Market Pricing of Differentiated Internet Services

CU/CTR/TR 503-98-37

N. Semret, R. R.-F. Liao, A. T. Campbell and A. A. Lazar
Dept. of Electrical Engineering, Columbia University
New York, NY, 10027-6699, USA.
{nemo, liao, campbell, aurel}@comet.columbia.edu

December 4, 1998

Abstract

This paper presents a decentralized auction-based approach to pricing of edge-allocated bandwidth in the differentiated services model of the Internet. The players in this architecture are users, one raw-capacity seller per network and one broker per service per network. With the Progressive Second Price auction mechanism as the basic building block, we conduct a game theoretic analysis, deriving optimal strategies for buyers and brokers, and the existence of network-wide equilibria. We investigate the system dynamics by simulating a scenario with two inter-connected networks, and three types of services built on the proposed standard expedited forwarding (EF) and assured forwarding (AF) per-hop-behaviors.

1 Introduction

The recent development of the differentiated service (diff-serv) Internet model is aimed at supporting service differentiation for aggregated traffic in a scalable manner [5, 3]. The tenet of diff-serv is to relax the traditional hard-QOS model (e.g. end-to-end per-flow guarantee of Int-serv, and ATM) in two dimensions: slower time-scale network mechanisms and coarser-grained traffic provisioning.

While the role of prices as essential resource allocation control signals has been established from the outset of diff-serv [2, 17], the precise development of pricing mechanisms is still at its very early stages. In the Simple

Integrated Media Access [13], the service charge for a user is proportional to the nominal bit rate subscribed by the user, and the price differentiation between different service classes is fixed. Similarly, in the User-Share Differentiation proposal [22], the pricing is based on the user share that is allocated over a long time scale. These schemes fall within the category of capacity-based pricing. Just as diff-serv aims to provide a range of "better than best-effort" services without the complexity and per-flow state of hard-QOS, capacity-based pricing schemes can be thought of as "better than flat-rates" (more rational and sustainable from the economic point of view), without the per-flow measurement and accounting required by usage-based pricing. Flat-rate pricing is the extreme of capacity pricing where the capacity equals the access line speed, while usage pricing can be thought of as the extreme where capacities are continuously adapted to fit the actual transmission rate of each flow at each moment in time. A pricing scheme which explicitly covers the range between these two, as well as the service-type dimension, is that of [12].

The space of network resource pricing schemes has many dimensions.[1] One is "where" the capacity abstraction takes place: at each hop inside the network or at the edges [19]. Another is how much *a priori* information on demand is required. At one extreme, the seller assumes perfect a priori knowledge of demand and does an offline calculation of optimal prices (e.g. time-of-day pricing based on historical traffic patterns). A less extreme approach is to assume only the functional form of demand, and adjust prices by on-line "tatonnement" algorithms [8].

Auctioning is the pricing approach with minimal information requirement. The more difficult it is for the seller to obtain demand information (or valuations), the stronger the case is for using auctions. In the Internet, because of the diverse and rapidly evolving nature of the applications, services, and population, the case is compelling. With suitably designed rules, auctions can achieve efficient (value maximizing) allocations with minimal a priori information.

In this paper, we investigate the feasibility of auctioning capacity in the diff-serv model. We show through game theoretic analysis and simulation that the Progressive Second Price (PSP) auction introduced in [14] can provide stable pricing in the diff-serv bandwidth market. The PSP mechanism achieves economic objectives (incentive compatibility, and efficiency), while being realistic in the engineering sense (small signalling load and compu-

[1] This is discussed more fully in [14].

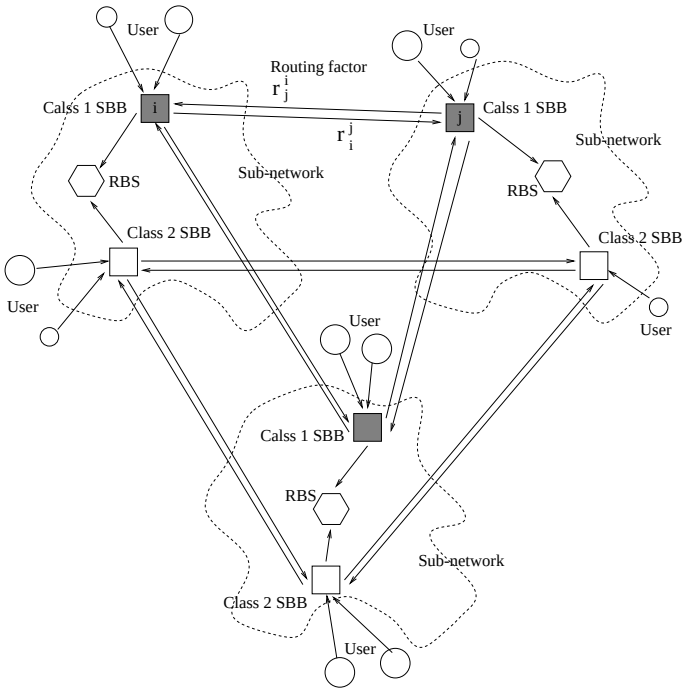

Figure 1: 2-tier Auction Pricing Framework for Diff-serv Internet

tationally simple allocation rule). As such, it provides a useful baseline for understanding the conditions for the economic feasibility of wide-area differentiated services.

The structure of this paper is as follows. In Section 2 we describe the framework of our pricing scheme. In Section 3 we present a game theoretical analysis of the scheme. Following this, in Section 4, we use simulations to study the capacity market dynamics.

2 Market-based Pricing Framework

Our network model assumes that each network can be abstracted into a single bottleneck capacity (e.g. as a "Norton-equivalent" [10]). The capacity may be represented by an absolute amount of bandwidth, or some relative metrics like user share in the User-Share Differentiation proposal [22] or resource token in Location Independent Resource Accounting proposed by Stoica and Zhang [20]. Large networks can be modeled by subdivision into

a set of interconnected networks, each of which can be abstracted into a bottleneck capacity. The degree of subdivision that is necessary depends on traffic, topology and size constraints as well as the desired level of accuracy. Within each network, the routing of aggregated traffic to each peer[2] is stable over the resource allocation time scale (e.g. in the order of hours) .

Figure 1 presents the architecture of our proposed auction pricing framework for a set of interconnected networks as described above.

A two-tier whole-seller/retailer market model is used to accommodate a network of goods (i.e. bandwidth) with multiple differentiated service classes. We define three kinds of players: users, service bandwidth brokers (SBBs) and raw bandwidth sellers (RBSs), to play the roles of end-users, retailers and whole-sellers respectively. Each network has a single RBS and a separate SBB for each class of service being offered. The RBS can be thought of as the bearer, and the SBBs as service providers. The RBS and SBBs on the same network may or may not be owned by the same entity. The idea is that the competition among SBBs results in a dynamic and efficient partition of the physical network resources among the services being offered, based on the demands from users. The users in our model are large/aggregated subscribers (web site, intra/extranet, virtual private networks) to a particular service offered by a particular provider, and buying capacity from an SBB.

3 Game Theoretic Modeling and Analysis

3.1 Message Process and Notation

Let the set of all players, including buyers, sellers and brokers (brokers are both buyers and sellers), be denoted by $\mathcal{I} = \{1, \ldots, I\}$. Following the notation in [14], a player's identity $i \in \mathcal{I}$ as a subscript indicates that the player is a buyer, and as a superscript indicates the seller.

Suppose player i is buying from player j. Then he places a **bid** $s_i^j = (q_i^j, p_i^j)$, meaning he would like to buy from j a quantity q_i^j and is willing to pay a *unit* price p_i^j. Without loss of generality, we assume that all players bid in all auctions, with the understanding that if a player i does not need to buy from j, we simply set $s_i^j = (0, 0)$.

[2] In this paper, we use the term "peer" in the most general sense, i.e. any network which inter-connects with a given network, and not just those that choose to exchange all traffic free of charge.

A seller j places an **ask** $s_j^j = (q_j^j, p_j^j)$, meaning he is offering a quantity q_j^j, with a reserve unit price of p_j^j. In other words, when the subscript and superscript are the same, the bid is understood as an ask.

When sub/superscripts are omitted, the notation refers to the vector obtained by letting it range over all values. For example, if q_i^j denotes the quantity player i wants to buy from player j, then q_i is the $1 \times I$ vector $(q_i^1, \ldots, q_i^I)$, and q is the $I \times I$ matrix. A subscript with a minus sign indicates a vector with that component deleted $s_{-i} \equiv (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_I)$, and $(x_i; s_{-i})$ denotes the profile obtained by replacing s_i with x_i.

Based on the profile of bids $s^j = (s_1^j, \ldots, s_I^j)$, seller j computes an allocation $(a^j, c^j) = A^j(s^j)$, where a_i^j is the quantity given to player i and c_i^j is the total cost charged to the player i. A^j is the **allocation rule** of seller j. It is feasible if $a_i^j \leq q_i^j$, and $c_i^j \leq p_i^j q_i^j$. One possible allocation rule is the progressive second price (PSP) auction – see Section 3.4.

3.2 Sellers

Suppose player $k \in \mathcal{I}$ is an RBS. Then its strategy consists of always *ask*ing $s_k^k = (q_k^k, p_k^k)$, with q_k^k equal to the physical bottleneck capacity of its network, and p_k^k equal to the unit cost of operation. Since it is a passive seller of physical bandwidth, k does not buy from anyone, i.e $s_k^j = 0, \forall j \neq k$.

Suppose $j \in \mathcal{I}$ is an SBB. It offers a capacity q_j^j for sale to its users. In order to honor its contracts, the quantity offered must be constrained by the capacities that j can actually obtain. First, it must get enough bandwidth from k, the RBS in its own network, to carry the total capacity it allocates to its customers, i.e.

$$\sum_i a_i^j \leq a_j^k. \tag{1}$$

Second, since it is selling wide-area service, j must get enough capacity from the SBBs offering the same service in each peer network. Let l denote one such peer SBB, and r_j^l be the "fraction of traffic" generated by j's customers that is routed to the network where player l is the peer SBB (see Remark below for interpretations of r_j). Then, j must satisfy

$$r_j^l \sum_{i \neq l} a_i^j \leq a_j^l, \tag{2}$$

for all peers l.[3] For notational convenience, fix $r_j^k = 1$, when k is j's RBS.

[3] We assume that service providers block "loop-back" traffic, i.e. traffic going from l

Since $a_k^j = 0$, (2) includes (1) as the special case $l = k$. If l is neither a peer of j, nor its RBS, then we set $r_j^l = 0$.

Define, for any allocation a,

$$e_j^l(a) \stackrel{\triangle}{=} \frac{a_j^l}{r_j^l} + a_l^j.$$

Proposition 1 *(Broker's sell-side constraints) Let $j \in \mathcal{I}$ be a SBB, and fix its buy-side allocation (a_j, c_j). Then, on the sell-side, the quantity offered must satisfy*

$$q_j^j \leq \min_{l \neq j} e_j^l(a)$$

For a broker who does not sell at a loss, the reserve price must satisfy

$$p_j^j \geq \frac{1}{q_j^j} \sum_l c_j^l.$$

Proof: Suppose $\exists l \neq j$ such that $q_j^j > e_j^l$. Then when all the offered quantity is bought, we have $\sum_i a_i^j = q_j^j > e_j^l = \frac{a_j^l}{r_j^l} + a_l^j \Leftrightarrow \sum_{i \neq l} a_i^j > \frac{a_j^l}{r_j^l}$, and condition (2) is violated. This proves the first assertion.

Since $\sum_l c_j^l$ is the total cost of the capacity that j is buying, the second assertion follows immediately from the our assumption that the broker will not sell at a loss. □

We will call $\min_{l \neq j} e_j^l(a)$ the *expected bottleneck* capacity for the service offered by j.

Remark: r_j is a vector of route-provisioning coefficients and can accommodate different types of brokers.

- Expected capacity service agreement: one type of SBB may offer service level agreements (SLAs) for expected capacity, i.e. on average, users will get the capacity they pay for, even when the traffic enters peer networks. This could include for example services built on the diff-serv assured forward (AF) per-hop behaviors [9]. In this case, r_j^l is the expected[4] fraction of the total traffic entering j that is routed

through j and back to l. If that is not the case, then the summation would be over all i.

[4]Thus r represents aggregate flow patterns. r is measured over a time-scale slow enough to make quasi-static estimates which average out instantaneous micro-flows.

to l. r_j^j is the fraction of traffic that terminates with one of j's own customers, and $\sum_{l \neq k} r_j^l = 1$, where k is the RBS of j.[5]

- Worst-case capacity service agreement: another type of SBB may offer service agreements for worst-case bandwidth, i.e. each user always gets the amount of bandwidth they pay for, even if all of the traffic is routed to the same peer. This could include for example services built on the diff-serv expedited forwarding (EF) per-hop behavior [11]. In this case $r_j^l = 1$ for all peers l.

- For an SBB which offers SLAs valid only within its own network, $r_j^j = 1$ and $r_j^l = 0, \forall l \neq j$.

Depending on the scheduling and buffer management algorithms used to provide the PHBs, some amount of over-provisioning may be required [11]. Degrees of over-provisioning must also be used to differentiate among AF classes for example [9]. These engineering needs can be represented in this model by simply factoring the over-provisioning into r (see Section 4 for a specific scenario).

3.3 Buyers

A general buyer $i \in \mathcal{I}$, seeks to maximize its utility

$$u_i = \theta_i(a) - \sum_j c_i^j, \tag{3}$$

possibly subject to a budget constraint

$$\sum_j c_i^j \leq b_i.$$

Remark: θ_i is the buyer's *valuation* function, which is private information. In particular, the seller only sees the buyers' bids, and not the valuations that lead buyers to make those bids.

If the buyer is a user, then (3) has the simpler form $u_i = \theta_i(a_i^j) - c_i^j$, where j is that user's SBB. The valuation is a function of the player's own allocation only, and expresses the amount the user is willing to pay for

[5] Note that for expected capacity, a user i whose traffic is entirely within the allocated profile a_i^j when it enters its broker j's network could temporarily be out of profile in the peer network l, if j miscalculated r_j^l, or if there is a sudden surge of traffic from many of j's customers to l.

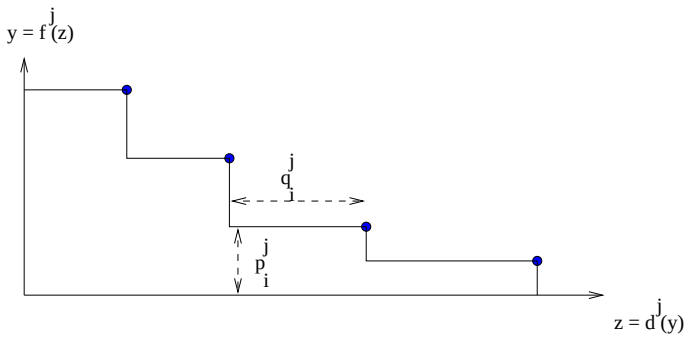

Figure 2: Demand curve for a broker j.

each possible quantity of resource. It can be based on economic and/or information theoretic considerations (see the appendix in [14]).

If the buyer is a broker, the natural utility is the potential profit $u_j = \theta_j(a) - \sum_k c_j^k$, where θ_j, the broker's buy-side valuation, is the potential revenue from the sale (on the sell-side) of the capacity obtained on the buy-side. The potential revenue is derived from the demand on the sell-side: let $\forall y \geq 0$,

$$d^j(y) \stackrel{\triangle}{=} \sum_{p_k^j \geq y} q_k^j,$$

the demand at unit price y. Its "inverse" function is defined by,

$$f^j(z) \stackrel{\triangle}{=} \sup \left\{ y \geq 0 : d^j(y) \geq z \right\}.$$

See Figure 2. Note that we chose f^j to be continuous from the left. For a given demand function $d^j(.)$, $\forall z \geq 0$, $f^j(z)$ represents the highest unit price at which j *could* sell the z-th unit of capacity (the actual prices charged to users depend on the specific allocation mechanism A^j used).

Proposition 2 *(Broker's buy-side valuation) Let $j \in \mathcal{I}$ be a broker with inverse demand $f^j(z)$, and fix its sell-side offering to the maximum quantity allowed by Proposition 1, $q_j^j = e_j^m$, where $m = \arg\min_{l \neq j} e_j^l$. Then, on the buy-side, its valuation is*

$$\theta_j(a) = \int_0^{e_j^m(a)} f^j(z)\, dz.$$

Proof: If e_j^m decreases by δ, then q_j^j must be reduced by δ. The value to j of the lost quantity is the revenue j would have gotten from it. By definition, this lost potential revenue is $f^j(e_j^m)\delta$. Thus, by abuse notation, writing θ_j as a function of e_j^l,

$$\theta_j(e_j^m) - \theta_j(e_j^m - \delta) = f^j(e_j^m)\delta$$

and the result follows. □

3.4 Analysis

The progressive second price auction (PSP) was introduced in [14]. It is defined by: $\forall i, j \in \mathcal{I}$,

$$a_i^j(s) \equiv a_i^j(s^j) = q_i^j \wedge \left[q_j^j - \sum_{p_k^j \geq p_i^j, k \neq i} q_k^j \right]^+, \tag{4}$$

$$c_i^j(s) \equiv c_i^j(s^j) = \sum_{k \neq i} p_k^j \left[a_k^j(0; s_{-i}^j) - a_k^j(s_i^j; s_{-i}^j) \right], \tag{5}$$

where $\wedge$ means taking the minimum. Note that each seller computes allocations from local information only (the bids for that resource). Define,

$$P_i^j(z) \stackrel{\triangle}{=} \inf \left\{ y \geq 0 : q_j^j - \sum_{p_k^j > y, k \neq i} q_k^j \geq z \right\}.$$

Note that we define P_i^j to be continuous from the left. Under PSP, P_i^j is the market price function from the point of view of user i. Indeed, it can be shown that,

$$c_i^j = \int_0^{a_i^j} P_i^j(z)\, dz. \tag{6}$$

Remark: Except at points of discontinuity, we have $P_i^j(z) = f^j(q_j^j - z)$. This mechanism generalizes Vickrey ("second-price") auctions [21] which are for non-divisible objects. PSP bears some similarity to Clarke-Groves mechanisms [4, 7]. The fundamental difference with the latter is that PSP is designed with a message (bid) space of two dimensions (price and quantity) in which each message is a single point, rather than an infinite dimensional space of valuation functions where each message is a revelation of the whole valuation curve (see [16, 14] for an explanation of the "revelation principle").

This reduction of the message space is crucial in the context of communication networks, where limiting the size and complexity of the exchanged messages (signalling) is very important.

Define *elastic demand* as follows: $\forall i$, θ_i is continuous, concave, and smooth (θ_i' is continuous); and for some (possibly infinite) maximum capacity $\bar{q}_i \leq \infty$, θ_i' is strictly decreasing on $[0, \bar{q}_i]$, and non-increasing on $[\bar{q}_i, \infty)$.

Under elastic demand analyzed as a complete information game, the PSP auction has the following properties which are proven in [14]:

- incentive compatible: truth-telling (setting the bid price equal to the marginal valuation) is a dominant strategy;
- stable: it has a "truthful" ϵ-Nash equilibrium [6], for any positive seller reserve price;
- efficient: at equilibrium, allocations maximize total user value (social welfare) to within $O(\sqrt{\epsilon})$; and
- enables a direct trade-off between engineering and economic efficiency (measured respectively by convergence time and total user value), by the parameter ϵ, which has a natural interpretation as a bid fee.

In the rest of this paper, we assume all the sellers in the network are using PSP as the allocation mechanism.

For users, the best strategy consists simply of bidding for the largest quantity such that the marginal valuation is higher than the market price, and setting the bid price equal to the marginal valuation (i.e. "truth-telling" is optimal).

Proposition 3 *(User's strategy) Let $i \in \mathcal{I}$ be a user such that θ_i that is differentiable and θ_i' continuous from the left. Let $l \in \mathcal{I}$ be that user's broker. For a fixed profile s^l_{-i}, an ϵ-best reply for player i is $t^l_i = (v^l_i, w^l_i)$, such that*

$$v^l_i = \sup\left\{ z \geq 0 : \theta_i'(z) > P^l_i(z) \text{ and } \int_0^z P^l_i(\eta)\, d\eta \leq b_i \right\} - \epsilon/\theta_i'(0),$$

and

$$w^l_i = \theta_i'(v^l_i).$$

That is, $\forall s^l_i$, $u_i(t^l_i; s^l_{-i}) \geq u_i(s^l_i; s^l_{-i}) - \epsilon$.

Proof: This is a special case of Proposition 4, with $\forall l, a_l^i = 0$, $r_i^i = 1$, and $\forall l \neq i$, $r_i^l = 0$. This was derived separately in [14]. □

Consider now a broker, participating in many auctions simultaneously. By the nature of its valuation (Proposition 2), capacity allocations are valuable to the broker only insofar as they increase its expected bottleneck capacity $\min_{l \neq i} e_i^l$. Thus, a broker must coordinate its buy-side bids (one submitted to each of its peers and its RBS) to maximize its overall utility.

Note that for Proposition 3, we do no require that θ_i be smooth. Concavity and non-increasingness suffice, along with the purely technical condition of continuity from the left. These are satisfied by the broker's valuation (Proposition 2). Thus, we can expect that the same principle (optimality of truth-telling) should hold.

Indeed, as we will now show, it turns out that the optimal strategy is very similar to that of a single user. But instead of searching directly for the optimal capacity, the broker finds the optimal expected bottleneck e, which is the largest one such that the marginal value is just greater than the market price. The role of the market price is played by the average of the market prices at the different auctions, weighted by the route provisioning factors. The actual bids are obtained by transforming the desired optimal expected bottleneck e back into the corresponding quantities v_i^l to bid at each buy-side market. As with a user, truth-telling is optimal for the broker, so at each buy-side market, the broker sets the bid price to the marginal value.

Proposition 4 *(Broker's buy-side strategy) Let $i \in \mathcal{I}$ be a broker, and fix all the other players' bids s_{-i}, as well as the broker's sell-side s_i^i (thus a^i is fixed). Let*

$$e = \sup \left\{ h \geq 0 : f^i(h) > \sum_{l \neq i} P_i^l \left((h - a_l^i) r_i^l \right) r_i^l \right\} - \epsilon / f^i(0), \tag{7}$$

and for each $l \neq i$,

$$v_i^l = (e - a_l^i) r_i^l,$$

and

$$w_i^l = \frac{1}{r_i^l} f^i(e).$$

Then a (coordinated) ϵ-best reply for the broker is $t_i = (v_i, w_i)$, i.e., $\forall s_i$, $u_i(t_i; s_{-i}) + \epsilon \geq u_i(s_i; s_{-i})$.

Proof: Since f^i is non-increasing and $\forall l$, P_i^l is non-decreasing, (7) implies $f^i(e) > \sum_{l\neq i} P_i^l(v_i^l) r_i^l$, and therefore $\forall l \neq i$,

$$\begin{array}{rrcl} & w_i^l & > & P_i^l(v_i^l) = f^l(q_l^l - v_i^l) \\ \Rightarrow & v_i^l & \leq & q_l^l - d^l(w_i^l), \\ \Rightarrow & a_i^l(t_i; s_{-i}) & = & v_i^l, \\ \Rightarrow & e_i^l \circ a(t_i; s_{-i}) & = & e. \end{array}$$

Therefore

$$\begin{aligned} u_i(t_i; s_{-i}) &= \int_0^e f^i(\eta)\, d\eta - \sum_{l\neq i} \int_0^{v_i^l} P_i^l(z)\, dz \\ &= \int_0^e f^i(\eta)\, d\eta - \sum_{l\neq i} \int_{a_l^i}^e r_i^l P_i^l \left((\eta - a_l^i) r_i^l\right)\, d\eta. \end{aligned}$$

Now suppose $\exists s_i = (q_i, p_i)$ such that $u_i(s_i; s_{-i}) > u_i(t_i; s_{-i}) + \epsilon$. Let $\xi = \min_{k\neq i} e_i^k \circ a(s)$, and $\forall l \neq i$, $\zeta_i^l = \left[\xi - a_l^i\right] r_i^l$ and $\sigma_i = (\zeta_i, p_i)$. From (9) in Lemma 1, $a_i^j(\sigma_i; s_{-i}) = \zeta_i^j$, therefore

$$u_i(\sigma_i; s_{-i}) = \int_0^\xi f^i(\eta)\, d\eta - \sum_{l\neq i} \int_{a_l^i}^\xi r_i^l P_i^l \left((\eta - a_l^i) r_i^l\right)\, d\eta.$$

By Lemma 1 (given in the Appendix), $u_i(\sigma_i; s_{-i}) \geq u_i(s_i; s_{-i})$. Therefore, $u_i(\sigma_i; s_{-i}) > u_i(t_i; s_{-i}) + \epsilon$, which by Proposition 2, is equivalent to

$$\int_e^\xi f^i(\eta)\, d\eta - \sum_{j\neq i} \int_e^\xi r_i^l P_i^j \left((\eta - a_l^i) r_i^l\right)\, d\eta > \epsilon. \tag{8}$$

Let $\overline{e} = e + \epsilon / f^i(0)$. Since f^i is non-increasing $\int_e^{\overline{e}} f^i(\eta) \leq f^i(0)(\overline{e} - e) = \epsilon$. That, along with the fact that P_i^j is non-negative, and (8), implies

$$\int_{\overline{e}}^\xi f^i(\eta)\, d\eta - \sum_{j\neq i} \int_{\overline{e}}^\xi r_i^l P_i^j \left((\eta - a_l^i) r_i^l\right)\, d\eta > 0.$$

If $\xi > \overline{e}$, then for some $\delta > 0$, $f^i(\overline{e} + \delta) > \sum_{j\neq i} r_i^l P_i^j \left((\overline{e} + \delta - a_l^i) r_i^l\right)$, which contradicts (7).

If $\xi \leq \overline{e}$, then $f^i(\overline{e}) < \sum_{j\neq i} r_i^l P_i^j \left((\overline{e} - a_l^i) r_i^l\right)\, d\eta$. But, since both f^i and P_i^j are continuous from the left,(7) implies that $f^i(\overline{e}) \geq \sum_{j\neq i} r_i^l P_i^j \left((\overline{e} - a_l^i) r_i^l\right)\, d\eta$, which is a contradiction. □

As stated above, for stability of PSP, we assume that demand is elastic for all players. However, the broker does not satisfy the smoothness (continuous derivative) condition. From Proposition 2, the broker's valuation,

as a function of the (scalar) expected bottleneck capacity $\min_{l \neq i} e_i^l$, is piecewise linear and concave (the derivative is the "staircase" function shown in Figure 2). Thus, we need to assume that brokers apply some smoothing in deriving the buy-side valuation from the sell-side demand, e.g. by fitting a smooth concave curve to the piecewise linear one.

Proposition 5 *(Stability) In a game consisting of arbitrarily networked PSP auctions, where all the players have utility functions of the form (3), if all players have elastic demand, then for any $\epsilon > 0$, there exists a (truthful) network-wide ϵ-Nash equilibrium.*

Proof: See [15]. □

4 Dynamics

The static game analysis of the previous section establishes the optimal strategies and the existence of a stable operating point, but does not give any indication as to which particular point will be reached. In [14], it is shown that with static sellers, the equilibrium reached by the buyers is efficient (maximizes value). Here, however, the dynamics of the brokers may affect the outcome.

For example, if ever a broker j gets a zero allocation at one of its peers' markets, then by Proposition 1, it cannot offer any bandwidth for sale. This would lead its peers to get a zero allocation from j, and they must in turn reduce their offered quantity to zero. This can converge to an equilibrium where all the brokers of a given class offer zero capacity (our experiments indicate this effect is significantly more likely if the network topology has cycles as the case in Figure 1 where there exists connectivity from network 1 to 2, to 3 and then back to 1.).

Such occurrences are not merely artifacts of PSP or of any particular pricing mechanism. They are a fundamental issue of peering. They may correspond to what is observed in the real world[6], and has been studied in the literature [1]. However, an important feature of any pricing architecture is whether it enables simple and efficient stabilizing mechanisms to avoid

[6] In recent years, some large ISPs have decided it is not in their interest to peer with some others because they would do better by selling the bandwidth directly to their own customers. This has occurred mainly with large backbone providers in relation to small access networks.

the risk of "dis-peering", i.e. a drop to zero traffic exchange between peers, *in the cases where a mutually beneficial positive allocation could be reached instead.*

In what follows, we will use simulation to study the effect of different stabilizing mechanisms on pricing under a realistic service provisioning scenario. We consider two classes of services, and hence, two SBBs in each sub-network:

- class 2 is for reliable and high quality service (e.g. the virtual leased line service considered by the EF PHB), and;
- class 1 is for adaptive multimedia applications with less stringent quality requirements (e.g. the class two considered by the AF PHB Groups);

The class 0 best-effort service gets the remaining capacity. It is charged on flat rate and does not participate in the bandwidth auction market.

The simulation network has a mesh topology of three networks as shown in Figure 1, two are access networks: argo and bongo and one is backbone network: maraca. The amount of bandwidth of sale in each network are: argo: 40 Mbps, bongo: 40 Mbps, and maraca: 150 Mbps. Each service class has 30 end users: 20 with maximum bandwidth requirement of T1 rate ($\overline{q}_i$ = 1.5 Mbps) and the other 10 with T3 rate ($\overline{q}_i$ = 40 Mbps). To simulate the dynamics of subscribers switching among service providers, each user is modulated by an ON-OFF Markov process. During ON period, the user is connected randomly to one of the three three networks based on a uniform load distribution among argo, bongo and maraca networks. During OFF period, the user terminates subscription to the current network and remains idle. In simulation, ON and OFF intervals are exponentially distributed with mean of 720 and 7200 time units. This ON-OFF process allows us to simulation the dynamics caused by an average 9% change in subscribers.

The different degree of over-provisioning for the two service classes is reflected in the routing factors r_j^l that are set according to Table 1. For class 1, a 30% over-provisioning is added to the average traffic load. For class 2, the routing parameters are set based on the worst case traffic load, i.e. the allocated quantity at all the inter-network links should be large enough for the aggregated traffic demand. Besides this, an extra over-provisioning of 10% is added to all routing parameters to reduce delay jitter (assuming that the capacity required for class 2 service is the peak rate).

buyer	*seller*	*service classes*	
		class 1 (AF)	class 2 (EF)
argo SBB	argo RBS	1.30	1.10
	argo SBB	0.39	1.10
	bongo SBB	0.26	1.10
	maraca SBB	0.65	1.10
bongo SBB	bongo RBS	1.30	1.10
	argo SBB	0.26	1.10
	bongo SBB	0.39	1.10
	maraca SBB	0.65	1.10
maraca SBB	maraca RBS	1.30	1.10
	argo SBB	0.13	1.10
	bongo SBB	0.13	1.10
	maraca SBB	1.04	1.10

Table 1: Inter-network Routing Factor: r_i^j

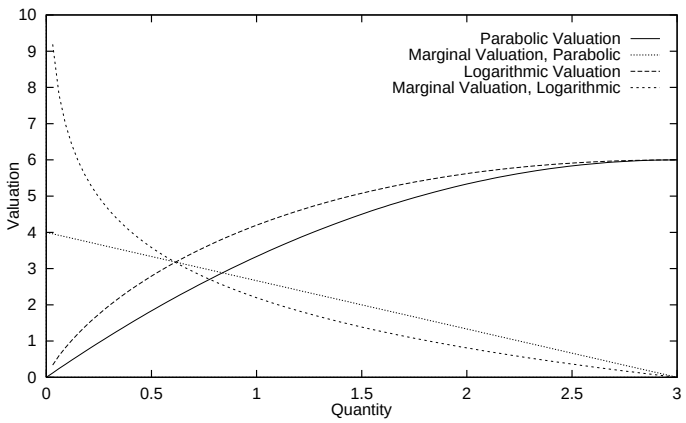

Figure 3: Valuation and Marginal Valuation Functions, $\overline{a}_i = 3$ and $\overline{\theta}_i = 6$

In Section 3.4, we assumed a very general form (i.e. elastic demand) for a user's valuation. Further specification of users' valuations requires a market study on actual Internet users (see for example [18]).[7] In the simulations, we give our users a parabolic valuation

$$\theta_i(a) = -\frac{\overline{\theta}_i}{\overline{a}_i{}^2}(a \wedge \overline{a}_i)^2 + \frac{2\overline{\theta}_i}{\overline{a}_i}(a \wedge \overline{a}_i),$$

where $\overline{a}_i$ is the line rate, $\overline{\theta}_i$ is the maximum valuation. The marginal valuation $\theta_i'(a)$ has the linear form with maximum at $\theta_i'(0)$:

$$\theta_i'(a) = \frac{2\overline{\theta}_i}{\overline{a}_i{}^2}(\overline{a}_i - (a \wedge \overline{a}_i)).$$

In simulation, $\overline{\theta}_i$ is generated randomly from a uniform distribution on [0, 10]*$\overline{a}_i$ so that $\theta_i'(0)$ is uniformly distributed in [0, 20].

As mentioned at the end of Section 3.4, the broker's buy-side valuation must be smoothed. We select a logarithmic form:

$$\theta_i(a) = \frac{\overline{\theta}_i}{\overline{a}_i}(a \wedge \overline{a}_i)(1 + \ln \frac{\overline{a}_i}{a \wedge \overline{a}_i})$$

To fit the curve to the demand, the broker dynamically sets $\overline{a}_i = \sum_j q_j^i$ and $\overline{\theta}_i = \sum_j q_j^i p_j^i$. The marginal price function $\theta'(a)$ has the form:

$$\theta'(a) = \frac{\overline{\theta}_i}{\overline{a}_i} \ln \frac{\overline{a}_i}{a \wedge \overline{a}_i}$$

where as a approaches zero, the marginal valuation approaches infinity (see Figure 3.) In some circumstances, this last feature can be crucial. A finite marginal valuation would make it possible for the broker to be completely shut-out (i.e. $a_j^l = 0$ at some peer l where enough users have very higher valuations), leading to a downward spiral as described above.

However, our experiments show this feature alone is not sufficient. Figure 4 illustrates the buy-side and sell-side bandwidth quantity at class 2 SBB in argo network over a time interval of 2500 time unit. The brokers and users bid once per time unit, whereas a single time unit can represent, e.g., one hour or one day depending on the implementation time-scale. As

[7] Recall that the difficulty in developing realistic models is one of the reasons why auctions are advantageous in the first place, since the mechanism itself does not need to know the valuations.

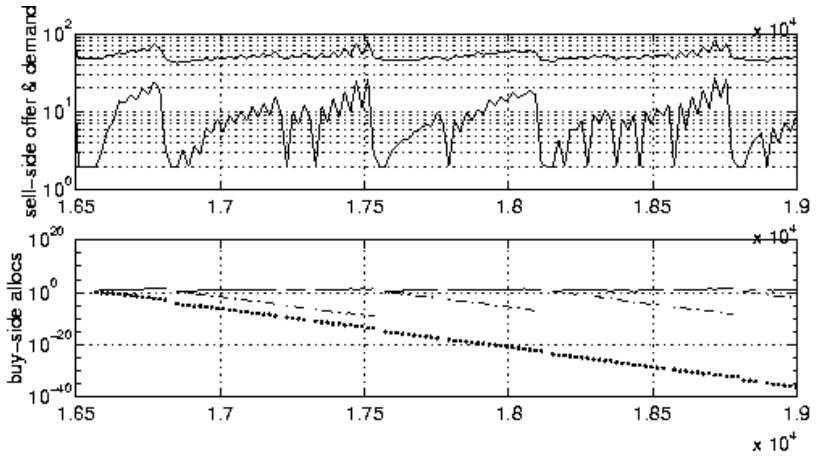

Figure 4: Example of Degenerated Equilibrium

shown in the lower part of the graph, the purchased bandwidth at argo's two peer networks degenerates to below 10^{-20} and leads to a shut-down at argo's own market. Even after we temporarily stop peering at maraca network, the recover is still short-lived.

To stabilize the market, we add the following additional heuristic mechanisms:

- "Buy-back": First, rather than simply setting $q_j^j = \min_{l \neq j} e_j^l(a)$ as required by Proposition 1, the seller can leave q_j^j equal to the maximum physical capacity, and place in its own market an artificial "buy-back" bid equal to

$$q_0^j = (q_j^j - e)^+ \ , \ p_0^j = \theta_j'(e),$$

where e broker j's expected bottleneck after the buy-side best-reply computation, as in Proposition 4. It is relatively straightforward to see that this buy-back bid effectively limits j's users to precisely the capacity that j can honor in forward to its peers. In other words, the buy-back bid ensures that the quantity constraint of Proposition 1 is automatically satisfied. However, unlike reducing q_j^j, it leaves open the possibility of increasing it back again, if there is demand at prices greater than $\theta_j'(e)$. $\theta_j'(e)$ is precisely the price at which j could obtain more capacity at its bottleneck to a peer network. If this bid comes from one of the peers, it will be satisfied and hence prevent the dis-peering phenomenon.

- As a further safe-guard against temporary drops in inter-peer allocations leading to drops in offered quantity, temporary reduction of the route-factor r_j^l helps j maintain a positive bottleneck. We find that the following formula

$$\hat{r}_j^l = r_j^l \exp\left(\frac{e_j^l - q_j^j}{q_j^j}\right)$$

  achieves the goal of filtering out temporary drops, while maintaining in the long term $\hat{r}_j^l \approx r_j^l$. This could even be physically interpreted as a market-responsive re-routing decision, when the broker recognizing that the peer network is over-priced.

Analyzing the dynamics of these mechanisms is beyond the scope of this paper. They are presented to convey the insight that some kind of mechanisms of this type are necessary and need further investigation

The performance traces at the six SBBs (two in each of the three networks) are presented in Figures 5 to 10. Each figure contains four graphs showing in clockwise the traces of market price, market demand and sell-side offered bandwidth, buy-side purchased bandwidth, and profit. In the graph showing demand and sell-side offered bandwidth, the upper curve shows the demand while the lower curve shows the offered bandwidth. In the graph showing buy-side purchased bandwidth, the top curve illustrates the trace of allocation at RBS while the bottom two are the traces of allocations at the two peer SBBs.

It is clear that with the additional stabilizing mechanisms, we prevent the dis-peering phenomenon shown earlier. In addition, one can observe that:

- as shown in the buy-side allocation graph in Figure 5 and 6, at the backbone maraca network, the two class SBBs are able to purchase exactly the same buy-side allocations from the two peers (i.e. the two curves are superimposed). This is the optimality for a broker because of no waste at buy-side;

- in both argo and bongo networks, class 1 consistently gets more allocation than class 2, i.e. 25 Mbps vs 15 Mbps. This because the route provisioning factors for class 1 are less than that of class 2, and the demands are equal. To get an equal allocation, users of the more resource intensive EF-based services would have to have a higher willingness to pay;

- in the access networks bongo and argo, where demand greatly exceeds the available capacity, class 2 which is more highly provisionned has higher prices as expected; however competition at the RBS level keeps the prices close, and class 2 is penalized mainly by lower allocation;
- in the backbone maraca network, 60% of the capacity is left unpurchased even though the total market demand exceeds the capacity. This is because the allocations for class 1 and 2 are constrained by the bottlenecks to the two peer networks that have much less capacity. This demonstrates that best-effort class 0 traffic can get positive leftover allocation because it does not need any guaranteed bandwidth provisioning;
- the occasional price spikes in all the figures are part of control mechanism, caused by a SBB's artificial bids and the reduction of route-factor to stabilize the offered quantity at sell side;
- an SBB broker may not always be profitable because it needs to take risk to stabilize the sell-side offered quantity. If one SBB broker's profit is negative, then its peer SBB or its RBS is making the profit;
- there is no clear conclusion on which class is more profitable. In maraca network class 1 is more profitable while in argo network class 2 is more profitable.

5 Conclusion

In this paper, we present a decentralized auction-based pricing approach for differentiated service Internet. Our game theoretical analysis identifies the best strategies for end users and bandwidth brokers. Even though the analysis proves the existence of efficient equilibrium in the form of static game, our initial simulation results indicate that the interactions of the brokers' strategies can form negative-feedback control loops and lead the game to zero equilibrium. However, with our propose stabilizing mechanisms, we demonstrate that the market price can be maintained at non-zero equilibrium. Our future work will be to study the interactions of brokers' strategies under multilateral peering scenario and investigate the conditions to maintain a stable market.

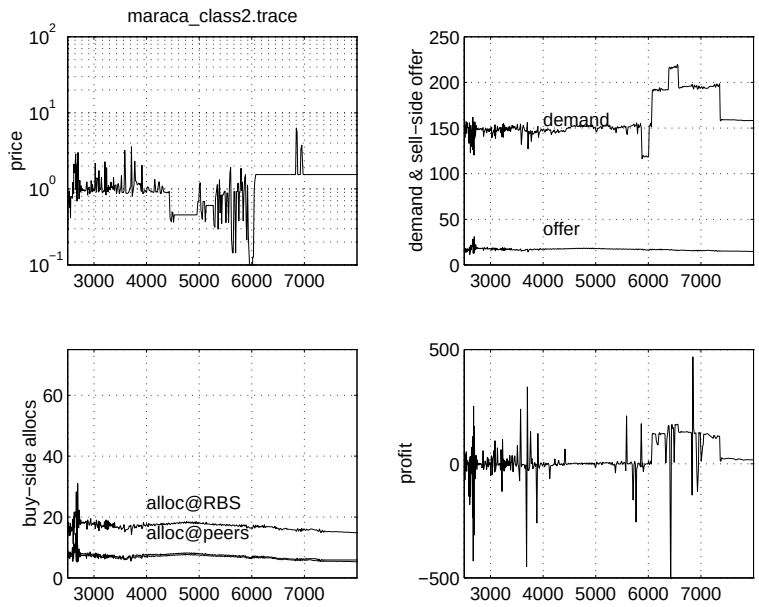

Figure 5: Class 2 Trace at maraca

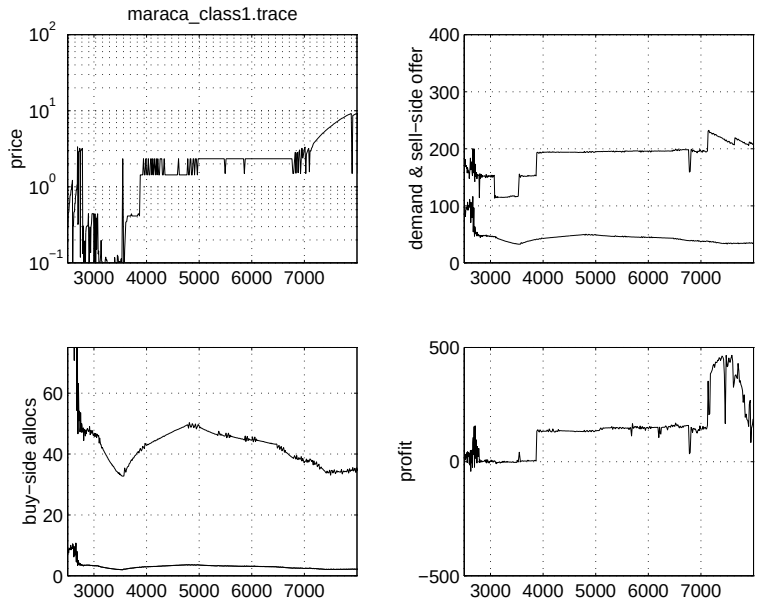

Figure 6: Class 1 Trace at maraca

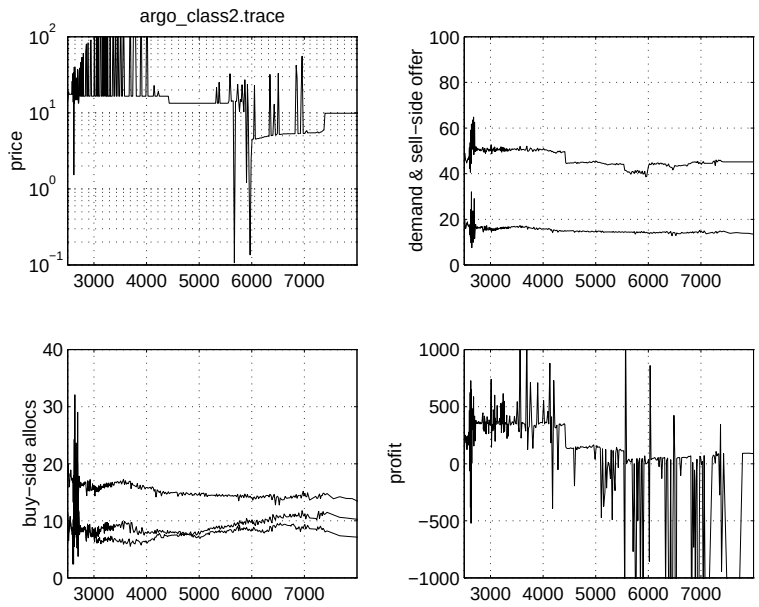

Figure 7: Class 2 Trace at argo

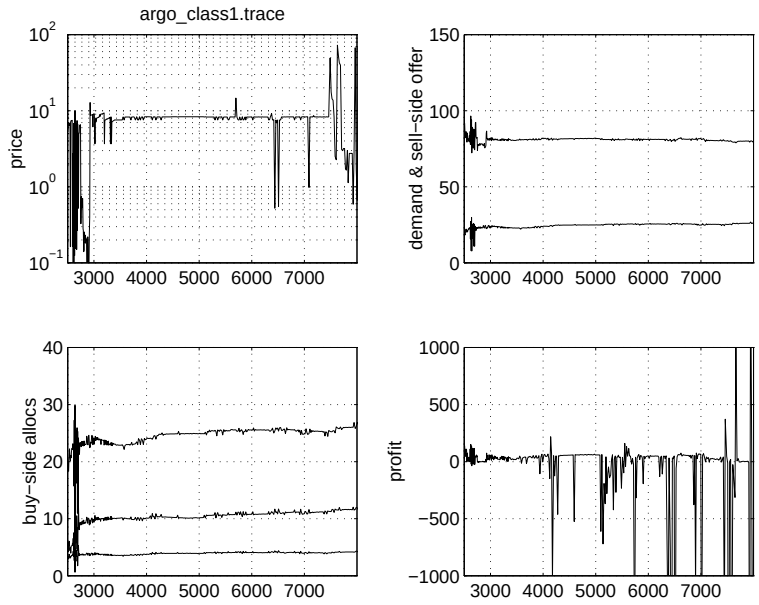

Figure 8: Class 1 Trace at argo

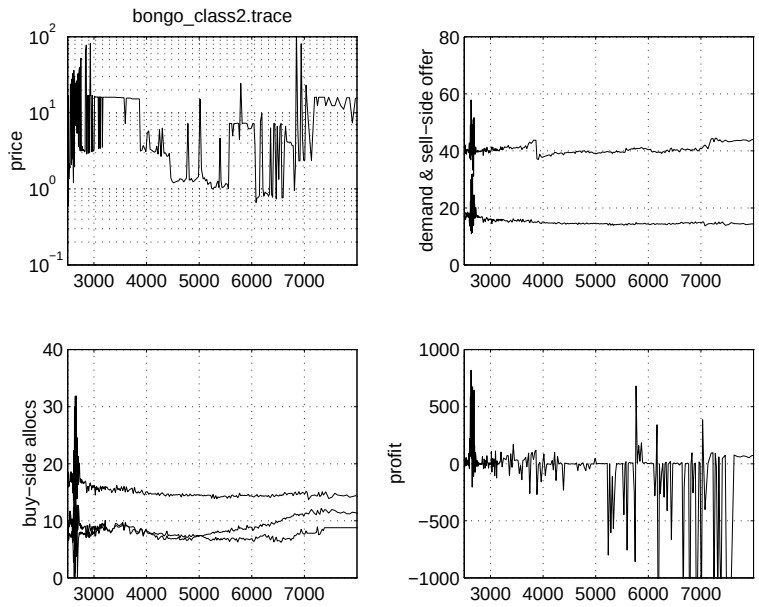

Figure 9: Class 2 Trace at bongo

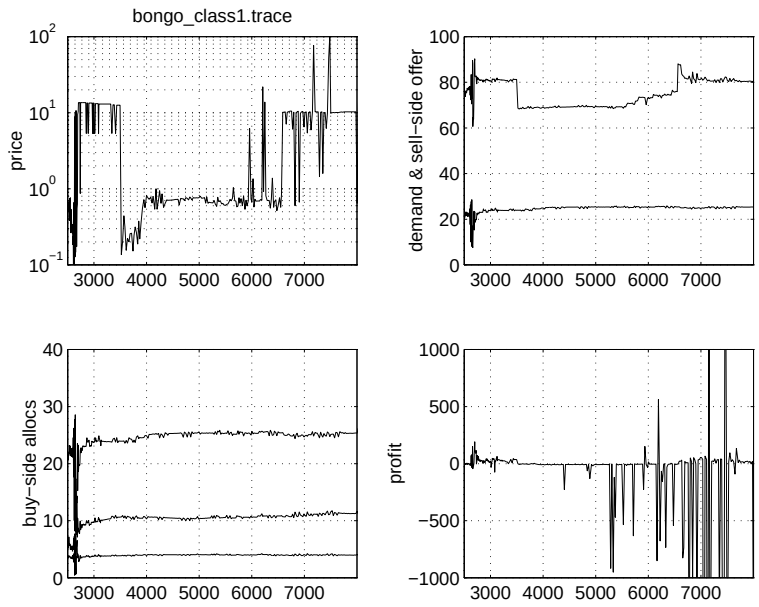

Figure 10: Class 1 Trace at bongo

References

[1] P. Baake and T. Wichmann. On the economics of internet peering. *Netnomics*, 1(1), 1999.

[2] D. Clark. Internet cost allocation and pricing. In L. W. McKnight and J. P. Bailey, editors, *Internet Economics*. MIT Press, 1997.

[3] D. Clark and W. Fang. Explicit allocation of best-effort packet delivery service. *IEEE/ACM Trans. Networking*, 6(4):362–373, Aug 1998.

[4] E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 8:17–33, 1971.

[5] Internet Engineering Task Force. Differentiated services working group. http://www.ietf.org/html.charters/diffserv-charter.html.

[6] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[7] T. Groves. Incentives in teams. *Econometrica*, 41(3):617–631, July 1973.

[8] A. Gupta, D. 0. Stahl, and A. B. Whinston. Priority pricing of integrated services networks. In L. W. McKnight and J. P. Bailey, editors, *Internet Economics*. MIT Press, 1997.

[9] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. Assured forwarding PHB group. Internet Draft – work in progress, Nov 1998. http://www.ietf.org/internet-drafts/draft-ietf-diffserv-af-03.txt.

[10] M. T. T. Hsiao and A. A. Lazar. Optimal decentralized flow control of markovian queueing networks with multiple controllers. *Performance Evaluation*, 13(3), Dec 1991.

[11] V. Jacobson, K. Nichols, and K. Poduri. An expedited forwarding PHB. Internet Draft – work in progress, Nov 1998. http://www.ietf.org/internet-drafts/draft-ietf-diffserv-ef-01.txt.

[12] F. P. Kelly. Charging and rate control for elastic traffic. *European Trans. Telecommunications*, 8, 1997.

[13] K. Kilkki, J. Ruutu, and O. Strandberg. High quality and high utilization - incompatible objectives for internet? In *IEEE/IFIP* 6^{th} Int. Workshop on Quality of Service, Napa Valley, CA, May 1998.

[14] A. A. Lazar and N. Semret. Design, analysis and simulation of the progressive second price auction for network bandwidth sharing. Technical Report CU/CTR/TR 497-98-21, Columbia University, 1998. http://comet.columbia.edu/publications.

[15] A. A. Lazar and N. Semret. The progressive second price auction mechanism for network resource sharing. In 8^{th} *Int. Symp. on Dynamic Games and Applications*, Maastricht, July 1998.

[16] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, February 1981.

[17] K. Nichols, V. Jacobson, and L. Zhang. A two-bit differentiated services architecture for the internet, 1997. ftp://ftp.ee.lbl.gov/papers/dsarch.pdf.

[18] B. Rupp, R. Edell, H. Chand, and P. Varaiya. INDEX: A platform for determining how people value the quality of their internet access. In *IEEE/IFIP* 6^{th} Int. Workshop on Quality of Service, Napa Valley, CA, May 1998.

[19] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *ACM Comput. Commun. Review*, 26(2):19–43, 1996.

[20] I. Stoica and H. Zhang. LIRA: An approach for service differentiation in the internet. In *Proc. of NOSSDAV'98*, Cambridge, England, July 1998.

[21] W. Vickrey. Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance*, 16, 1961.

[22] Z. Wang. A case for proportional fair sharing. In *IEEE/IFIP* 6^{th} Int. Workshop on Quality of Service, Napa Valley, CA, May 1998.

A Broker's buy-side coordination

Lemma 1 *(Broker coordination) Let $j \in \mathcal{I}$ be a broker. For any profile s, $s_j = (q_j, p_j)$, let $a \equiv a(s)$ be the allocations that would result, and $m = \arg\min_k e_j^k(a)$. Then, a better reply for the broker is $x_j = (z_j, p_j)$, where $\forall l \neq j$*

$$z_j^l = \left[e_j^m(a) - a_l^j\right] r_j^l.$$

That is, $u_j(x_j; s_{-j}) \geq u_j(s)$. *Moreover,*

$$a_j^l(z_j, p_j) = z_j^l. \tag{9}$$

Proof: To avoid cluttered notation, since s_{-j} is fixed, we will omit it, writing, e.g., $u_j(.,.) \equiv u_j((.,.); s_{-j})$. Also, the argument of the function will be omitted when it is simply s, so that $u_j \equiv u_j(s_j) \equiv u_j(s_j; s_{-j})$. Note that, since we are holding all the other players fixed, and varying only the buy-side of player j, only the quantities with subscript j will change. In particular, a_l^j remains the same throughout.

We will show that

$$u_j \equiv u_j(q_j, p_j) \leq u_j(z_j, p_j) \tag{10}$$

Now, $\forall l \in \mathcal{I}$,

$$\begin{aligned} z_j^l &= \left[e_j^m(a) - a_l^j\right] r_j^l \\ &\leq \left[e_j^l(a) - a_l^j\right] r_j^l = a_j^l \\ &\leq \left[q_l^l - \sum_{p_k^l \geq p_j^l, k \neq j} q_k^l\right]^+, \end{aligned} \tag{11}$$

where the last line follows from (4). Now using (4) again, we get

$$a_j^l(z_j, p_j) = \left[q_l^l - \sum_{p_k^l \geq p_j^l, k \neq j} q_k^l\right]^+ \wedge z_j^l = z_j^l = \left[e_j^m(a) - a_l^j\right] r_j^l,$$

where the second equality follows from (11), and the last is by definition. This proves (9). Thus, we have $e_j^l(a(z_j, p_j)) = a_j^l(z_j, p_j)/r_j^l + a_l^j = e_j^m(a)$, and this holds $\forall l \neq j$. Therefore, by Proposition 2, $\theta_j(a(z_j, p_j)) = \theta_j(a)$, i.e., changing the bids from (q_j, p_j) to (z_j, p_j) does not change j's bottleneck value. Therefore,

$$\begin{aligned} u_j(z_j, p_j) - u_j &= \sum_{l \neq j} c_j^l - c_j^l(z_j, p_j) \\ &= \sum_{l \neq j} \int_{a_j^l(z_j, p_j)}^{a_j^l} f^l(q_l^l - z)\, dz. \end{aligned}$$

Now $\forall l$, $e_j^m(a) \leq e_j^l(a) \Rightarrow z_j^l/r_j^l + a_l^j \leq a_j^l/r_j^l + a_l^j \Rightarrow a_j^l \geq z_j^l \geq a_j^l(z_j, p_j)$, where the last inequality follows from (4). That along with the fact that $f^l \geq 0$ implies $u_j(z_j, p_j) - u_j \geq 0$. □