Time-Dependent Broadband Pricing

Carlee Joe-Wong
Department of Mathematics
Princeton University
Email: cjoe@princeton.edu

Sangtae Ha
Department of Electrical Engineering
Princeton University
Email: sangtaeh@princeton.edu

Mung Chiang
Department of Electrical Engineering
Princeton University
Email: chiangm@princeton.edu

***Abstract*—Charging different prices for Internet access at different times induces users to spread out their capacity, or bandwidth,g consumption across times of the day. Is it feasible and how much benefit can it bring? We design an architecture for time-dependent broadband pricing. We develop an efficient way to compute the cost-minimizing time-dependent prices for an Internet service provider (ISP), using both a static session-level model and a dynamic session model with stochastic arrivals. Our formulations of the optimization problem remain computationally tractable on a large scale. We show simulations illustrating the use and limitations of time-dependent pricing, and discuss next steps in implementing and deploying the theory in the TUBE Project.**

I. Introduction

A. Motivation and Related Work

Internet service providers (ISPs) practicing flat rate pricing face a dilemma: unlike its cost, an ISP's revenue does not scale with users' demand for more bandwidth. Usage-based pricing has been adopted by ISPs outside the United States and entered the U.S. wireless market last year (e.g. [1], [2]), driven by the tremendous growth of both wireline and wireless network traffic. This growth is out-pacing the increase of capacity and turning ISPs' attention to pricing as the ultimate tool to regulate demand. Yet pricing based just on monthly bandwidth usage leaves a timescale mismatch: monthly usage drives an ISP's revenue, but peak-hour congestion dominates its cost structure. Ideally, ISPs would like bandwidth consumption to be spread evenly over all the hours of the day.

Time-dependent usage pricing (TDP) charges a user based on both "how much" bandwidth is consumed and "when" it is consumed, as opposed to **time-independent usage pricing** (TIP), which only considers monthly consumption amounts. TDP has the potential to even out time-of-the-day fluctuations in bandwidth consumption [3]. Since it does not differentiate based on traffic type, protocol, or user class, TDP also sits lower on the "radar screen" of network neutrality scrutiny. In fact, wireless operators have long used a simple, 2 period TDP scheme, and small ISPs in the U.S. have begun experimenting with TDP. However, in their current implementation, users do not react to the time-dependent prices, and the prices are not optimized accordingly.

An effective TDP system must leverage different demographics' and applications' time-sensitivities into setting the right prices, which in turn shifts part of the demand across times of the day. A TDP system requires the integration of traffic measurement, optimal price determination, and user interface design in a real-time system. This paper addresses these questions with TUBE (Time-dependent Usage-based Broadband-price Engineering), an end-to-end TDP system for ISPs. Figure 1 summarizes TUBE as a control loop.

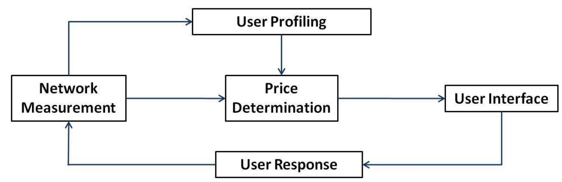

Fig. 1. Overall schematic of time-dependent pricing systems. We first discuss price determination and later explore user profiling, measurement, user interface and system integration.

The electricity industry has explored TDP over the years, e.g. [4]–[19]. There are several key differences from our pricing study. We extend these economic analyses to broadband pricing by incorporating a control-loop model that allows ISPs to adapt prices to user reactions in real time.[1] We also model TDP as users deferring part of their Internet usage, rather than the electricity market's model of users choosing the period in which to demand a resource. Previous broadband TDP models use "representative demand functions" to estimate peak and off-peak resource demands, while our models directly incorporate sessions' time-sensitivity and uncertainty in user behavior. We use n (e.g. $n = 48$ for half hour granularity) periods instead of 2; the multiple peaks and valleys in bandwidth usage over one day make 2 period TDP inadequate, although n periods complicate the model design.

This paper's formulation and methodology apply to both wireline and wireless pricing. In the U.S., wireless TDP will likely take off first, given its $10/GB usage price today, which is about 10 times wireline usage pricing.

B. Overview of Models and Summary of Results

When determining optimal prices, an ISP tries to balance the cost of demand exceeding capacity (e.g. the capital expenditure of capacity expansion) with the cost of offering reduced prices to users willing to move some of their sessions to later times. A user is a set of application sessions, each with a waiting

[1]Many prior works on TDP for electricity do not model real-time user reaction due to the lack of a convenient graphic user interface (GUI) and the relatively low elasticity of electricity usage. In contrast, broadband TDP can readily position GUIs on Internet access devices, and the elasticity of bandwidth consumption tends to be high for a good range of applications.

TABLE I
A SUMMARY OF THE MAIN NOTATION.

Symbol	Meaning: Static Model	Meaning: Dynamic Model
p_i	Reward for deferring to period i	Same
x_i	Usage in period i	Same
A	Maximum capacity	n/a
$f(x)$	Cost function, floored at 0	Same
X_i	Period i usage with TIP	Same
$w(p,t)$	Waiting function	Same
v_j	Volume of session j	n/a
$j \in i$	Sessions j originally in period i	n/a
$i-k$	$i-k$ mod n	Same
$w_\beta(p,t)$	The function $\frac{p}{(t+1)^\beta}$	n/a

function giving the probability or willingness to defer that session for some amount of time and some pricing incentive. Waiting functions may also represent an aggregate of users' willingnesses to wait over concurrent sessions. Pictorially, an ISP uses TDP to even out the peaks and valleys in bandwidth consumption over the day. The ISP's problem is to set its prices to balance these two types of costs, given its estimates of users' willingness to defer sessions at different prices.

The ISP's decision can equivalently be formulated in terms of rewards, as in our formulation. The ISP rewards users for deferring by the difference between TIP and optimal TDP prices. Without loss of generality, rewards are positive; their values reflect movement of the baseline usage price.

Section II develops the static model, which does not include stochastic arrival of new sessions. We prove that under certain reasonable conditions, price determination is computationally tractable. Section III extends to dynamic models with stochastic arrivals, which reduces to the static model for a single bottleneck network and demand under TIP equal to the amount of traffic arriving in each period.

While the waiting functions depend on the amount of time deferred, in our design all sessions in a given period are charged the same price. The ISP uses waiting function estimation to statistically model users' deferral behavior. Section IV shows sample waiting functions, illustrating the variation in time-sensitivities, and presents a waiting function estimation algorithm using aggregate TIP and TDP usage data.

Throughout this paper, we assume that ISPs are monopolies, facing an estimated distribution of users' waiting functions. Each session takes up a fixed amount of ISP capacity, e.g., the average over its short time-scale fluctuations, and TDP does not cause application sessions to disappear.

II. STATIC SESSION MODEL AND FORMULATION

The ISP's objective is to minimize the weighted sum of the cost of exceeding capacity and of offering reduced prices (i.e., rewards). The optimization variables are these rewards, which give users incentives to defer bandwidth consumption. Let X_i denote period i demand under TIP. The phrase "originally in period i" means that with TIP, this session occurs in period i.

Suppose that the ISP divides the day into n periods, and that its network has a single bottleneck link of capacity A. This link is often the aggregation link out of the access network, which has limited bandwidth compared to aggregate demand and is often oversubscribed by a factor of five or more. The cost of exceeding capacity in each period i, capturing both customer complaints and expenses for capacity expansion, is denoted by $f(x_i - A)$, where x_i is usage in period i. Capital expenditure cost is incurred over a large timescale; the f cost function represents the fraction due to daily capacity exhaustion.

Each period i runs from time $i-1$ to i. A typical period lasts a half hour. Sessions begin at the start of the period, an assumption readily modified to a distribution of starting times. The time between periods i and k is given by $i-k$, which is the number $b \in [1,n]$, $b \equiv i-k$ (mod n). If $k > i$, $i-k$ is the time between period k on one day and period i on the next.

For each session j originally in period i, define the **waiting function** $w_j(p,t) : \mathbb{R}^2 \to \mathbb{R}$, which measures the user's willingness to wait t amount of time, given reward p. Each session j has bandwidth requirement v_j, so $v_j w_j(p,t)$ is the amount of session j deferred by time t with reward p. To ensure that $w_j \in [0,1]$ and that the calculated usage deferred out of a period is not greater than demand under TIP, we normalize the w_j, dividing by the sum over possible times deferred t of $w_j(P,t)$. Here P is the maximum possible reward offered, or maximum marginal cost of exceeding capacity.

Proposition 1: The ISP's optimization problem for time-varying rewards can be formulated as

$$\mathbf{min} \sum_{i=1}^{n} p_i \left(\sum_{k=1,k\neq i}^{n} \sum_{j\in k} v_j w_j(p_i, i-k) \right) + f(x_i - A_i) \tag{1}$$

$$\mathbf{s.\,t.}\; x_i = X_i - \sum_{j\in i} v_j \sum_{k=1,k\neq i}^{n} w_j(p_k, k-i) + \sum_{k=1,k\neq i}^{n} \sum_{j\in k} v_j w_j(p_i, i-k), \tag{2}$$

$$\mathbf{var}.\; p_i;\, i = 1,\ldots,n.$$

Proof: See [20]. The key step uses the waiting function normalization to track aggregate usage deferred from and into each period. ■

In usage-based pricing, whether TDP or TIP, the ISP may charge a flat rate until users reach a certain cap, and after that charge a usage-based rate. Explicitly modeling this cap in TDP considerably complicates problem tractability, so we instead vary available capacity with time. In each period, the ISP subtracts from the network capacity usage from those users not reaching the cap, as well as an extra cushion of excess capacity against irrational users. The optimization problem then only involves sessions above the cap. Since A_i, the available capacity in period i, is independent of price, the model is essentially unchanged.

For efficient price determination in TDP, the optimization problem must have a scalable solution algorithm. The most useful criterion for this property is convexity: minimizing a convex function over a convex constraint set. We find mild conditions on the $w_j(p,t)$ that make the problem (1-2) convex and accommodate different price- and time-sensitivities.

Proposition 2: If the $w(p,t)$ are increasing and concave in p, and f is piecewise-linear with bounded slope, the ISP's optimization problem is convex.

Proof: See [20]. The key step is finding the cost function's Hessian matrix and observing that ISPs will not offer rewards greater than the marginal benefit of reduced capacity cost. ■

The conditions in Prop. 2 are readily satisfied: following the principle of diminishing marginal utility, w_j should be increasing and concave in p and decrease in t. Users prefer to defer for shorter times. ISP cost can also be readily represented with piecewise-linear functions of bounded slope.

III. Dynamic Session Models and Formulations

A. Offline Model

The dynamic model has offline and online versions. The offline model uses historical demand statistics, and for a single bottleneck network is proven equivalent to the static model.

We assume that sessions arrive according to a Poisson random process, with exponential file size distribution, and leave as a function of the amount of bandwidth allocated to each session. This stochastic model is similar to that in congestion control literature (e.g., see the extensive bibliography in [21]). Each session has a fixed size, e.g. file downloads, and stays in the network until completely processed. As with the static models, we assume a single bottleneck link. We use x to denote the number of sessions arriving on this link and $\Lambda(x)$ to denote the bandwidth allocated to the link by the ISP.

We assume that users defer only once. Consider one time period i, with start time $i-1$ and end time i, and define $N(t)$ as the number of active sessions at time $t \in [0,n]$. Since sessions may be partially processed, $N(t)$ can be non-integral. We assume Poisson session arrival within the period with parameter λ_i. Let $\Pi_i(t)$ denote the number of sessions arriving between time $i-1$ and time t. Session sizes are assumed to be exponentially distributed with mean b. Session arrival times are assumed to be uniformly distributed. Let $\mu(N(t))$ denote the bandwidth allocation in sessions per second.

Proposition 3: The ISP's optimization problem in the offline dynamic model can be formulated as

$$\mathbf{min} \sum_{i=1}^{n} \left(p_i \sum_{k=1, k\neq i}^{n} M_{k,i-k}(k) + f(bN(i)) \right) \tag{3}$$

$$\mathbf{s.\,t.}\; N(t) = N(i-1) - \sum_{k=1}^{n-1} M_{i,k}(t) + \sum_{k=1,k\neq i}^{n} M_{k,i-k}(k) + \Pi_i(t) - \int_{i-1}^{t} \mu(N(s))\, ds,\; t \in [i-1, i] \tag{4}$$

$$M_{i,k}(t) = \int_B \int_{i-1}^{t} \Pi_i(t) g_i(\beta) \times \frac{w_\beta(p_{i+k}, i-1+k-s)}{t-(i-1)}\, ds\, d\beta \tag{5}$$

$\mathbf{var}.\; p_i(k), i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n-1$,

where $M_{i,k}(t)$ denotes the number of sessions deferring from period i to period $i+k$ between time $i-1$ and time t, g_i is the probability density function of the waiting functions w_β parametrized as in the next section by $\boldsymbol{\beta}$, and B is the range of possible $\boldsymbol{\beta}$.

Proof: See [20]. It is similar to that for Prop. 1, but we must keep track of the number of sessions that have arrived and the number still in the network at time t. ■

For a single bottleneck network, $\mu(N)$ is just the access link's fixed capacity. This allows for a closed-form solution for $N(t)$, giving the following proposition:

Proposition 4: For a single bottleneck network, the dynamic model is equivalent to the static model with uniformly distributed arrival times and leftover sessions from one period carrying over into the next period.

Proof: See [20]. The key step compares Props. 1 and 3 using a closed-form solution for $N(t)$. The dynamic model thus retains the static model's computational tractability. ■

B. Online Model

Dynamic programming provides a way to solve the general problem in (3-5) with an online algorithm.

This system's state variables $\mathbf{s}$ consist of the rewards and the number of sessions remaining at the end of each period.[2] The ISP chooses these rewards to minimize the function $C_n(\mathbf{s})$, where C_i is the incurred cost up to period i. The reward p_n in period n is determined first, then p_{n-1}, etc.

We develop a low-complexity dynamic programming solution to the ISP's optimization problem and provide an online algorithm for determining rewards. While sub-optimal, this algorithm is easy to implement and avoids the high dimensionality of a full dynamic programming solution.

Online Price Determination Algorithm.

1: Start with a set of rewards for the next n periods, determined with the static model or offline dynamic model.
2: After the first period, use the static or offline dynamic model to compute the optimal reward for the nth period after this first period, given the other $n-1$ rewards.
3: After each subsequent period, compute the optimal reward for the nth period after the current one.

This algorithm's calculated rewards may not minimize the cost over several future periods. Section V's simulations show how much it can still improve the ISP's cost from that with TIP.

[2]The initial state comes from using some set of initial rewards, for instance determined by solving the problem in the static model.

IV. Waiting Function Estimation

In addition to price optimization as in the previous sections, a TDP system requires a user profiling module which estimates waiting functions and the size of their corresponding sessions. This section briefly describes a practical approach that requires only aggregate TIP and TDP usage data. These may be obtained during initial market trials before rolling out TDP.

The ISP chooses a parametrized family of waiting functions and then estimates each period's parameter distribution. From Prop. 2, these functions should be concave and increasing in p and decreasing in t. One reasonable choice is $w_\beta(t,p) = C_\beta \frac{p}{(t+1)^\beta}$, where the normalization constant C_β depends on the cost of exceeding capacity, number of periods, and β. The parameter $\beta \geq 0$ is a "**patience index**," with larger β indicating lower patience. In practice, since the ISP only sees aggregated sessions at any given time, there will be one β per type of application in each access network.

The ISP estimates waiting functions by observing the difference between demand under TIP and demand under TDP. Let T_i denote this difference in period i. Suppose there are m types of sessions. The parameters β_{j_i} then parametrize waiting functions for type j sessions in period i. The proportion of traffic taken up by each session type in period i is denoted by α_{j_i}. The patience indices and proportions can vary in different periods; in each period, there are m of the β_{j_i} and m of the α_{j_i}, for a total of $2mn$ parameters. The amount of traffic deferred from period i to period $k \neq i$ is then

$$Q_{ik} = X_i \left(\sum_{j=1}^{m} \alpha_{j_i} C \frac{p_k}{(k-i+1)^{\beta_{j_i}}} \right), \qquad (6)$$

where C is the appropriate normalization constant. Each T_i is thus a linear function of the Q_{ik}, yielding n linear equations in the $\frac{n(n-1)}{2}$ variables Q_{ik}. One equation is redundant, since we assume the sum of the T_i is zero (sessions never disappear). The ISP can estimate the parameters α_{j_i} and β_{j_i} as follows:

Waiting function estimation algorithm.

1: Compute the differences T_i between traffic under TIP and TDP, to obtain n linear equations for the Q_{ik}.
2: Solve for $n-2$ of the Q_{ik}, making sure that for each period j, at least one of the Q_{ik} is not solved for.
3: Plug these expressions back into the original equations for T_i, so that only one equation, linear in the Q_{ik}, remains.
4: This remaining equation then becomes a function of the offered rewards and the parameters α_{j_i} and β_{j_i}.
5: Use the TIP and TDP data for this function to estimate (e.g. with nonlinear least-squares) all the α_{j_i} and β_{j_i} parameters involved in this one equation.
6: The parameter estimates give us the waiting functions.

Table II shows the parameter values estimated by nonlinear least squares for a 3 period example with 2 session types. The percent difference between actual and estimated waiting functions for each period remains under 12 percent.

TABLE II
Actual and estimated parameter values in simulation of waiting function estimation.

Period	Actual Values			Estimated Values			Maximum Percent Error
	β_{1_i}	β_{2_i}	α_{1_i}	β_{1_i}	β_{2_i}	α_{1_i}	
1	1	2	0.17	1.03	2.48	0.46	11.8
2	1	2.33	0.5	1.02	2.49	0.45	9.0
3	1	2.67	0.83	0.90	2.15	0.71	0.5

This estimation algorithm uses a baseline measure of aggregate demand under TIP for each period. To account for changes in the baseline over time, we iterate our algorithm. The ISP uses TDP data from a relatively long period of time, e.g. one week, to estimate the waiting functions. It can then take these estimated parameters as given and solve for the demand under TIP, X_i, in each period i. Due to noise in the data, different sets of rewards may give different X_i; the ISP can take an average to determine the baseline X_i.

Since demand under TIP statistics are also used in the price determination, updated TIP estimates directly impact the optimal rewards. Estimation of waiting functions is not perfect no matter what statistical techniques are used, so the extended simulations in [20] stress-test by assuming incorrect waiting functions used by the ISP in its price optimization.

V. Simulation and Performance Evaluation

In this section, aggregate traffic data over times of the day (the blue dotted line in Fig. 3) comes from one week of empirical traces by a large ISP. User patience data is much harder to obtain from existing data sets, so we sweep the waiting function distribution over a range of typical values to quantify TDP's impact.

We first set the number of periods, each period's demand under TIP, sessions' waiting functions, and the ISP's cost function for exceeding capacity, and then set up the offline dynamic model's optimization problem in a standard convex optimization solver. Simulations of the dynamic models may be found in the technical report [20].

We parametrize session waiting functions as in Section IV:

$$w_\beta(p,t) = C_\beta \frac{p}{(t+1)^\beta}, \qquad (7)$$

where β = 0.5, 1, 1.5, …, 5. For simplicity, these w have a linear price- or reward-sensitivity. We assume the following simple cost function for exceeding capacity:

$$f(x_i - A_i) = 3 \max\left[x_i - A_i, 0\right].$$

For illustrative purposes, we use monetary units of $0.10.

We use 48 half hour periods, starting at 12am. Sessions are divided into the 10 waiting function types above; [20] gives the demand under TIP and waiting function distributions in each period. We set the single bottleneck link's capacity to a constant 180 Megabytes/second (MBps). The physical capacity of the bottleneck link may be larger, but ISPs often target the usage to be no more than 80% of the actual capacity, and we use that target as the value of A.

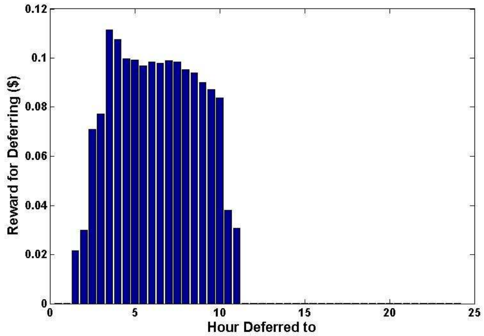

Fig. 2. Optimal rewards, static session model. Rewards have an upper bound of $0.15, and larger rewards roughly correlate with higher traffic.

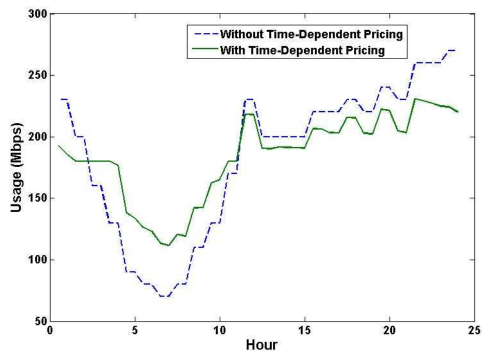

Fig. 3. Traffic profile, static session model. Traffic in over-capacity periods is deferred to under-capacity periods, even-ing out the overall profile.

The optimization yields an average daily cost per user of $3.26 with TDP and $4.26 with TIP (a 24% savings). Figures 2 and 3 respectively show the optimal rewards and traffic profile. Using Section II's propositions, these rewards are both globally optimal and efficiently computed. The optimization ran in under 10 seconds on a standard laptop, so it is easily scalable to a large number of periods and many different session models when run on powerful servers by an ISP. A discussion of the optimal rewards can be found in [20].

From Fig. 3, TDP for the 48 period model decreases the spread between the maximum and minimum usage from 200 to 119 MBps. Overused periods closer to underused ones have the greatest traffic reduction; users more easily defer for shorter times. However, some periods are still over and others still under capacity. TDP cannot effectively even out bandwidth usage fluctuations over a day if users are too impatient, sessions are too time-sensitive, or the cost of exceeding capacity is too low. To measure the even-ing out of traffic over time, we define **residue spread** as the area between a given traffic profile and one with the same total usage but with usage constant across periods. Figure 3 yields a residue spread of 472.5 GB with TDP and 923.4 GB with TIP. The area between the two profiles is 450.9 GB, so 24% of traffic is redistributed over a day.

VI. Next Steps

Based on the theory and algorithms presented thus far, we have implemented a test-bed at the Princeton EDGE Lab for time-dependent pricing. Additional features are being incorporated, such as an "auto-pilot" mode in which users do not need to actively decide when to consume bandwidth, congestion-dependent pricing with fast timescale price adaptation, and a scavenger class of wireless data users. Currently a complete system of TUBE is being prototyped for a customer trial at Princeton. Further details of the models, proofs, emulation and implementation can be found in [20].

References

[1] A. Dowell and R. Cheng, "AT&T Dials Up Limits on Web Data," *The Wall Street Journal*, Jun. 2010.

[2] R. Cox and R. Cyran, "Variable Pricing and Net Neutrality," *The New York Times*, Aug. 2010.

[3] V. Glass and P. U. Princeton Edge Lab, "United States Broadband Goals: Managing Spillover Effects to Increase Availability, Adoption and Investment," 2010, white paper. [Online]. Available: http://scenic.princeton.edu/paper/NECAPrincetonPaperJune2010.pdf

[4] S. Borenstein, "The Long-Run Efficiency of Real-Time Electricity Pricing," *The Energy Journal*, vol. 26, no. 3, pp. 93–116, 2005.

[5] C. R. Associates, "Impact Evaluation of the California Statewide Pricing Pilot," Charles River Associates, Tech. Rep., 2005. [Online]. Available: http://www.calmac.org/publications/2005-03-24_SPP_FINAL_REP.pdf

[6] A. Faruqui and K. Eakin, *Pricing in Competitive Electricity Markets*. Dordrecht, The Netherlands: Kluwer Academic Pub, 2000.

[7] A. Faruqui, R. Hledik, and S. Sergici, "Piloting the Smart Grid," *The Electricity Journal*, vol. 22, no. 7, pp. 55–69, 2009.

[8] A. Faruqui and L. Wood, "Quantifying the Benefits Of Dynamic Pricing In the Mass Market," Edison Electric Institute, Tech. Rep., 2008.

[9] J. Hausmann, M. Kinnucan, and D. McFaddden, "A Two-Level Electricity Demand Model: Evaluation of the Connecticut Time-of-day Pricing Test," *Journal of Econometrics*, vol. 10, no. 3, pp. 263–289, 1979.

[10] K. Herter, "Residential Implementation of Critical-peak Pricing of Electricity," *Energy Policy*, vol. 35, no. 4, pp. 2121–2130, 2007.

[11] S. Littlechild, "Wholesale Spot Price Pass-through," *Journal of Regulatory Economics*, vol. 23, no. 1, pp. 61–91, 2003.

[12] I. Matsukawa, "Household Response to Optional Peak-Load Pricing of Electricity," *Journal of Regulatory Economics*, vol. 20, no. 3, pp. 249–267, 2001.

[13] I. B. M. G. B. Services and eMeter Strategic Consulting, "Ontario Energy Board Smart Price Pilot Final Report," Ontario Energy Board, Tech. Rep., 2007.

[14] J. Wells and D. Haas, *Electricity Markets: Consumers Could Benefit from Demand Programs, But Challenges Remain*. Darby, PA: DIANE Publishing, 2004.

[15] F. Wolak, "Residential Customer Response to Real-Time Pricing: the Anaheim Critical-Peak Pricing Experiment," Stanford University, Tech. Rep., 2006. [Online]. Available: http://www.stanford.edu/wolak

[16] L. Jiang, S. Parekh, and J. Walrand, "Time-Dependent Network Pricing and Bandwidth Trading," in *IEEE Network Operations and Management Symposium Workshops*, 2008, pp. 193–200.

[17] G. Brunekreeft, "Price Capping and Peak-Load-Pricing in Network Industries," *Diskussionsbeiträge des Instituts für Verkehrswissenschaft und Regionalpolitik, Universität Freiburg*, vol. 73, 2000.

[18] H. Chao, "Peak Load Pricing and Capacity Planning with Demand and Supply Uncertainty," *The Bell Journal of Economics*, vol. 14, no. 1, pp. 179–190, 1983.

[19] W. Vickrey, "Responsive Pricing of Public Utility Services," *The Bell Journal of Economics and Management Science*, vol. 2, no. 1, pp. 337–346, 1971.

[20] C. Joe-Wong, S. Ha, and M. Chiang, "Time-Dependent Broadband Pricing: Feasibility and Benefits," Princeton University, Tech. Rep., 2010. [Online]. Available: http://www.princeton.edu/∼chiangm/timedependentpricing.pdf

[21] Y. Yi and M. Chiang, "Stochastic Network Utility Maximisation-a tribute to Kelly's paper published in this journal a decade ago," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 421–442, 2008.