Market Pricing of Differentiated Internet Services

Nemo Semret, Raymond R.-F. Liao, Andrew T. Campbell and Aurel A. Lazar
Center for Telecommunications Research, Columbia University
New York, NY, 10027-6699, USA.
mnl@comet.columbia.edu

Abstract

This paper presents a decentralized auction-based approach to pricing of edge-allocated bandwidth in a differentiated services model for 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 show the existence of network-wide equilibria. We investigate the system dynamics by simulating a scenario with three interconnected networks, and two types of services built on the proposed standard expedited forwarding (EF) and assured forwarding (AF) per-hop-behaviors.

1 Introduction

The recent development of a 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 [25], 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, 18], the precise development of pricing mechanisms is still at its early stages. In the Simple Integrated Media Access model [14], 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 remains fixed. Similarly, in the User-Share Differentiation proposal [28], pricing is based on the user share that is allocated over long time scales. 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 discussed in [13].

The space of network resource pricing schemes has many dimensions (for a complete taxonomy of network pricing see [22]). One is "where" the capacity abstraction takes place: at each hop inside the network or at the edges [24]. 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). In more sophisticated approaches, the seller assumes the functional form of demand and adjusts prices by on-line optimizations [9, 1, 15, 20, 7]. These pricing schemes are "model-based", in that the relationship between demand and price (and possibly time) is assumed in an a-priori formula. Knowledge of this model and its parameters is precisely the "information requirement" described above.

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 more compelling. With suitably designed rules, auctions can achieve efficient allocations (i.e. value maximizing) with minimal a priori information.

In this paper, we investigate the feasibility of auctioning capacity in a diff-serv model. We show through game theoretic analysis and simulation that the Progressive Second Price (PSP) auction of [16] can provide stable pricing in a diff-serv bandwidth market. The PSP mechanism achieves the economic objectives of incentive compatibility and efficiency, while being realistic in the engineering sense (small signalling load and computationally 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 theoretic analysis of the scheme. The results of this section extend analysis of the single sharable resource auction of [16] to the case of multiple networked resources, in an edge-capacity allocation framework. Following this, in Section 4, we use simulations to study the capacity market dynamics.

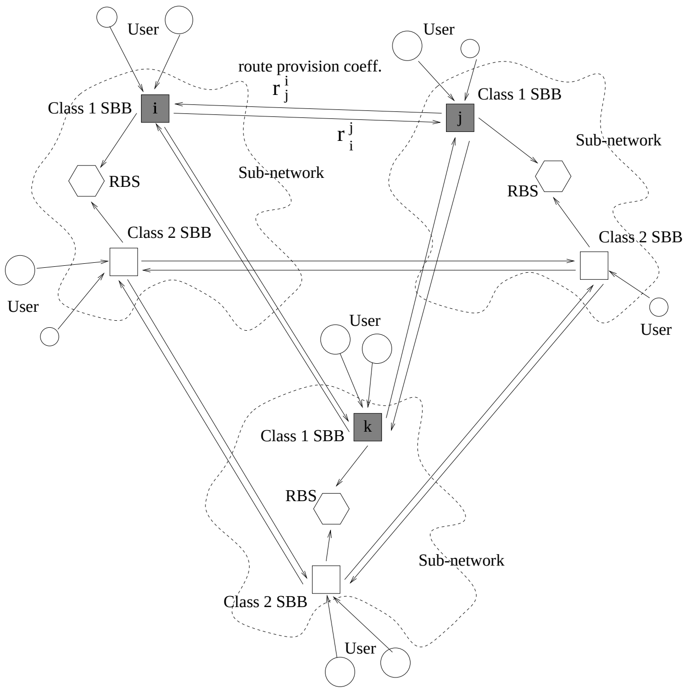

Figure 1: *2-tier Auction Pricing Framework for a Diff-serv Internet*

2 Market Pricing Framework

Our network model assumes that each network can be abstracted into a single bottleneck capacity (e.g., as a "Norton-equivalent" [11]). The capacity may be represented by an absolute amount of bandwidth, or some relative metrics such as user shares [28] or resource tokens [26]. 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[1] 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 similar to [7] 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 [19]. 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 (e.g., web site, intra/extranet, virtual private networks) to a particular service offered by a particular provider, and buying capacity from an SBB.

[1] 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.

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 [16], 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 $s_i^j = (0, 0)$.

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.

Unless otherwise indicated, when sub/superscripts are omitted, the notation refers to the vector obtained by letting it range over all values. For example, 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 as discussed in 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.[2] For notational convenience, fix $r_j^k = 1$, when k is j's RBS. 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,

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

We call

$$e_j \triangleq \min_{l \neq j} e_j^l(a) \tag{3}$$

the *expected bottleneck* capacity for the service offered by 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. □

Remark: The obvious way for a broker to satisfy Proposition 1 is simply setting $q_j^j = \min_{i \neq j} e_j^i(a)$. Alternately, 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 $e = \min_{i \neq j} e_j^i(a)$. Note that this artificial player $0 \notin \mathcal{I}$. 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 $\theta_j'(e)$. As we will become apparent through Proposition 4 below, $\theta_j'(e)$ is precisely the price at which j could obtain more capacity at its bottleneck to a peer network.

Remark: r_j is a vector of route-provisioning coefficients and can accommodate different types of brokers. We do not explicitly consider the per-hop behaviors per se, which of course are essential in assuring the service quality on the packet time-scale. On our level of abstraction, only the vector of provisioning coefficients r_j differentiates broker j and the service it offers. A broker is characterized by the type of Service Level Agreement (SLA) that it offers, e.g.:

- Expected capacity service agreement: one type of SBB may offer 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 Assured Forwarding (AF) per-hop behaviors [10]. In this case, r_j^l is the expected[3] fraction of the total traffic entering j that is routed 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.[4]
- 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 Expedited Forwarding (EF) per-hop behavior [12]. 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 [12]. Degrees of over-provisioning must also be used to differentiate among AF classes for example [10]. 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

We model buyers as **bottleneck buyers**, i.e. each buyer $i \in \mathcal{I}$ seeks to maximize its utility

$$u_i = \theta_i \circ e_i(a) - \sum_j c_i^j, \tag{4}$$

[2] We assume that service providers block "loop-back" traffic, i.e. traffic going from l through j and back to l. If that is not the case, then the summation in (2) would be over all i.

[3] 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.

[4] 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.

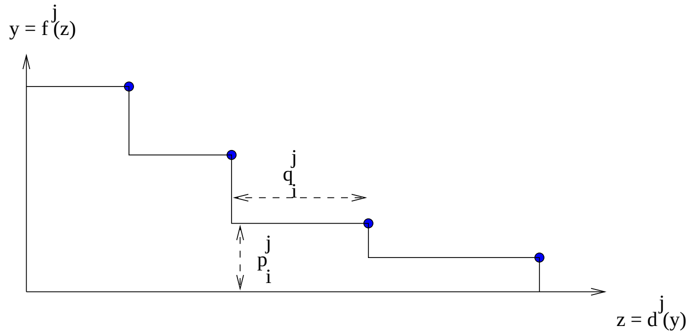

Figure 2: *Demand Curve for a Broker j.*

where e_i is as in 3, and θ_i is the buyer's **valuation** function, which is private information. As the name indicates, the value of an allocation to a buyer depends only on a scalar bottleneck $e_i(a)$ which is a function of the allocated quantities at all the resources. Other players (including the seller) only see the buyer's bid and not the valuation that lead buyer to make that bid.

If the buyer is a user, then $e_i = a_i^j$, and (4) 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 each possible quantity of resource. It can be based on economic and/or information theoretic considerations (see the appendix in [16]).

If the buyer is a broker, the natural utility is the potential profit so θ_j, the broker's buy-side valuation, is the potential revenue from the sale (on the sell-side) of the capacities obtained on the buy-side. The potential revenue is derived from the demand on the sell-side: let $\forall y \geq 0$,

$$d^j(y) \triangleq \sum_{p_k^j \geq y} q_k^j,$$

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

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

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 $f^j(z)$. Its buy-side valuation is*

$$\theta_j(x) = \int_0^x f^j(z)\,dz.$$

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

Proof: Since the broker seeks to maximize profit, for a given allocation a, it will sell as much as possible; thus by Proposition 1, $q_j^j = e_j$. If e_j decreases by δ, then q_j^j must be reduced by δ. The value to j of the lost quantity is the revenue j could 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,

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

and the result follows. □

3.4 Analysis

The design of our Progressive Second Price auction (PSP) appears in [16][5]. The mechanism is defined by: $\forall i, j \in \mathcal{I}$,

$$a_i^j(s) \equiv a_i^j(s^j) = q_i^j \wedge \left[q_j^j - \sum_{p_k^j \geq p_i^j, k \neq i} q_k^j\right]^+, \quad (5)$$

$$c_i^j(s) \equiv c_i^j(s^j) = \sum_{k \neq i} p_k^j \left[a_k^j(0; s_{-i}^j) - a_k^j(s_i^j; s_{-i}^j)\right], \quad (6)$$

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

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

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,

$$c_i^j = \int_0^{a_i^j} P_i^j(z)\,dz. \quad (7)$$

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 [27] which are for non-divisible objects. PSP bears some similarity to Clarke-Groves mechanisms [4, 8]. 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 [17, 16] 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 (signaling) is very important.

We 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 (i.e., $theta_i'' < 0$ if θ_i'' is well-defined) on $[0, \bar{q}_i]$, and non-increasing ($\theta_i'' \leq 0$) on $[\bar{q}_i, \infty)$.

[5] PSP was first presented at the *DIMACS Workshop on Economics, Game Theory, and the Internet*, Rutgers, NJ, April 1997, and a generalized analysis at the 8^{th} *International Symposium on Dynamic Games and Applications.*, Maastricht, The Netherlands, July 1998

Under elastic demand analyzed as a complete information game, the PSP auction for a single arbitrarily divisible resource (e.g., bandwidth on one link in a network) has the following properties which are proven in [16]:

- 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_{-i}^l, an ϵ-best reply for player i is $t_i^l = (v_i^l, w_i^l)$, such that*

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

and

$$w_i^l = \theta_i'(v_i^l).$$

That is, $\forall s_i^l$, $u_i(t_i^l; s_{-i}^l) \geq u_i(s_i^l; s_{-i}^l) - \epsilon$.

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 [16]. □

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*

$$e = \sup\left\{h \geq 0 : f^i(h) > \sum_{l \neq i} P_i^l\left((h - a_l^i)r_i^l\right) r_i^l\right\} - \epsilon/f^i(0), \tag{8}$$

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., $\forall s_i$, $u_i(t_i; s_{-i}) \geq u_i(s_i; s_{-i}) - \epsilon$.

Proof: Since f^i is non-increasing and $\forall l$, P_i^l is non-decreasing, (8) implies $f^i(e) > \sum_{l \neq i} P_i^l(v_i^l) r_i^l$, and therefore $\forall l \neq i$,

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

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 (10) 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

$$\int_e^\xi f^i(\eta)\,d\eta - \sum_{j \neq i} \int_e^\xi r_i^l P_i^j\left((\eta - a_l^i)r_i^l\right)\,d\eta > \epsilon. \tag{9}$$

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 (9), 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 (8).

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,(8) 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.

Unlike the proof of the the broker strategy, the proofs of the following results are not essential to intuitive understanding of the game and are omitted due to space constraints.

Proposition 5 *(Equilibrium) In a game consisting of arbitrarily networked PSP auctions, where all buyers have utilities of the form (4), and sellers are static, under elastic demand, for any $\epsilon > 0$, there exists a (truthful) network-wide ϵ-Nash equilibrium.*

Proof: See [22], Chapter 3. □

At such equilibria, the allocations are efficient (i.e. arbitrarily close to the value-maximizing allocations).

Proposition 6 *(Efficiency) Let a^* be the equilibrium allocations. Under elastic demand, if in addition $\forall i \in \mathcal{I}$, if θ_i'' exists and for some $\kappa > 0$, $\theta_i'' \geq -\kappa$,*

$$\max_{\mathcal{A}} \sum_i \theta_i \circ e_i(a) - \sum_i \theta_i \circ e_i(a^*) = O(\epsilon/\delta + \kappa\delta),$$

where $\mathcal{A} = \{a \in \prod_j [0, q_j^j]^I : \sum_i a_i^j \leq q_j^j\}$, for any $\delta \leq \min_i\{e_i(a^*) : e_i(a^*) > 0\}$.

Proof: See [22], Chapter 3. □

The bound $\epsilon/\delta + \kappa\delta$ is minimized when $\delta = \sqrt{\epsilon/\kappa}$. Thus, the strongest statement that can be made here is that as long as $\min_i\{e_i(a^*) : e_i(a^*) > 0\} > \sqrt{\epsilon/\kappa}$ we get an inefficiency which is $O(\sqrt{\epsilon/\kappa})$.

In a dynamic auction game, $\epsilon > 0$ can be interpreted as a *bid fee* paid by a bidder each time they submit a bid. Indeed, in Propositions 3 and 4, the user will send a best reply bid as long as it improves her current utility by ϵ, and the game can only end at an ϵ-Nash equilibrium.

4 Dynamics

The strategic game analysis of the previous section establishes the optimal strategies and the existence of a stable and efficient operating point, but does not give any indication as to which particular equilibria will be reached.

In what follows, we will use simulation to further study the diff-serv PSP framework 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);

At all times, best-effort traffic can use any unallocated or allocated but unused 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. At the beginning of an 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 120 and 12 time units.

		service classes	
buyer	*seller*	class 1 (AF)	class 2 (EF)
argo SBB	argo RBS	1.0	1.0
	argo SBB	0.3	0.5
	bongo SBB	0.2	0.5
	maraca SBB	0.5	0.5
bongo SBB	bongo RBS	1.0	1.0
	argo SBB	0.2	0.5
	bongo SBB	0.3	0.5
	maraca SBB	0.5	0.5
maraca SBB	maraca RBS	1.0	1.0
	argo SBB	0.1	0.5
	bongo SBB	0.1	0.5
	maraca SBB	0.8	0.5

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

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.

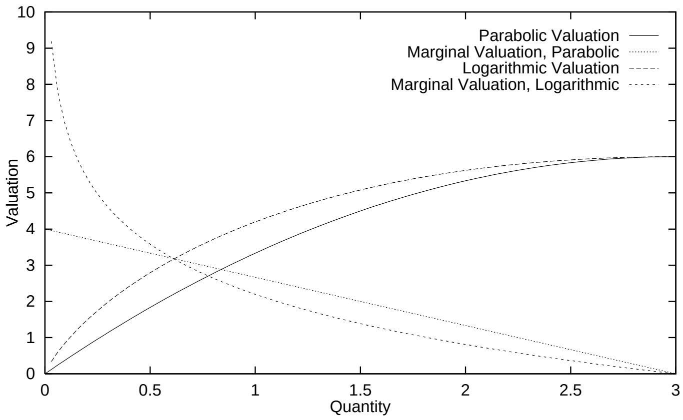

Figure 3: *Valuation and Marginal Valuation Functions,* $\overline{a}_i = 3$ *and* $\overline{\theta}_i = 6$

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 [21]).[6] 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 our simulation, for class 1, $\overline{\theta}_i$ is generated randomly from a uniform distribution on [0, 10]*$\overline{a}_i$ so that $\theta_i'(0)$ is uniformly distributed in [0, 20]. For class 2, the range is $[20, 40]$. Note that, from Proposition 3, $\theta_i'(0)$ is the highest unit price user i would ever be willing to pay, and, since the valuation is concave (elastic demand), that occurs when the user has been "squeezed" to almost zero allocation. We give class 2 users higher valuations because, as the simulations will show, due to the more conservative provisioning required, market forces naturally make class 2 more expensive.

As mentioned in 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:

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

[6] Recall that the difficulty in developing realistic models is one of the reasons why auctions are advantageous in the first place, since the (run-time) mechanism itself (5)-(6) does not need to know the valuations.

where as a approaches zero, the marginal valuation approaches infinity (see Figure 3.) In some circumstances, this last feature can be useful. A finite maximum 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), and when one broker is shut-out, so are all its peers, and the service is no longer offered.

The simulation traces of the state of the six SBBs (two in each of the three networks) are presented in Figures 4 to 9. Each figure contains four plots showing the market price, the demand, the broker's buy-side purchased bandwidth (from peers), and profit. For each broker, in the demand plot, the solid line indicates the total requested quantity of bandwidth (solid line) and the number of users and peers (dotted line). In the plot showing buy-side purchased bandwidth, the top line illustrates the trace of allocation at the RBS while the bottom two are the traces of allocations at the two peer SBBs. Not that since $r_i^k = 1$ when k is the RBS, *the allocation from the RBS exactly equals the bottleneck capacity that the broker is offering for sale.* In the price and profit plots, the dashed line shows the average during the whole run.

We observe that:

- as shown in the buy-side allocation plot in Figure 4 and 5, 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 to be expected since the route provisioning coefficients are the same for the two peers, and the maraca SBBs will not buy any capacity that does not improve the bottleneck;

- in all networks, despite the higher valuations, class 2 gets no more than 1 Mbps, even in the backbone network. Thus, a small difference in provisioning can lead to very large differences in the allocated capacity among the classes. To collectively get an equal allocation, users of the more resource intensive EF-based services would have to have a much higher willingness to pay. In this scenario, the market price for the class 2 is 20 times higher than for class 1.

- the occasional price spikes in all the figures are caused by arrivals or departures of T3 users that greatly change the demand. This is part of the brokers' strategic behaviour, which is in essence a distributed control mechanism.

- a broker may occasionally be unprofitable because the markets on its sell- and buy-side are not synchronized. If one SBB's profit is negative then its RBS or a peer SBB is making the profit. However, in the long run, setting the sell-side according to Proposition 1 ensures that brokers will not operate at a loss. In general, competition among SBBs in the same network for the resources offered by the RBS ensures that their profit

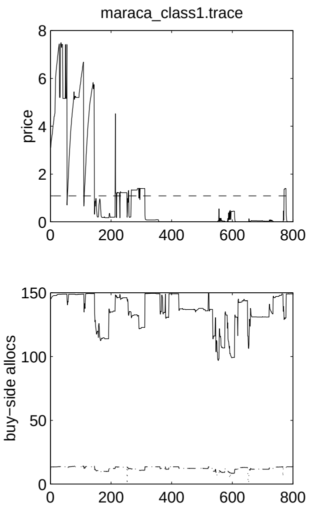

Figure 4: *Class 1 Trace at Maraca*

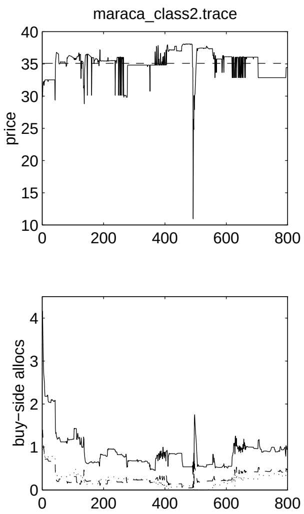

Figure 5: *Class 2 Trace at Maraca*

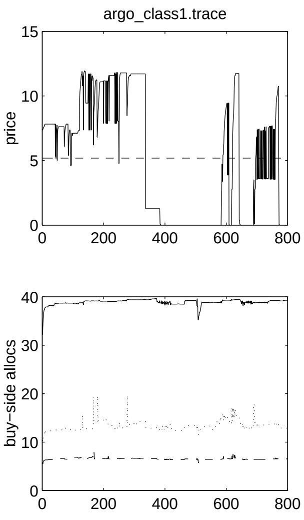

Figure 6: *Class 1 Trace at Argo*

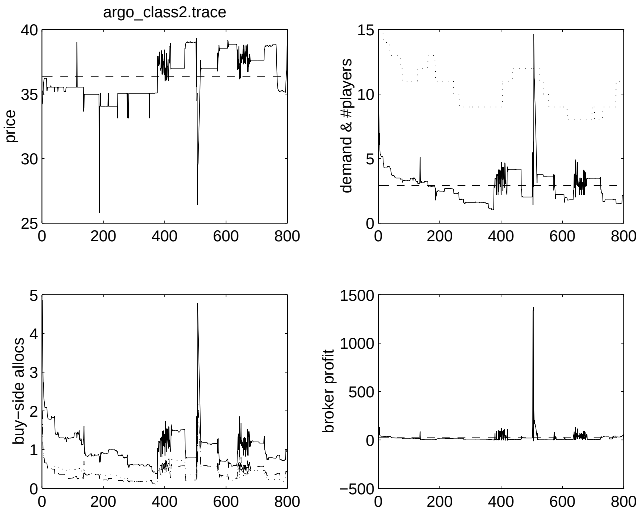

Figure 7: *Class 2 Trace at Argo*

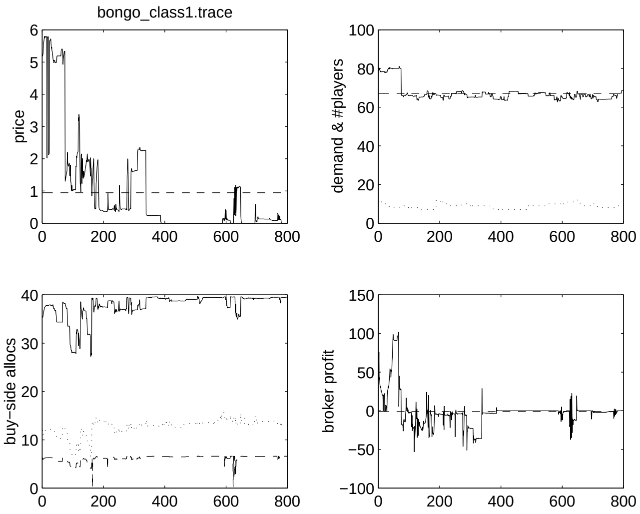

Figure 8: *Class 1 Trace at Bongo*

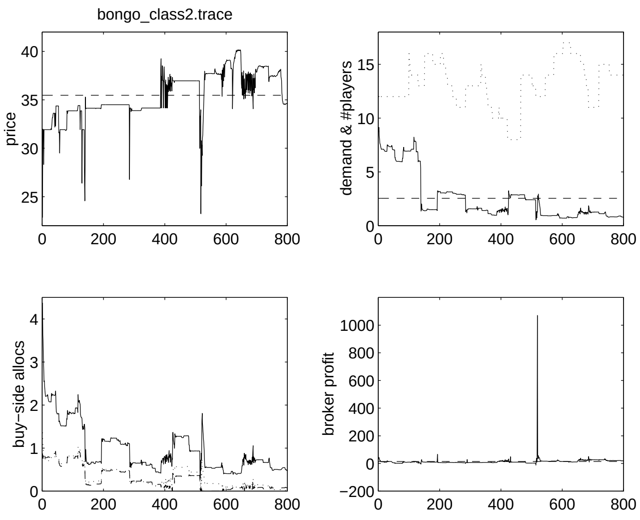

Figure 9: *Class 2 Trace at Bongo*

margins are very thin. Thus, the RBS' (bearers) reap most of the profits, indicating that the providers of differentiated services will tend to be vertically integrated (i.e. the bearers will likely also be the service providers).

5 Conclusion

We have presented a decentralized auction-based pricing approach for differentiated Internet services. Our game theoretic analysis identifies the best strategies for end users and bandwidth brokers. The analysis proves the existence of efficient stable operating points and the simulations indicate that even an aggregate 50% difference in the degree of provisioning between two services can lead to one order of magnitude difference in the market price of services, and partitioning of bandwidth between services.

The interaction between brokers has a much richer dynamic than discussed in this paper. For example, not all configurations of provisioning coefficients in the wide area network lead to convergence and stable allocations. Depending on the topology and degree of over-provisioning, the interaction between brokers can lead to oscillating allocations. On the other hand, stable operating points may lead to zero allocations for some brokers resulting in certain classes of service not being offered at all. These are not mere artifacts of the PSP or any particular pricing mechanism but are fundamental issues of peering and provisioning under edge-capacity allocation. The former case is analytically related to classical problems such as route-flapping using decentralized routing algorithms. The latter case relates to empirical evidence in a best-effort Internet where market forces abandon traditional "free-for-all" peering between networks of unequal size. These issues are analyzed in our subsequent work [23], where we present necessary and sufficient conditions for wide-area feasibility of service classes in terms of the provisioning coefficients.

References

[1] N. Anerousis and A. A. Lazar. A framework for pricing virtual circuit and virtual path services in ATM. In *Proc. IEE Int'l Teletraffic Congress*, Washington, DC, June 1997.

[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] E. Fulp, M. Ott, D. Reininger, and D. Reeves. Paying for qos: An optimal distributed algorithm for pricing network resources. In *IEEE/IFIP* 6^{th} Int. Workshop on Quality of Service, Napa Valley, CA, May 1998.

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

[9] A. Gupta, D. O. 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.

[10] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. Assured forwarding PHB group. Internet Draft – work in progress, Feb 1999. http://www.ietf.org/internet-drafts/draft-ietf-diffserv-af-06.txt.

[11] M. T. T. Hsiao and A. A. Lazar. An extension to Norton's equivalent. *Queueing Systems, Theory and Applications*, 5:401–412, 1989.

[12] 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.

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

[14] 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.

[15] Y. A. Korilis, T. A. Varvarigou, and S. R. Ahuja. Incentive-compatible pricing strategies in noncooperative networks. In *Proc. IEEE Infocomm.*, 1998.

[16] A. A. Lazar and N. Semret. Design and analysis of the progressive second price auction for network bandwidth sharing. *Telecommunication Systems* – Special issue on Network Economics, 1999.

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

[18] 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.

[19] National Research Council NRenaissance Committee. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C., 1994.

[20] I. C. Paschalidis and J. N. Tsitsiklis. Congestion-dependent pricing of network services. Technical report, Boston University, 1998.

[21] 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.

[22] N. Semret. *Market Mechanisms for Network Resource Sharing.*
PhD thesis, Columbia University, 1999. Available at http://comet.columbia.edu/publications/thesis.html.

[23] N. Semret, R. R.-F. Liao, A. T. Campbell, and A. A. Lazar. Peering and provisioning of differentiated internet services. Technical report, Columbia University, March 1999. Submitted for publication. Available at comet.columbia.edu/publications/techreports.html.

[24] 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.

[25] S. Shenker, C. Partridge, and R. Guerin. Specification of guaranteed quality of service. IETF RFC 2122, Sep 1997. ftp://ftp.isi.edu/in-notes/rfc2212.txt.

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

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

[28] 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,

$$a_j^l(z_j, p_j) = z_j^l. \tag{10}$$

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

$$u_j \equiv u_j(q_j, p_j) \leq u_j(z_j, p_j) \tag{11}$$

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

$$\begin{aligned} z_j^l &= \left[e_j^m(a) - a_l^j\right] r_j^l \\ &\leq \left[e_j^l(a) - a_l^j\right] r_j^l = a_j^l \\ &\leq \left[q_l^l - \sum_{p_k^l \geq p_j^l, k \neq j} q_k^l\right]^+, \end{aligned} \tag{12}$$

where the last line follows from (5). Now using (5) 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 (12), and the last is by definition. This proves (10). 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 (5). That along with the fact that $f^l \geq 0$ implies $u_j(z_j, p_j) - u_j \geq 0$. □