Pricing Data Services: by Hours, by Gigabytes, or by Mega Bytes per Second?

Ying-Ju Chen* Ke-Wei Huang†

September 15, 2009

1 Introduction

Internet is inarguably the most important innovation of mankind in the past decade. Since 2000, the explosion of data travelling on the Web has brought in tremendous revenue for Internet Service Providers (ISPs) and telecommunication companies. Recently, the same trend has been propagated to the cell phone industry. While the percentage of cellular carriers' revenue from data plans remains relatively low (less than 20%, even including lucrative text messaging services), the expected year-over-year growth rate is high. Primary drivers for this trend include higher data throughput around 1 Mbps, availability in most major metropolitan areas, multiple device options (smartphones, PC Card modems, and embedded options for laptops), and an increasing selection of mobile applications and middleware. Typical industry analysts see no roadblocks of the long-term adoption of cellular broadband, particularly after iPhone 3G has successfully made its consumers addicted to mobility of computing.

How can a data services seller capitalize on such a rising trend? Anecdotal evidence from various ISPs and cell phone companies has clearly provided the answer: pricing is the key. A simple approach to find out the optimal pricing strategy is to learn from the history of ISP pricing. However, as we look back, the historical pricing plans exhibit substantially different features, particularly in the *pricing variables* (the basis or units for price discrimination). Historically, dial-up by 56k modems and ISDN ISP services were charged based on per-hour pricing. Nowadays, most of the residential ADSL services or cable-modem based ISPs charge consumers based on download/upload speeds. As we move towards the era of mobile commerce, leading mobile companies offer pricing plans that are based on the total data usage limit. Such a discrepancy is also pervasive across countries. While most of the mobile broadband service providers charge consumers by gigabytes (GB), pricing plans based on

*University of California, 4121 Etcheverry Hall, Berkeley, CA 94720; e-mail: chen@ieor.berkeley.edu.
†National University of Singapore, Singapore, Singapore, 117541; e-mail: huangkw@comp.nus.edu.sg.

download/upload speeds (e.g., in Singapore and Malaysia) or time usage (e.g., optional per-hour pricing in Taiwan and optional per-day pricing in New Zealand) are also adopted contemporarily.

Recent evidence strengthens our conjecture that practitioners may be unable to choose pricing variables for data services scientifically, as several dominant fixed-line residential ISPs realized that the prevalent per-Mbps pricing may not be profit-maximizing. For example, in 2008, Time Warner experimented in Texas a trial of "consumption-based billing", which is essentially a per-GB pricing, in order to boost the revenue by "avoiding the consequences of unfairness pricing," which could be an political language for profit maximization. Bell Canada, which meters services in some plans, charges consumers who go over the limit $7.50 per additional gigabyte. Comcast set a 250GB monthly cap on its residential Internet service in 2009.

This paper attempts to investigate the selection of pricing variables and the corresponding ISP pricing plans. In pursuit of this goal, we construct a stylized model in which a monopoly seller intends to provide the data services to heterogeneous consumers. Our primary goal is to provide a unified framework that allows us to investigate the sophisticated strategic considerations of data services pricing in the presence of consumer heterogeneity and supply restrictions. The remainder of this paper is organized as follows. In Section 2, we introduce the model. Section 3 summarizes our main results, and Section 4 discusses possible extensions and future directions.

2 The model

Our model setup is built upon the standard monopoly nonlinear pricing model in which a seller intends to provide the data services to heterogeneous consumers ([3]). Consumers are heterogeneous with regard to their willingness to pay for the total usage of the service, enjoy using the service at a higher connection speed, and incur utility loss while spending more time in using the services. Specifically, we adopt the following functional form for the consumer's utility:

$$u(B, M, \theta) = \begin{cases} B \times M \times \theta - \frac{1}{2}\left(B \times M\right)^2 - \alpha M + \beta B, & \text{if } B \times M \leq \theta \\ \frac{1}{2}\theta^2 - \alpha M + \beta B, & \text{if } B \times M > \theta \end{cases}, \tag{1}$$

where θ indicates the consumer's willingness to pay per unit data usage, B is the connection speed (i.e., Mbps), M is the usage in terms of unit time (i.e., minute, hour, or day), and $B \times M$ therefore represents the total usage in terms of gigabytes (GB) in this model. In (1), the term αM captures the disutility of spending more time in using the service, and βB represents the additional utility that a consumer obtains from having a higher connection speed, where α and β are known constants that are common among consumers.

We assume that the functional form (1) and the coefficients α and β are common knowledge.

However, the seller is unable to observe the consumer's willingness to pay (θ). Thus, θ also represents the consumer's "type." From the seller's perspective, θ is a random variable, and for simplicity we assume that it is uniformly distributed over $[0, 1]$. As in the majority of the literature on information goods/services, the variable cost of the monopoly ISP is negligible and therefore we normalize it to zero. Moreover, we assume that there is an upper bound for the connection speed, K, that reflects the seller's technological limitation.

As the seller is assumed to be risk neutral, his goal is to maximize his expected payoff while confronted with the heterogeneous consumers. Since the consumers are otherwise homogeneous except with regard to their willingness to pay that is unobservable to the seller, the seller can only adopt the indirect (second-degree) price discrimination scheme. Given that the consumers' utility depends on the usage and the connection speed, there are three options for the seller to conduct the second-degree price discrimination: by hours (M), by Mbps (B), and by GB ($B \times M$). This is the main departure from the existing literature on nonlinear pricing/ digital goods pricing which typically takes the pricing variables as exogenously given (see, e.g., [1], [2], and [3]). After the seller has chosen the pricing variables, the remaining problem is to design an appropriate pricing plan to maximize his expected profit. That is, the seller needs to design the pricing plans $P_M(M)$, $P_B(B)$, and $P(B \times M)$, respectively, in different scenarios.

From the above discussions, our model requires *two-stage* strategic decision making for the seller: the selection of pricing variables and the design of the pricing plan. In general, the (second-stage) problem of designing the pricing plan is fairly complicated since it requires the seller to characterize the entire function. In this paper, we adopt the mechanism design approach following the literature on the agency theory and nonlinear pricing. Specifically, we invoke the revelation principle to replace the pricing plan by a menu of contracts, and under this menu of contracts the seller can simply request the consumers to report their types and choose the contract accordingly on their behalves. For example, if the seller chooses to charge by hours, the pricing plan $P_M(M)$ can be transformed into a menu $\{(M(\theta), P_M(\theta))\}$, where $M(\theta)$ specifies the minutes used by the type-θ consumer and $P_M(\theta)$ is the corresponding price. Likewise, we can also transform the pricing plans $P_B(B)$ and $P(B \times M)$ into $\{(B(\theta), P_B(\theta))\}$ and $\{([B \times M](\theta), P(\theta))\}$, respectively. We normalize the reservation utilities of the seller and the consumers to zero if no transaction takes place.

3 Summary of our results

In this section, we characterize the seller's optimal pricing scheme. The sequence of events proceeds as follows. *Stage 1:* The seller first chooses one pricing variable (time usage, connection speed, or data usage). *Stage 2:* Given the pricing variable, the seller announces the menu of pricing contracts. *Stage 3:* Each consumer chooses from the menu. *Stage 4:* Each consumer determines the variables that are not specified in the pricing menu. For example, given a pricing plan by time usage, each consumer can self-select a bandwidth (which directly affects her total data usage) to maximize her utility.

In line with the standard procedure to solve this type of principal-agent problem, by backward induction, we first take the pricing variable as given and derive the optimal pricing plans under different scenarios (Stage 2's solutions). Following this, we compare the seller's maximum profits and determine the optimal pricing variables in Stage 1. Below, we omit the detailed formulations of the seller's optimization problems under different scenarios and simply report our main results.

First, we observe that under the pricing by Mbps, the seller should simply offer a single contract, which can be explained as follows. Recall that when the seller charges consumers by Mbps, they are still allowed to select their time usage on top of the connection speed choice. Thus, since the consumers can self-select time usage and change their data usage accordingly, the major function – the price discrimination – embedded in the nonlinear pricing scheme becomes ineffective. On the other hand, the cannibalization problem (that the seller must avoid the consumers from choosing contracts intended for other types) is so severe that renders the versioning strategy to be suboptimal. Our model therefore suggests that, the consumers' self-selection behaviors may completely eliminate the seller's price discrimination incentive, even though he retains the monopoly power in the market.

Second, we show that pricing by gigabytes is generally equivalent to pricing by hours. This is because in both cases, the consumers tend to use the maximum connection speed to maximize their utility from the services. As a result, there exists a one-to-one mapping from time spent on Internet and total gigabytes downloaded. This is reminiscent to the general insight regarding the equivalence between "watching the stock" and "watching the clock." As long as the connection speed (flow) is fixed, it does not matter whether we measure the aggregate usage or simply the time passed by.

Furthermore, our results imply that pricing by connection speed tends to be inferior to the other two options. To understand this result, recall that the consumers still have the discretion of selecting the time usage under the pricing plan (by connection speed). Thus, this provides too much leeway for the consumers to self-select. Consequently, this pricing plan is relatively ineffective in screening among heterogeneous consumers, thereby leaving too much surplus for the consumers that ultimately hurt the seller's revenue. On the contrary, if the seller prices consumers by hours or by gigabytes, every consumer self-selects the maximum connection speed $B = K$; as it is rationally expected by the seller, the consumers' self-selection behaviors are inconsequential under the other two pricing plans. Our result suggests that, from purely the demand-side consideration, residential ADSL or cable-modem-based ISPs may find it profitable to move from pricing by Mbps to other alternatives. This seems to coincide with the strategic concerns raised by Time Warner and Comcast.

4 Discussions and extensions

In this section, we consider some variants of our model characteristics to evaluate the robustness of our results.

4.1 Uncertain usage

In our model, we assume that the consumers can perfectly predict how much they will use the services. In some scenarios, it is possible that while signing the contract, the consumers face future uncertainties and consequently ex post usages may be different from their ex ante estimation. This effect can be conveniently incorporated into our framework by adding one noise term in the consumer's total usage. Specifically, in the presence of uncertainty, we can adopt the following alternative utility function:

$$u(B, M, \theta, \epsilon) = \begin{cases} [B \times (M+\epsilon) \times \theta - \frac{1}{2}[B \times (M+\epsilon)]^2 - \alpha(M+\epsilon) + \beta B, & \text{if } B \times M \leq \theta \\ \frac{1}{2}\theta^2 - \alpha(M+\epsilon) + \beta B, & \text{if } B \times M > \theta \end{cases},$$

where ϵ is a random variable with zero mean that captures the unpredictable usage after the contracting stage. In such a scenario, as the consumer is risk neutral, we can remove the additional constant term $-\frac{1}{2}\epsilon^2$ and conclude that all our results can be applied directly.

4.2 General Utility Function and Distribution of Types

Let us now relax the assumptions regarding the consumer utility and the type distribution. Specifically, we adopt a more general setting in which the consumers' utility takes the general form $U(B, M, \theta)$ and the type follows a general density function $f(\theta)$.

Equivalence of pricing by time usage and pricing by data usage. We first evaluate the robustness of the equivalence between pricing by hours and pricing by gigabytes. With this generalized model characteristics, we are able to establish the strong equivalence between the two pricing plans as long as $\frac{\partial U(B,M,\theta)}{\partial B} \geq 0$ holds. In other words, as long as each consumer unambiguously prefers higher speed than lower one, there is a strong equivalence between pricing by hours and pricing by gigabytes in terms of the induced consumer behavior and the resulting profitability for the seller.

Dominance of pricing by data usage over pricing by connection speed. Moreover, we are able to characterize the sufficient condition under which per-Mbps pricing is less profitable than per-GB pricing. If we denote the optimized utility function at Stage 4 as $U^*(B,\theta) = U(B(\theta), M^*(B,\theta), \theta)$, we can conveniently represent each consumer's utility as a function of two arguments: B and θ (after the time usage selection). Given this definition, we find that if $\frac{\partial^2 U^*(B,\theta)}{\partial B^2} > 0$ and $\frac{\partial^2 U^*(B,\theta)}{\partial B^2 \partial \theta} \leq 0$, pricing by Mbps yields a lower profit for the seller than the other two pricing plans.

4.3 The seller's bandwidth constraint

In our basic model, we focus primarily on the demand-side concern, leaving several supply side issues unaddressed. A notable issue is the presence of bandwidth capacity constraint in some practical situations. While our results apply to the cases in which data services sellers are endowed with ample capacity (such as fixed-line ADSL or Cable modems or WiMax ISPs), these capacity constraints are certainly crucial for other data services sellers (e.g., dialup or 3G mobile broadband). Taking into account such a bandwidth constraint, we are able to establish that 1) per-GB pricing is equivalent to per-Minutes pricing, and 2) it is possible that pricing by Mbps may outperform the other two pricing plans.

4.4 The seller's operating cost

In our model, we abstract away the cost of operating the ISP to obtain a simple closed-form expression. In reality, however, the seller must incur some operating cost while providing the services. In such a scenario, we can add some associated cost in the seller's objective. Given the structure of our problem, it can be verified that all the results remain unchanged if the seller incurs a constant marginal cost, as it simply gives rise to an additional markup for the prices. Naturally, the seller may also discard some more consumers that have relatively low willingness to pay due to the cost benefit analysis. However, the structural properties of the optimal pricing plans are unaffected by this modification. The case with general cost structure is much more involved, and it is conceivable that the optimal pricing variables and the corresponding pricing plan crucially depend on the cost structure. It therefore requires a thorough investigation to better understand the interaction between the supply side (the operations) and the marketing side (the pricing scheme).

References

[1] LM Hitt and PY Chen. Bundling with customer self-selection: A simple approach to bundling low-marginal-cost goods. *Management Science*, 51(10):1481–1493, OCT 2005.

[2] S Jain and PK Kannan. Pricing of information products on online servers: Issues, models, and analysis. *Management Science*, 48(9):1123–1142, SEP 2002.

[3] A Sundararajan. Nonlinear pricing of information goods. *Management Science*, 50(12):1660–1673, DEC 2004.