The Price of Simplicity*

Srinivas Shakkottai[†], R. Srikant[‡], Asuman Ozdaglar[§] and Daron Acemoglu[¶]

Abstract

We study revenue-maximizing pricing by a service provider in a communication network and compare revenues from simple pricing rules to the maximum revenues that are feasible. In particular, we focus on flat entry fees as the simplest pricing rule. We provide a lower bound for the ratio between the revenue from this pricing rule and maximum revenue, which we refer to as the Price of Simplicity. We characterize what types of environments lead to a low Price of Simplicity and show that in a range of environments, the loss of revenue from using simple entry fees is small. We then study the Price of Simplicity for a simple non-linear pricing (price discrimination) scheme based on the Paris Metro Pricing. The service provider creates different service classes and charges differential entry fees for these classes. We show that the gain from this type of price discrimination is small, particularly in environments in which the simple entry fee pricing leads to a low Price of Simplicity.

1 Introduction

With the considerable increase in the commercial use of the Internet, the issue of how Internet services and network resources should be priced has become an important topic of analysis both in economics and in engineering communities. The traditional model of the media, often adopted in analysis of the Internet and other two-sided markets, presumes

*This work was supported in part by the NSF grant CCF 06-34891

[†]Dept. of MS&E, Stanford University, sshakkot@stanford.edu

[‡]Dept. of ECE, UIUC, rsrikant@uiuc.edu

[§]Dept. of EECS, MIT, asuman@mit.edu

[¶]Dept. of Economics, MIT, daron@mit.edu

that content providers generate revenues through advertising or other sources related to viewership (e.g., Rochet and Tirole [1]). In addition, in the present context, the platforms, the Internet ISPs (Internet Service Providers), can be best approximated as charging fixed prices regardless of the value of the content that is being uploaded or downloaded. This "dumb pipe" model—with fixed pricing by carriers and no pricing by companies providing the services regardless of the content—is changing rapidly, however, as more content providers start charging customers for accessing their services and ISPs providing different qualities of service to different groups of users (see, for example, Odlyzko [2, 3]). The issue of different qualities of service (QoS) naturally raises the question of whether a scheme of non-linear pricing can be and will be implemented in communication networks in general and in the Internet in particular.

One drawback of differentiated pricing schemes in general is that the resulting pricing schemes tend to be highly nonlinear and not resemble the prices observed in practice. For example, in the context of an optimal auction (mechanism design) problem, Cremer and McLean [4] show that when the values of different buyers are correlated, a revenue-maximizing monopolist may be able to extract all of the consumer surplus. But the form of this auction (or the corresponding differentiated prices) are much more complicated than the pricing schemes we observe in auctions or in industries with price discrimination. Whether simple pricing schemes can be used profitably in the context of the Internet is important, since complicated schemes may be too costly to implement or may be prone to be manipulated. Odlyzko [3], for example, argues that such differentiated pricing is typically resisted by consumers, but also suggests that it will be eventually adopted if it is not too complicated. On this basis, we may expect greater use of revenue-maximizing behavior by ISPs and non-linear prices, which may improve the allocation of resources in communication networks, but only to the extent that there are profitable pricing schemes that are "simple".

In this paper, we study whether simple pricing schemes can approximate revenue-maximizing non-linear prices in communication networks. We focus on a stylized communication network in which the allocation of resources takes place in real time. In particular, at any given point in time, a service provider has a certain amount of network resources (e.g., bandwidth) to be allocated among a number of users. We assume that there are N user classes, each represented by a utility function designating their utility and willingness to pay for a given amount of the resource (bandwidth). Using this model, we first study the profitability of the simplest pricing scheme, which charges a simple entry fee to all potential users and then divides the resource equally among users who pay the entry fee. We then compare revenue from this simple entry fee scheme to the maximum revenue that the service provider could secure. In particular, we know that maximum revenue is upper bounded by the maximum value of the consumer surplus (which is typically a non-tight bound, but in certain problems such as those studied by Cremer and McLean [4], a complicated pricing scheme might reach this bound). Therefore our strategy is to compare the profitability of the simple entry fee pricing to the maximal value of consumer surplus. We call the ratio of these two objects the *Price of Simplicity* (PoS). When the PoS is small, simple pricing schemes are unlikely to be revenue-maximizing and thus the actual pricing in practice has to be determined by an explicit comparison of the profitability of different pricing schemes and their costs in terms of complication. In contrast, when PoS is close to one, a simple pricing scheme can approximate revenue-maximization. In this case, we may expect actual pricing to take a relatively simple form and emerge relatively quickly.

We consider the revenue of a monopolist supplier of communication bandwidth in a decentralized network. For this problem, we characterize an explicit lower bound on the PoS as a function of the number of user classes (each with different utility functions). Somewhat surprisingly, in many cases this lower bound is quite high, indicating that using simple pricing schemes may not be too costly for service providers in communication networks. The worst-case (lower bound) corresponds to a situation which the utility functions are piece-wise linear (with a steeply-increasing first portion followed by a flat portion). In this case, the PoS is still bounded away from zero for any finite number of differentiated user classes, but as the number of user classes tends to infinity the PoS tends to zero.

We next consider the simplest scheme of differentiated pricing, which, following Odlyzko's [5] work, is commonly referred to as the Paris Metro Pricing (PMP). In PMP a network is partitioned into several service classes, with each service class having a fixed fraction of the entire network capacity. The fractions handle traffic using the same protocols and give no formal QoS guarantee to users. In the context of our model, this means that the resource in each service class is distributed equally among the users that take part in that service class. Since different service classes will allocate different amount of resources to users, the entry fee differs across service classes, leading to a simple form of differentiated (non-linear) pricing scheme. According to this pricing scheme, service classes with higher prices will naturally have higher QoS to compensate (marginal) users for the higher prices that they are paying by offering them greater use of resources and greater utility. In the second part of the paper, we analyze the properties of a PMP scheme and characterize the PoS for this case. We first establish a single-crossing type result, showing that user classes with greater valuation of bandwidth are more likely to choose high-priced service classes than use the classes with lower valuation. Using this result, we then characterize the revenue-maximizing PMP scheme and show that for the piece-wise linear utility function, the PMP scheme generates a similar PoS as the simple entry fee pricing. Consequently, the gain from this particular differentiated pricing scheme is relatively limited. This result, combined with the relatively good performance of the simple access fee pricing for general concave utility functions, suggests that simple pricing rules might work quite well in modern communication networks. We view this result as encouraging for the development of efficiency-enhancing pricing schemes in the Internet and other communication networks, since it indicates that most service providers will not have an incentive for designing complicated pricing schemes.

Our paper is a first attempt at understanding whether simple pricing schemes can be used by revenue-maximizing service providers in communication networks. As such, our model is stylized and one can obtain sharper results by making more specific assumptions on utility functions or the number of user classes. We believe that this is an interesting area for future work.

While we are not aware of other studies investigating the price of simplicity (or other concepts related to the costs of using simple pricing schemes), there is a large related literature upon which we are building. First, as already mentioned above, there is a large literature in economics on optimal (revenue-maximizing) auctions and mechanisms. A classic paper by Myerson [6] characterized optimal auctions in situations where potential buyers have identically distributed private values. This corresponds to one of the simplest cases of non-linear pricing. Non-linear pricing in general is studied in Wilson [7], while the large literature on optimal auctions is surveyed in Krishna [8]. Particularly notable are the papers by Cremer and McLean [4], mentioned above, which show how full surplus extraction may be possible in certain situations. Second, the environment we analyze is closely related to that in Acemoglu *et al.* [9]. They consider entry pricing schemes in communication networks where the service provider also determines the resource allocation rule (for example, the allocation of power according to signal quality in wireless networks). They show that the revenue-maximizing policy for the service provider satisfies the marginal user principle, whereby the resource allocation rule is chosen to maximize the utility of the marginal user, that is, the user who is just indifferent between participating and not participating in the network. In the current context, this result is straightforward when there is no price discrimination, since we assume equal allocation of network resources among participants (thus there is no choice of resource allocation rule). When we turn to networks with different service classes (the PMP), we provide a simple generalization of this result.

Third, there is now a large literature investigating game-theoretic equilibria in communication networks (see the survey by Ozdaglar and Srikant [10]). One branch of this literature focuses on the strategic interactions among users [11, 12, 13]. This literature typically characterizes worst-case efficiency losses from strategic interactions, which is referred to as the Price of Anarchy (PoA) following the work by Papadimitriou [14]. Another branch integrates the strategic interactions between users and service providers as well as among users [15, 16, 17, 18, 19, 20, 21]. This work has not focused on price discrimination or non-linear pricing, though our framework is considerably simpler than some of the models studied in this literature, since it has a single service provider rather than competition among service providers.

The rest of the paper is organized as follows. In Section 2 we study the case of a single service class, characterize the PoS for a class of utility functions and determine a tight lower bound. We then consider a multi-class service structure in Section 3 and show that its worst case revenue efficiency is identical to that of the single service-class regime. Finally, we conclude in Section 4.

2 Single Service Class

Let us consider the case where there is a single class of service, and the service provider sets a single entry price. Let C denote the bandwidth of the link, and let p be the entry price. Let $\mathcal{N} = \{1, \ldots, N\}$ denote the set of potential users. We consider the case of ordered utilities, where the utility function of user i is given by

$$u_i(x) = \alpha_i u(x),$$

where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N > 0$. Such utility functions might be seen in practice when all users are running identical applications, but their valuations of the applications are different. This would mean that the shape of the QoS curve would be the same for all of them up to a scaling factor. Throughout the paper we adopt the following assumption on the utility function $u(x)$.

Assumption 1 *The utility function $u : [0, \infty) \to [0, \infty)$ is concave, nondecreasing, and satisfies $u(0) = 0$.*

We assume that the ISP divides C equally among the users entering the system. Thus, if $\tilde{N}$ users enter the system, then each user would obtain a bandwidth of $C/\tilde{N}$. Users enter the system only if the utility they gain from the bandwidth allocation x exceeds the entry price, i.e., $u_i(x) \geq p$. Under these conditions, we can characterize the revenue that would be obtained by the ISP using the following result.

Result 1 (Marginal user principle [9]) *For a given vector $\alpha = \{\alpha_1, \ldots, \alpha_N\}$ with $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$, the maximum revenue that the service provider can obtain by posting a single entry price, denoted by $\mathcal{R}_{MUP}(\alpha, u)$, is given by*

$$\mathcal{R}_{MUP}(\alpha, u) = \max_{i \in \mathcal{N}} \left\{ i \alpha_i u\Big(\frac{C}{i}\Big) \right\}.$$

This result is straightforward since we assume equal allocation of network resources among participants. It is a special case of the result in [9], where the authors were concerned with the determination of the allocation rule.

However, for a given vector α, the maximum revenue that a service provider can obtain by any selling mechanism is upper bounded by the entire consumer surplus that he can extract under complete information i.e., as the optimal solution of the following problem.

$$\begin{aligned} \text{maximize}_{x \geq 0} \quad & \sum_{i \in \mathcal{N}} \alpha_i u(x_i) \\ \text{subject to} \quad & \sum_{i \in \mathcal{N}} x_i \leq C. \end{aligned} \tag{1}$$

We denote the optimal value of the preceding problem by $\mathcal{R}_S(\alpha, u)$. Since the utility function u is concave by assumption, a vector x^S is an optimal solution of problem (1) if and only if there exists a scalar $\mu \geq 0$ such that $\mu\Big(\sum_{i\in\mathcal{N}} x_i^S - C\Big) = 0$ and

$$\alpha_i u'(x_i^S) \;=\; \mu, \quad \text{if } x_i^S > 0, \tag{2}$$

$$\geq\; \mu, \quad \text{if } x_i^S = 0. \tag{3}$$

Cremer and Mclean [4] show that under some assumptions on the correlation of values of customers, using an individual lottery for each customer can be used to extract the entire surplus. However, in this paper we are interested in the highest loss of revenue incurred by using the simple selling mechanism that posts a single entry price. We next define our revenue loss metric.

Definition 1 (Price of Simplicity) *Given a utility function u, we define the price of simplicity (PoS) as the worst-case ratio of the revenue under single posted price to the maximum revenue that the service provider can obtain under complete information; i.e.,*

$$PoS(u) = \inf_{\alpha\geq 0} \frac{\mathcal{R}_{MUP}(\alpha, u)}{\mathcal{R}_S(\alpha, u)}. \tag{4}$$

We can assume without loss of generality that the infimum in the preceding definition is taken over all α such that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$ and $\alpha_1 = 1$. We have the following simple result.

Lemma 1 *Let the utility function u satisfy Assumption 1. Let $\alpha$and $\tilde{\alpha}$ be two nonnegative vectors in $\mathbb{R}^N$ such that $\alpha \geq \tilde{\alpha}$ (i.e., $\alpha_i \geq \tilde{\alpha}_i$ for all $i \in \mathcal{N}$). Then,*

$$\mathcal{R}_S(\alpha, u) \geq \mathcal{R}_S(\tilde{\alpha}, u).$$

Proof. Let x^S and $\tilde{x}^S$ be an optimal solution of the problem in (1) for vectors α and $\tilde{\alpha}$, respectively. Since $\tilde{x}^S$ is a feasible solution for the problem in (1) with the scaling vector α, we have

$$\mathcal{R}_S(\alpha, u) = \sum_{j\in\mathcal{N}} \alpha_j u(x_j^S) \geq \sum_{j\in\mathcal{N}} \alpha_j u(\tilde{x}_j^S) \geq \sum_{j\in\mathcal{N}} \tilde{\alpha}_j u(\tilde{x}_j^S) = \mathcal{R}_S(\tilde{\alpha}, u),$$

where the second inequality follows by the assumption that u is a nonnegative function [since u is nondecreasing and $u(0) = 0$]. ■

Using the preceding monotonicity result, we can show that at the optimal solution of the problem defined in (4), all terms in the numerator of the objective function are equalized.

Proposition 1 *Let the utility function u satisfy Assumption 1. There exists an optimal solution $[1, \alpha_2^*, \dots, \alpha_N^*]$ of the problem defined in (4) such that*

$$u(C) = i\alpha_i^* u\Big(\frac{C}{i}\Big), \qquad i = 2, \dots, N. \tag{5}$$

Proof. Let $[1, \alpha_2^*, \dots, \alpha_N^*]$ be an optimal solution of the problem defined in (4). Assume that the relation given in (5) does not hold for all i. Let j be the smallest integer such that

$$j\alpha_j^* u\Big(\frac{C}{j}\Big) \neq u(C).$$

There are two cases to consider:
Case 1: Assume that $j\alpha_j^* u\Big(\frac{C}{j}\Big) > u(C)$. Let

$$K = \frac{j\alpha_j^* u\Big(\frac{C}{j}\Big)}{u(C)}.$$

Then, by Lemma 1, it follows that the vector $[K, \dots, K\alpha_{j-1}^*, \alpha_j^*, \dots, \alpha_N^*]$ is an optimal solution of the problem in (4), at which we have $j\alpha_j^* u\Big(\frac{C}{j}\Big) = u(C)$.
Case 2: Assume that $j\alpha_j^* u\Big(\frac{C}{j}\Big) < u(C)$. In view of the concavity of the utility function u, for all $i = 1, \dots, N$, we have

$$iu\Big(\frac{C}{i}\Big) \geq (i-1)u\Big(\frac{C}{i-1}\Big). \tag{6}$$

To see this, consider the linear function $l(x) = (i-1)u\Big(\frac{C}{i-1}\Big)x$. Since u is concave and $u(0) = 0$ by Assumption 1, $l(x)$ underestimates the function $Cu(x)$ in the interval $x \in \Big[0, \frac{C}{i-1}\Big]$, and the result follows. Let

$$K = \frac{(j-1)\alpha_{j-1}^* u\Big(\frac{C}{j-1}\Big)}{j\alpha_j^* u\Big(\frac{C}{j}\Big)}.$$

By (6), we have $K\alpha_j^* \leq \alpha_{j-1}^*$. Together with Lemma 1, this implies that the vector

$$[1, \dots, \alpha_{j-1}^*, K\alpha_j^*, \alpha_{j+1}^* \dots, \alpha_N^*]$$

is an optimal solution of the problem in (4), at which we have $j\alpha_j^* u\left(\frac{C}{j}\right) = u(C)$.

Repeating the preceding arguments recursively and normalizing at the end such that $\alpha_1^* = 1$, we construct an optimal solution of the problem in (4) with the desired characteristic, which completes the proof. ■

The following corollary follows naturally from the above characterization of the worst case vector α.

Corollary 1 *Let the utility function u satisfy Assumption 1. We can express the PoS (cf. Definition 1) as*

$$PoS(u) = \frac{u(C)}{\sum_{i \in \mathcal{N}} \frac{u(C)}{iu(C/i)} u(x_i^S)}, \tag{7}$$

where x^S is an optimal solution of the problem in (1) for $\alpha_i = \frac{u(C)}{iu(C/i)}$.

We next investigate the PoS for different kinds of utility functions. Our goal is to provide a tight lower bound on the PoS.

Worst-case utility function

We provided an explicit characterization of the loss of revenue when using a single entry price above. We next provide a lower bound on the revenue loss. We first provide a few examples using standard utility functions and then go on to lower bound. We assume without loss of generality that $C = 1$. **Example 1:** Consider the utility function

$$u(x) = \log(1 + x)$$

We can find an explicit expression for the PoS in this case using simple Lagrange multiplier techniques. Here the multiplying coefficients of the different classes (again assuming without loss of generality that $C = 1$), are of form $\frac{\log(2)}{i \ \log\left(1+\frac{1}{i}\right)}$. It is obvious that given these multiplying coefficients, the allocation maximizing the sum utility would be such that the slopes of the different classes are identical at the optimal allocation. Using this fact, we numerically find that the number of classes obtaining a nonzero allocation is 14. We then write out the expression for the PoS to be

$$\frac{\log(2)}{\sum_{i=1}^{14} \frac{\log(2)}{i \ \log\left(1+\frac{1}{i}\right)} \ \log\left(\frac{15 \ i \log\left(1+\frac{1}{i}\right)}{\sum_{j=1}^{14} \frac{\log(2)}{j \ \log\left(1+\frac{1}{j}\right)}}\right)}.$$

Evaluating this expression yields the value for the PoS to be 87.8012%.

Example 2: Consider the utility function

$$u(x) = 1 - e^{-x}$$

Using the same method as above, we evaluate the number of classes which get a nonzero allocation as 12. Again, we may write out a closed form expression, which we evaluate as 84.4756%.

The above utility functions can be said to represent so called "elastic-traffic" (such as data transfers), in which user utility is strictly concave-increasing in the allocated bandwidth. Many of the currently used data transfer protocols such as TCP, can be modeled using such utility functions. The examples seem that the to suggest that the PoS may not be too high. We provide a lower bound on the price of simplicity, which shows that for large N, the PoS can actually be arbitrarily bad.

Proposition 2 *Let the utility function u satisfy Assumption 1. Then for any N, we have*

$$PoS(u) \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{i}},$$

and the bound is tight (satisfied with equality) for the piece-wise linear utility function

$$u_{pwl}(x, N) \triangleq \begin{cases} Nx & \text{if } 0 \leq x \leq \frac{1}{N} \\ 1 & \text{if } x \geq \frac{1}{N}. \end{cases} \tag{8}$$

Proof. We first show the following relation.

$$\frac{u(x_i^S)}{u(1/i)} \leq 1 \qquad \forall\, i \in \mathcal{N}, \tag{9}$$

where x^S is an optimal solution of the problem in (1) for $\alpha_i = \frac{u(C)}{iu(C/i)}$. To see this, we show that $x_i^S \leq \frac{1}{i}$. Assume to arrive at a contradiction that $x_i^S > \frac{1}{i}$. Since $\alpha_i \geq \alpha_j$ for all $j \leq i$ [cf. Eq. (6], this implies by the structure of problem (1) that $x_j^S > \frac{1}{i}$ for all $j \leq i$. Therefore, we have $x_1^S + \cdots + x_i^S > 1$ contradicting the fact that $C = 1$. Hence, we have $x_i^S \leq \frac{1}{i}$. In view of the assumption that u is a nondecreasing function, this implies the relation in (9). Using relation (9) in Corollary 1, we have that the denominator in (7) is upper bounded as

$$\sum_{i=1}^{\tilde{N}} \frac{1}{iu(1/i)} u(x_i^S) \leq \sum_{i=1}^{\tilde{N}} \frac{1}{i} \leq \sum_{i=1}^{N} \frac{1}{i}, \tag{10}$$

establishing the relation

$$PoS(u) \geq \frac{1}{1 + \frac{1}{2} + \cdots + \frac{1}{N}}. \tag{11}$$

To establish tightness, consider the piece-wise utility function $u_{pwl}(x, N)$ defined in (8). We see from Corollary 1 that the vector $\alpha = [1, \frac{1}{2}, \cdots, \frac{1}{N}]$ is associated with the lowest revenue efficiency of $u_{pwl}(x, N)$. It is then obvious that for $u_{pwl}(x, N)$, the maximum value of the denominator in (7) occurs when we allocate $1/N$ to each of the N users. Thus, we can write out the expression for the PoS as

$$PoS(u_{pwl}) = \frac{1}{1 + \frac{1}{2} + \cdots + \frac{1}{N}}, \tag{12}$$

which is identical to the lower bound (11) calculated above. ■ The utility function $u_{pwl}(x, N)$ would be representative of real-time traffic, wherein the user utility increases with increasing bandwidth up to a point, after which excess bandwidth is irrelevant. Thus, the single-class scheme is not efficient for real-time traffic.

3 Generalized Metro Pricing

We next consider an extension of the pricing scheme described in the previous section and assume that the ISP divides the available bandwidth (or rate) into M parts or *classes* and charges a separate price p_j for each class. We refer to this pricing scheme as *generalized metro pricing* since it is reminiscent of Odlyzko's Paris Metro Pricing scheme [5]. Our goal in this section is twofold. First, we provide a model for generalized metro pricing when prices are set by a single revenue-maximizing provider. We provide an explicit characterization for revenue-maximizing prices and show that the optimal prices satisfy a generalized marginal user principle. Second, we investigate the worst-case loss of revenue for the case of ordered utility functions that satisfy Assumption 1 with the objective of understanding the effect of charging multiple prices on the revenue of the service provider.

Our model is as follows. The service provider divides the available bandwidth C into M classes with class set $\mathcal{M} = \{1, \ldots, M\}$ and assigns price p_j to class j. We assume without loss of generality that $C = 1$. There are N potential users with user set $\mathcal{N} = \{1, \ldots, N\}$.

Given the price vector set by the service provider $p = [p_j]_{j \in \mathcal{M}}$, we assume that the payoff function of user i that is assigned to class j is given by

$$\pi_i(x) = \alpha_i u(x) - p_j,$$

where $u(x)$ is a utility function that satisfies Assumption 1, and $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N > 0$.

For each class j, the service provider chooses a price p_j, the amount of bandwidth that would be available per user x_j, and the number of users to include n_j to maximize his revenue. As in Paris Metro Pricing, we assume that for $k \geq j$, $x_k \leq x_j$. Given the price vector $p = [p_j]_{j \in \mathcal{M}}$, the allocation vector $x = [x_j]_{j \in \mathcal{M}}$, and the vector of number of users admitted in each class $n = [n_j]_{j \in \mathcal{M}}$, the revenue of service provider is given by

$$\Pi(p, x, n) = \sum_{j=1}^{M} p_j n_j.$$

The service provider should set the prices, the allocation vector, and the number of users admitted in each class such that for user i designated for class j, the individual rationality constraints given by

$$\alpha_i u(x_j) - p_j \geq 0,$$

and the incentive compatibility constraints given by

$$\alpha_i u(x_j) - p_j \geq \alpha_i u(x_k) - p_k \qquad \text{for all } k \neq j,$$

are satisfied.

The following arguments use the ordered utility assumption and allow us to write down the incentive compatibility constraints in a compact form.

Lemma 2 *Let j and $\bar{j}$ be two classes such that $j \leq \bar{j}$ (i.e., $x_j \geq x_{\bar{j}}$) with $p_j \geq p_{\bar{j}}$. Let i and $\bar{i}$ be two users such that $\alpha_i \geq \alpha_{\bar{i}}$. Then we have the following results on when the users have no incentive to change their service classes.*
Satisfaction with better QoS: If $\alpha_{\bar{i}} u(x_j) - p_j \geq \alpha_{\bar{i}} u(x_{\bar{j}}) - p_{\bar{j}} \Rightarrow \alpha_i u(x_j) - p_j \geq \alpha_i u(x_{\bar{j}}) - p_{\bar{j}}$, i.e., users i and $\bar{i}$ would prefer better QoS class j to worse QoS class $\bar{j}$.
Satisfaction with worse QoS: If $\alpha_i u(x_{\bar{j}}) - p_{\bar{j}} \geq \alpha_i u(x_j) - p_j \Rightarrow \alpha_{\bar{i}} u(x_{\bar{j}}) - p_{\bar{j}} \geq \alpha_{\bar{i}} u(x_j) - p_j$, i.e., users i and $\bar{i}$ would prefer worse QoS class $\bar{j}$ to better QoS class j.

Proof. Suppose that $\alpha_{\bar{i}} u(x_j) - p_j \geq \alpha_{\bar{i}} u(x_{\bar{j}}) - p_{\bar{j}} \Rightarrow \alpha_{\bar{i}}(u(x_j) - u(x_{\bar{j}})) \geq p_j - p_{\bar{j}}$. Since $x_j \geq x_{\bar{j}}$ and $p_j \geq p_{\bar{j}}$ by assumption, both sides of the inequality are non-negative. Then the first part of the lemma follows since $\alpha_i \geq \alpha_{\bar{i}}$. The second part can be proved in an identical fashion. ■

Suppose that users $\mathcal{N}_j \subset \mathcal{N}$ are assigned to class $j \in \mathcal{M}$, with user $i \in \mathcal{N}$ being associated with α_i. Define $i_{\max}(j) = \arg\max_i \alpha_i, \; i \in \mathcal{N}_j$ and $i_{\min}(j) = \arg\min_i \alpha_i \; i \in \mathcal{N}_j$. Suppose that the set of users $\mathcal{N}'_j = \{i : i_{\min}(j) \leq i \leq i_{max}(j)\}$ were assigned to class j. Lemma 2 implies that if user $i_{max}(j)$ has no incentive to move to a better QoS class, and user $i_{\min}(j)$ has no incentive to move to a worse QoS class for all classes j, then if $\mathcal{N}_j = \mathcal{N}'_j$ for all j, none of the users assigned to any class would have an incentive to leave. On the other hand, if $\mathcal{N}_j \neq \mathcal{N}'_j$ for some j, then some user in some class would have an incentive to change. Thus, given the vector of number of users in each class $n = [n_j]_{j \in \mathcal{M}}$, the previous lemma implies that the following allocation of users to classes is a necessary condition for equilibrium.

Condition 1

$$\textit{Users } 1, \ldots, n_1 \rightarrow \; \textit{class } 1,$$

$$\textit{Users } n_1 + 1, \ldots, n_1 + n_2 \rightarrow \; \textit{class } 2,$$

$$\vdots$$

$$\textit{Users } n_1 + n_2 + \cdots + n_{M-1} + 1, \ldots, n_1 + n_2 + \cdots + n_M \rightarrow \; \textit{class } M.$$

This motivates us to define the following notation. For any user $i \in \mathcal{N}$, let j_i denote the class that user i is assigned to, i.e., j_i is an integer in $\mathcal{M}$ that satisfies $n_1 + \cdots + n_{j_i - 1} + 1 \leq i \leq n_1 + \cdots + n_{j_i}$. Note that $j_{n_1 + \ldots + n_j} = j$. Also note that by the preceding discussion for any user with index $i \leq q$, we must have $j_i \leq j_q$. We then have the following lemma.

Lemma 3 (Single Crossing Property) *If the assignment of users to classes satisfies Condition 1 and*

$$\alpha_i u(x_{j_i}) - p_{j_i} \;\; \geq \;\; \alpha_i u(x_{j_i+1}) - p_{j_i+1} \quad \forall \; i \in \mathcal{N}, \; j_i < M \textit{ with equality for } i = n_1 + \ldots + n_j,$$

then no user has an incentive to deviate from his assigned class to any other class.

Proof. *Case 1:* The condition for user i to not switch to a worse QoS level is $\alpha_i u(x_{j_i}) - p_{j_i} \geq \alpha_i u(x_k) - p_k$ for all $k \geq j_i$, which can be re-written as

$$\alpha_i(u(x_{j_i}) - u(x_k)) \geq p_{j_i} - p_k \quad \forall\ k \geq j_i$$

Now, since we know that by Condition 1 that for any user $i \leq q$ we have $\alpha_i \geq \alpha_q$ and class of service $j_i \leq j_q$. Also, $x_{j_i} \geq x_k$ for any $k \geq j_i$. Thus,

$$\begin{aligned}
\alpha_i(u(x_{j_i}) - u(x_k)) &= \alpha_i \sum_{l=j_i}^{k-1} u(x_l) - u(x_{l+1}) \\
&\geq \alpha_i(u(x_{j_i}) - u(x_{j_i+1})) + \sum_{l=j_i+1}^{k-1} \alpha_{q_l}(u(x_l) - u(x_{l+1})), \text{ where } q_l \in \mathcal{N}_l \\
&\geq \sum_{l=j_i}^{k-1} p_l - p_{l+1} = p_{j_i} - p_k,
\end{aligned}$$

which is the desired condition.
Case 2: For any user i the condition for not switching to a better QoS is $\alpha_i u(x_{j_i}) - p_{j_i} \geq a_i u(x_k) - p_k$ for all $k \leq j_i$, which can be re-written as

$$\alpha_i(u(x_k) - u(x_{j_i})) \leq p_k - p_{j_i} \quad \forall\ k \leq j_i$$

We follow a similar argument as before:

$$\begin{aligned}
\alpha_i(u(x_k) - u(x_{j_i})) &= \alpha_i \sum_{l=k+1}^{j_i} (u(x_{l-1}) - u(x_l)) \\
&\leq \sum_{l=k+1}^{j_i-1} \alpha_q(u(x_{l-1}) - u(x_l)) + \alpha_i(u(x_{j_i-1}) - u(x_{j_i})), \text{ where } q_l \in \mathcal{N}_l \\
&\leq \sum_{l=k+1}^{j_i} \alpha_{n_1+..+n_{l-1}}(u(x_{l-1}) - u(x_l)) \\
&= \sum_{l=k+1}^{j_i} p_{l-1} - p_l = p_k - p_{ji}
\end{aligned}$$

as desired. Note that we made use of the fact that $\alpha_{n_1+\ldots+n_j} u(x_j) - p_j = \alpha_{n_1+\ldots+n_j} u(x_{j+1}) - p_{j+1}$ for all j. ■

The next proposition provides an explicit characterization of the price of each class.

Proposition 3 *If the assignment of users to classes satisfies Condition 1 and*

$$p_j = \sum_{k=j}^{M} \alpha_{n_1+\ldots+n_k} \left(u(x_k) - u(x_{k+1})\right) \;\; \forall \; j \leq M,$$

where n_j is as defined in Condition 1, and by definition, $u(x_{M+1}) = 0, p_{M+1} = 0$, then no user has an incentive to deviate from his assigned class to any other class.

Proof. The proof follows directly from Lemma 3 as follows. We have for any user i

$$\begin{aligned} p_{j_i} &= \sum_{k=j_i}^{M} p_k - p_{k+1} \\ &= \sum_{k=j_i}^{M} \alpha_{n_1+\ldots+n_k} \left(u(x_k) - u(x_{k+1})\right), \end{aligned}$$

where the above follows from the equality condition of Lemma 3. Also, we can verify that p_{j_i} above also satisfies the inequality condition of Lemma 3, which yields the proof. ■
Note that the proposition automatically requires that $p_k \geq p_l$ for all service classes $k \leq l$. Coupled with our requirement that $x_k \geq x_l$ for all service classes $k \leq l$, this means that the condition is that a better QoS corresponds to a higher price, which is what we would intuitively expect. We can then write the revenue-maximization problem as

Result 2 (Generalized Marginal User Principle) *For a given vector $\alpha = \{\alpha_1, \ldots, \alpha_N\}$ with $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$, and the number of service classes $M \leq N$ the maximum revenue that the service provider can obtain by generalized metro pricing, denoted by $\mathcal{R}_{GMP}(\alpha, u)$, satisfies*

$$\begin{aligned} \text{maximize}_{x,n} \quad & \sum_{j=1}^{M} p_j n_j \\ \text{subject to} \quad & p_j = \sum_{k=j}^{M} \alpha_{n_1+\ldots+n_k} \left(u(x_k) - u(x_{k+1})\right) \;\; \forall \; j \leq M. \\ & \text{with } x_{M+1} = 0 \\ & \sum_{j=1}^{M} n_j \leq N, \quad \sum_{j=1}^{M} x_j n_j \leq 1, \end{aligned}$$

We refer to the above as the *generalized marginal user principle* since the price vector p depends only on the users $n_1, n_1 + n_2, n_1 + n_2 + n_3, .., n_1 + n_2 + ... + n_M$, i.e., on the users who lie at the boundary of each service class.

Worst Case Utility Function

In the preceding discussion, we characterized the maximum revenue that could be obtained using GMP. Let us define a revenue efficiency metric

$$\eta_{GMP}(\alpha, u, M) = \frac{\mathcal{R}_{GMP}(\alpha, u, M)}{\mathcal{R}_S(\alpha, u)}, \tag{13}$$

where $\mathcal{R}_S(\alpha, u)$ is as defined in Problem 1 and by definition $\inf_{\alpha>0} \eta_{GMP}(\alpha, u, 1) = PoS(u)$. We would like to know whether there is an increase in revenue efficiency due to the extra degrees of freedom that GMP provides.

We first consider a few examples of the revenue efficiency associated with the utility functions considered earlier. For simplicity, for a given utility function u, let us assume that the vector α satisfies (5), $M = 2$ and the capacity $C = 1$ is divided equally among the two service classes. Under these circumstances, we can show numerically that for

$$\begin{aligned} u(x) &= \log(1+x), \quad \eta_{GMP}(\alpha, \log(1+x), 2) = 95.2\% \\ u(x) &= 1 - e^{-x}, \qquad \eta_{GMP}(\alpha, 1 - e^{-x}, 2) = 94.8\%, \end{aligned}$$

which are both higher than the PoS values calculated for these utility functions in Examples 1 and 2. This suggests that perhaps GMP might be able to extract a greater surplus than a single service class for all utility functions.

We next consider the revenue obtained by charging multiple prices with the piece-wise linear utility function $u_{pwl}(x, N)$ defined in (8). We will show that in fact, the GMP scheme is identical to a single class scheme as far as revenue is concerned regardless of the number of service classes, and the maximum revenue is obtained by assigning all the capacity to the first service class. We have the following proposition.

Proposition 4 *Suppose that there are N potential users and $u(x)$ is piece-wise linear as defined in (8), with $\alpha = [1, \frac{1}{2}, ..., \frac{1}{N}]$ (satisfies (5)). The revenue efficiency obtained using GMP with any number of service classes $M \leq N$ is*

$$\eta_{GMP}(\alpha, u_{pwl}, M) \leq \frac{1}{1 + \frac{1}{2} + \cdots + \frac{1}{N}},$$

with equality for $M = 1$.

Proof. The total revenue obtained using GMP may be determined from Result 2 to be

$$\begin{aligned}\mathcal{R}_{GMP}(\alpha, u_{pwl}) &= n_1^*\alpha_{n_1^*}(u(x_1^*) - u(x_2^*)) \\ &+(n_1^* + n_2^*)\alpha_{n_1^*+n_2^*}(u(x_2^*) - u(x_3^*)) \\ &+... \\ &+(n_1^* + n_2^* + ... + n_{M-1}^*)(\alpha_{n_1^*+n_2^*+...+n_{M-1}^*})(u(x_{M-1}^*) - u(x_M^*)) \\ &+(n_1^* + n_2^* + ... + n_M^*)(\alpha_{n_1^*+n_2^*+...+n_M^*})(u(x_M^*) - 0)\end{aligned}$$

Here the * refers to the maximizing solution. Note that $x_k^* = 0 \Rightarrow x_l^* = 0$ for all $l \geq k$ (by assumption) and $n_k^* = 0 \Rightarrow n_l^* = 0$ for all $l \geq k$. Since $\alpha_i = \frac{1}{i}$ for $i \leq N$, we obtain

$$\mathcal{R}_{GMP}(\alpha^*, u_{pwl}, M) = u(x_1^*)$$

Using the fact that $x_1^* \leq 1$, and (following the same argument as Proposition 2) $\mathcal{R}_S(\alpha, u_{pwl}) = 1 + \frac{1}{2} + \cdots + \frac{1}{N}$ we obtain the proof. ■

We conclude that in the worst case situation studied, GMP has no effect on the revenue obtained, and an ISP would have little incentive to implement such a scheme if user utilities resemble the piece-wise linear functions that we constructed.

4 Conclusions

In this paper we presented results on price-based discrimination for bandwidth allocation in wire-line communication networks. In general, the problem of mechanism design for resource allocation is very complex, and our focus was on studying simple mechanisms that show promise of widespread adoption in the arena of Internet pricing. Our objective was to study the revenue efficiency of single and multi-class pricing schemes as compared to the maximum possible revenue, and to the best of our knowledge, our work is the first attempt to provide a rigorous analysis of such pricing. Our assumption was that all the potential users would be running the same application, with each user valuing the application differently. This causes all the user utilities to have the same shape up to a scaling factor, i.e., the utility functions are of the form $u_i(x) = \alpha_i u(x)$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N > 0$. We further assumed that the utility functions are concave and increasing in the amount of bandwidth allocated. Finally, we assumed that users are individually rational, and would not join the system unless they had non-negative payoffs.

We first studied the case of a single service class with a fixed entrance fee. We showed that the worst-case ratio of revenues occurs when the set $\{\alpha_i\}$ is such that the ISP obtains the same revenue, regardless of the number of users who enter the system. The relation converts the determination of the worst case from a functional optimization to a that of a parameter optimization. We showed that for some typical utility functions representing elastic traffic (such as data transfers), the revenue efficiency is of the order of $85 - 90\%$. Hence, the current simple flat-rate pricing adopted by many ISPs is quite efficient in extracting revenue for elastic traffic. We also determined a tight lower-bound on the efficiency, which depends only on the number of served users. We showed that the utility function for which our lower bound is tight, is piece-wise linear with an initial increasing phase followed by a horizontal phase, and could result in an arbitrarily low revenue efficiency. Such a utility function would be representative of real-time traffic such as video streaming.

Motivated by the example of public transportation systems where two or more service classes are common, we studied the case of a multi-class pricing and service scheme. Here, the available bandwidth would be divided into chunks, and the entry fee for each chunk would be determined in advance. We showed how the incentive compatibility constraints of the users results in a *generalized marginal user principle*, in which only the marginal user in each class is instrumental in setting the price. Using this characterization of prices, we showed that even the simple artifice of dividing the available capacity into two parts results in an increased revenue for elastic traffic. However, we also proved the surprising result that regardless of the number of service classes, the revenue efficiency when the users have piece-wise linear utility functions is identical to that of a single class scheme.

In the future we would like to extend our work in several directions. We would like to know for what class of utility functions the PoS is bounded. This would give us an insight into what the worst case would be for particular applications. Secondly, our focus in this paper was on all users running the same application. We would like to study the case of users running different applications, which would result in multiple classes of users with utility functions of different shapes. We would like to answer the question of whether such multiple classes of utility functions should correspond to multiple classes of service for efficient revenue extraction.

References

[1] J.-C. Rochet and J. Tirole, "Platform Competition in Two-Sided Markets," *Journal of. the European Economic Association*, vol. 1, no. 4, pp. 990–1029, 2005.

[2] A. M. Odlyzko, "Pricing and architecture of the Internet: Historical perspectives from telecommunications and transportation," March 2004, available at: http://www.dtc.umn.edu/∼odlyzko/doc/networks.html.

[3] ——, "The evolution of price discrimination in transportation and its implications for the Internet," *Review of Network Economics*, vol. 3, no. 3, pp. 323–346, September 2004.

[4] J. Cremer and R. P. McLean, "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, vol. 56, no. 6, pp. 1247–1258, 1988.

[5] A. M. Odlyzko, "Paris metro pricing for the Internet," in *Proceedings of the ACM Conference on Electronic Commerce (EC'99)*, 1999, pp. 140–147.

[6] R. B. Myerson, "Optimal Auction Design," *Mathematics of Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.

[7] R. B. Wilson, *Nonlinear Pricing.* Oxford University Press, Oxford, UK, 1997.

[8] V. Krishna, *Auction Theory.* Academic Press, MA, U.S.A, 1997.

[9] D. Acemoglu, A. Ozdaglar, and R. Srikant, "The marginal user principle for resource allocation in wireless networks," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[10] A. Ozdaglar and R. Srikant, "Incentives and pricing in communication networks," in *Algorithmic Game Theory,* N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors. Cambridge University Press, Cambridge, UK, 2007.

[11] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, Trier, Germany, March 1999.

[12] T. Roughgarden and E. Tardos, "How bad is selfish routing?" in *IEEE Symposium on Foundations of Computer Science*, 2000, pp. 93–102.

[13] T. Roughgarden, "The price of anarchy is independent of the network topology," in *Proceedings of the 34th ACM Symposium on the Theory of Computing*, 2002, pp. 428–437.

[14] C. Papadimitriou, "Algorithms, games, and the Internet," in *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (STOC)*, Heraklion, Crete, Greece, July 2001.

[15] J. K. McKie-Mason and H. R. Varian, "Internet interconnection and the off-net-cost pricing principle," *RAND Journal of Economics*, vol. 34, no. 2, pp. 370–390, Summer 2003.

[16] R. E. Azouzi, E. Altman, , and L. Wynter, "Telecommunications network equilibrium with price and quality-of-service characteristics," in *Proceedings of the International Teletraffic Congress*, Berlin, September 2003.

[17] L. He and J. Walrand, "Dynamic provisioning of service level agreements between interconnected networks," in *Proceedings of the Allerton Conference on Control, Communications and Computing*, 2002.

[18] B. Hajek and G. Gopal, "Do greedy autonomous systems make for a sensible Internet?" in *Conference on Stochastic Networks*, 2002.

[19] G. Shrimali and S. Kumar, "Bill-and-keep peering," in *CoCombine Conference on Telecommunications Policy*, Cambridge, England, December. 2005.

[20] S. Shakkottai and R. Srikant, "Economics of network pricing with multiple ISPs," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1233–1245, December 2006.

[21] D. Acemoglu and A. Ozdaglar, "Competition and Efficiency in Congested Markets," *Mathematics of Operations Research*, vol. 32, no. 1, pp. 1–31, 2007.