $T+1$ Walrasian auctions, one for each period in which bandwidth must be allocated. A Walrasian auction is one corresponding to the following algorithm[7]

1. At the beginning of period 0, the administrator announces a price vector $P_0^i = (p_{0,0}^i, p_{1,0}^i, ..., p_{T,0}^i)$;

2. users communicate to the administrator their desired amount of bandwidth for each time slot t according to (2.5);

3. the administrator communicates to the users a new price vector P_0^{i+1}, whereby each element of P_0^{i+1} is determined according to the following rule

$$
\begin{aligned}
Z_{t,0}(P_0^i) > 0 \qquad &\Rightarrow \qquad p_{t,0}^{i+1} = p_{t,0}^i + \epsilon \\
Z_{t,0}(P_0^i) < 0 \qquad &\Rightarrow \qquad p_{t,0}^{i+1} = p_{t,0}^i - \epsilon \\
Z_{t,0}(P_0^i) = 0 \qquad &\Rightarrow \qquad p_{t,0}^{i+1} = p_{t,0}^i
\end{aligned} \tag{2.9}
$$

where ϵ is a small positive scalar and $t = 0, 1, ..., T$;

4. go to step 2.

It must be stressed that this procedure does not guarantee that an equilibrium price vector is found. In fact, restrictions must be put upon the characteristics of aggregate demand if one wants to be sure that (2.9) converges towards an equilibrium within a finite number of steps[8].
Finally, note that no transaction occurs out of equilibrium. That is, until P_0^* has been announced by the administrator, users are not allocated any bandwidth. One of the aims of the administrator is therefore to minimise the number of iterations needed to reach an equilibrium. We come back to this issue in section 4.

[7] A similar type of auction has been used in various other contributions (see, for example, Low [9], Murphy and Murphy [13], and Jiang [6]).
[8] For a discussion see Mas-Colell *et al.* [12] and Varian [14].

2.4 Intertemporal equilibrium in period $h > 0$

In the previous section, we derived the equilibrium allocation at the beginning of period 0. We now assume that at the beginning of each subsequent period $h = 1, 2, ...$ bandwidth is allocated for the following T periods. This means that at the beginning of period 1, period $T + 1$ bandwidth must be distributed. In fact, all bandwidth until period T has already been allocated at the beginning of period 0. So, at the beginning of each period $h > 0$, the administrator distributes among the users the anticipated period $T + h$ bandwidth, $\overline{S}_{T+h,h}$. The amount received by user k is denoted by $s^*_{k,T+h,h}$, with $\sum_{k=1}^{K} s^*_{k,T+h,h} = \overline{S}_{T+h,h}$.
Although only bandwidth for one period needs distributing, we allow users to revise their entire consumption plans. In particular, they are allowed to exchange some of the bandwidth they had previously purchased. The exchange occurs through a Walrasian auction at the beginning of each period. This implies that, as in period 0, at the beginning of period h, $T+1$ simultaneous auctions are run, one for each time interval. The first T auctions concern bandwidth that had already been allocated in previous periods, while the remaining auction regards period $T+h$ bandwidth, which had not been auctioned before.
The reason why users may want to change their consumption plan is related to uncertainty. At the beginning of period h, users' objective functions are as follows

$$U_{k,h} := U_{k,h}(\hat{X}_{k,h-1}, x_{k,h}, x_{k,h+1}, ..., x_{k,T+h}), \tag{2.10}$$

with $k = 1, 2, ..., K$ and where $\hat{X}_{k,h-1} := (\hat{x}_{k,0}, \hat{x}_{k,1}, ..., \hat{x}_{k,h-1})$ represents past bandwidth utilization by user k.
Like $U_{k,0}$, $U_{k,h}$ is random, so that user k maximises its anticipation, $\overline{U}_{k,h}$. $\overline{U}_{k,h}$ differs from $\overline{U}_{k,0}$ under two aspects: it is a function of bandwidth from periods beyond T and incorporates all information that has become available between period 0 and period h. Both these features will, in general, induce users to change allocation at the beginning of each period. In fact, even if $U_{k,h}$ were a function of bandwidth only for periods until T, as it is $U_{k,0}$, users would still be willing to change the allocation established in period 0. This is because in each period users gain additional information about their future bandwidth needs, so that their anticipated utility function changes and with it the optimal bandwidth allocation.
As for the budget constraint, this has the same form as (2.4). Specifically

$$I_{k,h} = \sum_{s=h}^{T+h} \frac{p_{s,h} x_{k,s,h}}{\beta^{s-h}}. \tag{2.11}$$

However, unlike in period 0, the beginning of period h income of user k, $I_{k,h}$, is not an amount of money but instead the value of her bandwidth holdings. In fact, in period 0 users have spent all their money to buy bandwidth so that all they are left with is bandwidth for each period and no money. So, if they require more bandwidth for some specific period t, they have to sell some of their bandwidth from other periods. User k's total bandwidth holding at the beginning of period h is equal to all the bandwidth she had purchased in previous periods plus some period $T+h$ bandwidth. $I_{k,h}$ is therefore defined as follows

$$I_{k,h} := \sum_{s=h}^{T+h-1} \frac{p_{s,h} x^*_{k,s,h-1}}{\beta^{s-h}} + \frac{p_{T+h,h} s^*_{k,T+h,h}}{\beta^T}. \tag{2.12}$$

The first term on the right-hand-side of (2.12) is the period h monetary value of k's total bandwidth holding in period $h-1$. In fact, $x^*_{k,s,h-1}$ denotes the equilibrium allocation of period s bandwidth determined in period $h-1$. That is, $x^*_{k,s,h-1}$ is k's bandwidth holding for period $s \geq h-1$ during period $h-1$.

The second term is the period h monetary value of $s^*_{k,T+h,h}$, the amount of period $T+h$ bandwidth assigned to k by the administrator.

Maximisation of (2.13) subject to (2.11) yields the following demand functions

$$x^D_{k,t,h} := x^D_{k,t,h}(P_h, I_{k,h}) \qquad t = h, h+1, ..., T+h \qquad k = 1, 2, ..., K, \tag{2.13}$$

where $P_h = (p_{h,h}, p_{h+1,h}, ..., p_{T+h,h})$. Hence, at the beginning of period h, $T+1$ new auctions are put in place to reallocate bandwidth across the users. The beginning of period h auctions will lead to a new equilibrium price vector $P^*_h = (p^*_{h,h}, p^*_{h+1,h}, ..., p^*_{T+h,h})$ and to a new allocation of bandwidth

$$X^*_h = (x^*_{1,h,h}, ..., x^*_{1,T+h,h}, x^*_{2,h,h}, ..., x^*_{2,T+h,h}, x^*_{K,h,h}, ..., x^*_{K,T+h,h}).$$

X^*_h has the same characteristics of X^*_0 (see Proposition 1).

3 Demand and supply shocks

In the next two sections we consider changes on the demand and supply sides. As for the latter, we discuss the case under which in some period h the anticipation of S_t changes ($\overline{S}_{t,h} \neq \overline{S}_{t,h-1}$, for some $t \geq h$). This means that either some additional bandwidth is anticipated to become available for period $t \geq h$ or, due for example to a fault in the system, the amount of bandwidth available for period $t \geq h$ is anticipated to fall.
As for the demand side, we have so far kept the total number of users, K, fixed. It is however thinkable that users join and leave the network as time goes by. We therefore look at the consequences of having some users leaving earlier than they had anticipated and some other users joining in some period $h > 0$.

3.1 A demand shock: change in users' number

A demand shock can be of two types: either a user leaves the network or a new one joins it.
The former case can be easily accomodated by the present framework. If some user decides in period h that she will stay connected only until period $j < T + h$, she will sell all her period $j+$ bandwidth holdings to buy period $j-$ bandwidth.
Slightly more complicated is the latter case. Suppose that in period v it becomes known that in period $h \geq v$ a new user N joins the network. She can not be given any of period j bandwidth, $j = h, h+1, ..., T+v$, since this has already been allocated to current users. Giving some money to N as she enters the network would not allow her to purchase bandwidth. This is because users' utility depends only on bandwidth and not on money. Hence they are not willing to exchange the former for the latter.[9] However, she can be given at the beginning of period $v+1$ a portion of period $T+v+1$ bandwidth, as this has not been allocated before (see the second term of (2.12)). The new user can then use her portion of $T+v+1$ bandwidth to acquire bandwidth for periods between h and $T+v$ from current users.
So, the problem of a new user joining the network reduces to deciding how much period $T+v+1$ bandwidth to allocate to her. Clearly, the larger her

[9] One way around this problem is to make users' utility contingent on money holdings. For example, we could have

$$U_{k,h} := U_{k,h}(x_{k,h}, x_{k,h+1}, ..., x_{k,T+h}; m_{k,h}, m_{k,h+1}, ..., m_{k,T+h}).$$

In this case, users would derive utility from holding money. Hence they would be willing to exchange their bandwidth for money.

portion and the higher its price (the price is determined through the auction), the more period j bandwidth, $j = h, h+1, ..., T+v$, N will be able to buy from current users at the beginning of period $v+1$.
A partial alternative to this approach would be to keep some spare bandwidth available for new users. In this case, in period 0 the administrator would not auction-off all of the anticipated bandwidth availability, $\overline{S}_{t,0}$, but only a fraction of it. This fraction could be made contingent on current and future expected numbers of users. Define $\overline{K}_{t,0}$ as the number of users that in period 0 are anticipated to be willing to use period t bandwidth, then the administrator may choose to allocate in period 0 only $\delta_{t,0}\overline{S}_{t,0}$ where

$$\delta_{t,0} := \begin{cases} K/\overline{K}_{t,0} & \text{if} \quad K < \overline{K}_{t,0} \\ 1 & \text{otherwise} \end{cases}$$

In this way, for $t = 1, 2, ..., T$, the administrator has available $(1 - \delta_{t,0})\overline{S}_{t,0}$ period t bandwidth to allocate to new users[10].
Obviously the two approaches can be used together, in the sense that the administrator can decide to keep some spare bandwidth for each period as well as giving to new users portions of future still unallocated bandwidth.

3.2 A supply shock: change in available bandwidth

A supply shock consists in either a reduction or an increase in available bandwidth. We are interested in the case under which the period h anticipation of bandwidth availability for period t, $\overline{S}_{t,h}$, changes with h. Let us denote by $X^*_{t,h}$ the sum of all users' period t bandwidth holdings as determined in period $h \leq t$. That is

$$X^*_{t,h} := \sum_{k=1}^{K} x^*_{k,t,h} \qquad t = h, h+1, ..., T+h.$$

A period t bandwidth loss/gain as determined in period h, $\overline{L}_{t,h}$, is then defined as the difference between $X^*_{t,h}$ and $\overline{S}_{t,h}$. So

$$\overline{L}_{t,h} := X^*_{t,h} - \overline{S}_{t,h} \qquad t = h, h+1, ..., T+h.$$

[10]Note that, given the definition of $\delta_{t,0}$, if $\overline{K}_{t,0} \leq K$, there will be no spare bandwidth for period t. It follows that, if the anticipation about the number of users who want to use bandwidth in period t is incorrect, in the sense that there are actually more users than it was anticipated, new users will not be able to access the network during period t.

A positive $\overline{L}_{t,h}$ implies an anticipated bandwidth loss while a negative one implies an anticipated bandwidth gain. In fact, when $\overline{L}_{t,h}$ is positive more bandwidth for time t has been allocated across the users than it is anticipated to be available. By contrast, when $\overline{L}_{t,h}$ is negative it means that some spare bandwidth is thought to become available.

The latter case poses no difficulties. If some additional period t bandwidth becomes available, the administrator simply auctions it off among the users. Less straightforward is the case in which a reduction in available bandwidth is anticipated ($\overline{L}_{t,h} > 0$). In this case, users hold more period t bandwidth than it is anticipated to be physically available. One consequence is that they must be subtracted bandwidth that they have already been awarded. So, the administrator has the problem of how to allocate the loss among users.

There is no optimal way to carry out this task. That is, since the administrator has imperfect information about users' utility functions, there is no way it can optimally allocate period t bandwidth loss among the users. However, this is not a major problem since, whichever way it decides to distribute the loss, the users will be trading[11] with each other to come to an optimal allocation of the remaining available bandwidth. This reallocation of bandwidth may well involve periods different from the one in which the bandwidth reduction has occurred.

The problem of allocating the loss is therefore a problem of fairness. In what follows we propose to subtract bandwidth holdings from the users in such a way that relative incomes remain unchanged. Specifically, k's income in period h, $I^*_{k,h}$, is defined as follows

$$I^*_{k,h} := \sum_{s=h}^{T+h} \frac{p^*_{s,h} x^*_{k,s,h}}{\beta^{s-h}}.$$

That is, in period h, k's income is equal to the current value of her current and future bandwidth holdings given the set of equilibrium prices established in period h. To keep relative incomes across current users unchanged after the shock, user k should be subtracted an amount of bandwidth, whose value is equal to

$$\Delta I^*_{k,h} = \frac{I^*_{k,h}}{\overline{F}^*_h} \Delta F^*_h \qquad k = 1, 2, ..., K, \tag{3.1}$$

where ΔF^*_h is the current value of the anticipated bandwidth loss, that is,

[11] By trading we mean that users will use the next run of auctions to change the allocation.

$$\Delta F_h^* := \frac{p_{t,h}^* \overline{L}_{t,h}}{\beta^{t-h}}$$

and $\overline{F}_h^*$ is the total current value of the allocated bandwidth, i.e.

$$\overline{F}_h^* := I_{1,h}^* + I_{2,h}^* + ... + I_{K,h}^*. \tag{3.2}$$

(3.1) says that the value of the total bandwidth to be subtracted from k must be proportional to her income share in period h.

One way to abide by this rule is to subtract from user k an amount of period t bandwidth equal to $\Delta S_{k,t,h}$, where $\Delta S_{k,t,h}$ is defined as follows

$$\Delta S_{k,t,h} := \frac{I_{k,h}^*}{\overline{F}_h^*} \overline{L}_{t,h}. \tag{3.3}$$

The major drawback of (3.3) is that some users may have not enough available period t bandwidth to contribute to the loss in proportion of their relative income. That is, it might well be that for some k we have $x_{k,t,h}^* < \Delta S_{k,t,h}$. For example, suppose that user C has not bought any period t bandwidth. Then she is unaffected by the anticipated loss of period t bandwidth and (3.3) can not be applied. It follows that other users have to give up an amount of bandwidth larger than the one implied by (3.3) for a value larger than the one implied by (3.1). In this case, to re-establish fairness, bandwidth from other periods for a value corresponding to $\Delta I_{C,h}^*$ should be taken away from user C and redistributed to those users whose loss of period t bandwidth exceeds the one implied by (3.3).

It should be stressed that (3.1) and (3.3) are fair rules not efficient ones. After these rules have been applied, it is up to the users to engage in trading so that an efficient allocation of bandwidth is achieved. This means that, even if the loss is distributed randomly among the users, as long as in the wake of the shock an auction is run, a Pareto optimal equilibrium is reached.

4 Engineering vs. economic efficiency

We saw in section 3 that an auction run according to (2.9), if it converges, leads to a Pareto efficient equilibrium. The equilibrium allocation X_0^* is therefore economic efficient, as it maximises a weighted sum of the individual utility functions (see Proposition 1).

However, a Walrasian auction requires a certain number of iterations before

it reaches the equilibrium price vector P_0^*. We shall denote this number by the letter q and interpret it as a measure of engineering efficiency. The larger is q the less efficient is the auction.
In designing the auction the administrator aims therefore at minimising q while preserving economic efficiency.
Intuition suggests that the size of q may depend on at least two factors. The size of the price adjustment, ϵ, and the level of economic efficiency that the administrator wants to achieve. We examine these two factors in turns.

q and economic efficiency

We assess economic efficiency by looking at the difference between supply and demand. Specifically, the more precisely demand and supply are matched by a given price vector, the more economically efficient is the resulting allocation. A trade-off between engineering and economic efficiency can therefore be established. In fact, the number of iterations needed to reach an equilibrium decreases with the accuracy of the supply-demand match. In other words, if the administrator is willing to accept allocations of bandwidth such that demand is slightly lower than supply, the number of eligible equilibrium price vectors will increase. Hence, it can expect to reach an equilibrium in a smaller number of iterations. This discussion leads to the following definition of $P_h^*(\gamma_h)$

Definition 1 *A price vector is denoted by $P_h^*(\gamma_h)$ if it satisfies the following conditions*

$$\overline{S}_{t,h} - D_{t,h}(P_h^*(\gamma_h)) \geq 0 \qquad t = h, h+1, ..., T+h$$

$$\max\{\overline{S}_{t,h} - D_{t,h}(P_h^*(\gamma_h))|\ t = h \dots T+h\} = \gamma_h$$

where $\gamma_h \geq 0$.

γ_h measures the maximum amount of unallocated bandwidth in any period t associated with the price vector $P_h^*(\gamma_h)$. The smaller is γ_h, the more efficient is the allocation. However, the smaller is γ_h, the larger is the number of iterations needed to find $P_h^*(\gamma_h)$.
So, if the administrator is willing to accept some degree of discrepancy between supply and demand to proceed to allocate bandwidth, it can expect to need a smaller number of iterations to reach an equilibrium.
We now analyse a second factor that affects the number of iterations needed to reach an equilibrium: the size of the price adjustment, ϵ.

q and the size of the price adjustment

Rule (2.9) implies a stepwise adjustment of bandwidth price. Let us denote the size of the i-th step in the auction for period t bandwidth run at the beginning of period h by $\epsilon^i_{t,h}$. So

$$\epsilon^i_{t,h} := |p^i_{t,h} - p^{i-1}_{t,h}|,$$

where $p^i_{t,h}$ is the i-th announced price for period t bandwidth during the auction run at the beginning of period h. The choice of the size of the price adjustment has an impact on the speed at which the equilibrium is reached, that is, on q. Specifically, if the first announced price $p^1_{t,h}$ is very large/small compared to the equilibrium price $p^*_{t,h}(\gamma_h)$, a small adjustment, i.e. a small $\epsilon^i_{t,h}$, implies a large number of iterations, that is a large q, before $p^*_{t,h}(\gamma_h)$ is reached. However, a large $\epsilon^i_{t,h}$, though it increases the speed at which $p^i_{t,h}$ approaches $p^*_{t,h}(\gamma_h)$, may cause the announced price to fluctuate *ad infinitum* around the equilibrium. The administrator faces therefore a trade-off. The larger (smaller) is $\epsilon^i_{t,h}$, the lower (higher) is q but also the lower (higher) the probability that the equilibrium price will ever be reached.

We deal with this trade-off by using a variable adjustment process. Define the excess demand function, $Z^i_{t,h}$, as follows

$$Z^i_{t,h} := D_{t,h}(P^i_h) - \overline{S}_{t,h}$$

then we set $\epsilon^i_{t,h}$ equal to $\hat{\epsilon}^i_{t,h}$, which is given by

$$\hat{\epsilon}^i_{t,h} := f_{t,h}\left(Z^{i-1}_{t,h}, Z^{i-2}_{t,h}\right)\hat{\epsilon}^{i-1}_{t,h}, \tag{4.1}$$

with the initial value of $\hat{\epsilon}^i_{t,h}$, $\hat{\epsilon}^2_{t,h}$, to be chosen by the administrator.
$f_{t,h}$ is a function with the following characteristics

$$\begin{aligned} \text{sign}\{Z^{i-1}_{t,h}\} = \text{sign}\{Z^{i-2}_{t,h}\} \qquad &\Rightarrow \qquad f_{t,h} > 1 \\ \text{sign}\{Z^{i-1}_{t,h}\} \neq \text{sign}\{Z^{i-2}_{t,h}\} \qquad &\Rightarrow \qquad f_{t,h} < 1 \end{aligned} \tag{4.2}$$

From (4.1) and (4.2) follows that $\hat{\epsilon}^i_{t,h}$ increases as long as $p^i_{t,h}$ is smaller/larger than $p^*_{t,h}(\gamma_h)$ and decreases when $p^i_{t,h}$ starts fluctuating around $p^*_{t,h}(\gamma_h)$. So, $p^i_{t,h}$ will approach $p^*_{t,h}(\gamma_h)$ in less iterations than under a fix size adjustment and will not fluctuate *ad infinitum* around it since a decreasing $\hat{\epsilon}^i_{t,h}$ ensures that the oscillations get smaller at each iteration so that $p^i_{t,h}$ converges to $p^*_{t,h}(\gamma_h)$.

5 Conclusion

In this paper we described a pricing mechanism for efficient bandwidth sharing. Users purchase bandwidth for immediate as well as future consumption and revise their consumption plans at the beginning of each period. The main reason why users wish to change allocation in each period is the presence of uncertainty. As time goes by, new information is revealed. Users respond to it by updating their anticipated bandwidth needs. This, in turn, leads to new demand functions and, eventually, to a new allocation of bandwidth. We also provided an auction design aimed at reducing the time needed to reach an equilibrium while maintaining an acceptable level of economic efficiency. Finally, we showed that our framework is well suited to deal with changing numbers of users as well as with faults in the network.

References

[1] Fulp, E. W. and D. S. Reeves (2000), *QoS Rewards and Risks: A Multi-Market Approach to Resource Allocation*, Proceedings of the IFIP TC6 Networking 2000 Conference, Paris, France

[2] Fulp, E. W. and D. S. Reeves (2001), *Optimal Provisioning and Pricing of Differentiated Services Using QoS Class Promotion*, Proceedings of the Informatik 2001, Vienna, Austria

[3] Gibbens, R.J. and F. P. Kelly (2000), "Resource Pricing and the Evolution of Congestion Control", *Automatica*, vol. 35, pp. 1969-1985

[4] Jiang, H. and S. Jordan (1995), "Connection Establishment in High Speed Networks", *IEEE Journal on Selected Areas in Communications*, 13(7), pp. 1150-1161

[5] Jiang, H. and S. Jordan (1995), "The Role of Price in the Connection Establishment process", *European Trans. Telecommunications*, 6(4), pp. 421-429

[6] Jiang, H., I. Sidhu and S. Jordan (1996), "A Pricing Model for Networks with Priorities", *Allerton Conference on Communication, Control and Computing*, pp. 269-275

[7] Kelly, F. P. (1997), "Charging and Accounting for Bursty Connections", in L. W. McKnight and J. P. Bailey (Eds.), *Internet Economics*, MIT Press

[8] Lazar, A.A. and N. Semret (2001), "The Design, Analysis and Simulation of the Progressive Second Price Auction Mechanism for Network Bandwidth Sharing", *Telecommunications Systems*, forthcoming

[9] Low, S. H. and P. P. Varaiya (1993), "A New Approach to Service Provisioning in ATM Networks", *IEEE/ACM Transactions on Networking*, 1(5), pp. 547-553

[10] MacKie-Mason, J. K. and H. R. Varian (1994), "Pricing the Internet", in B. Kahin and J. Keller (Eds.), *Public Access to the Internet*, Prenticce Hall

[11] MacKie-Mason, J. K., P. Thomas and D. Teneketzis (1999), "A Market-Based Approach to Optimal Resource Allocation in Integrated-Services Connection-Oriented Networks", http://www-personal.umich.edu/ jmm/papers.html

[12] Mas-Colell, A., M. D. Whinston, and J. R. Green (1995), *Microeconomic theory*, OUP, New York

[13] Murphy J. and L. Murphy (1994), "Bandwidth Allocation By Pricing In ATM Networks", *IFIP Transactions*, C-24, pp. 333-351

[14] Varian, H. R. (1992), *Microeconomic Analysis*, 3rd Edition, Norton, New York