Incentivizing Time-Shifting of Data: A Survey of Time-Dependent Pricing for Internet Access

Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang, Princeton University

ABSTRACT

The tremendous growth in demand for broadband data is forcing ISPs to use pricing as a congestion management tool. This changing landscape of Internet access pricing is evidenced by the elimination of flat rate data plans in favor of usage-based pricing by major wired and wireless operators in the US and Europe. But simple usage-based fees suffer from the problem of imposing costs on all users, irrespective of the network congestion level at a given time. To effectively reduce network congestion, appropriate incentives must be provided to users who are willing to time-shift their data demand from peak to off-peak periods. These pricing incentives can either be static (e.g., two-period daytime/nighttime prices) or computed dynamically (e.g., day-ahead pricing, real-time pricing). Data plans that offer such incentives to consumers fall under the category of time-dependent pricing (TDP). Many ISPs across the world are currently exploring various forms of TDP to manage their traffic growth. This article first outlines the sources of today's challenges, and then discusses current trends from regulatory and technological perspectives. Finally, we review representative pricing proposals for incentivizing the time-shifting of data.

INTRODUCTION

The demand for data is surging rapidly every year, with the Cisco Visual Networking Index for 2012 projecting a 18-fold increase in global mobile data traffic from 2011 to 2016. Mobile video is predicted to be the fastest growing consumer mobile service, increasing from 271 million users in 2011 to 1.6 billion users in 2016 [1]. Much of this demand growth is driven by the popularity of smart devices, bandwidth-hungry applications, cloud-based services, and media-rich web content [2]. This explosive growth in bandwidth demand is now forcing Internet service providers (ISPs) to invest in and expand their wired and wireless capacity by acquiring additional spectrum, deploying WiFi hotspots, and adopting new technologies such as fourth-generation (4G) Long Term Evolution (LTE) and femtocells. However, even these measures are proving insufficient for meeting today's challenges of network congestion.[1,2] Consequently, major U.S. ISPs like AT&T,[3] Verizon,[4] and Comcast[5] have been abandoning the traditional unlimited "flat rate" data plans in favor of such congestion-reducing data plans as throttling, data caps, and tiered usage-based (metered) pricing [3]. But such moves have been fraught with concerns for an open Internet, leading to a polarizing debate on how ISPs should manage their network traffic to create a sustainable Internet ecosystem.

This debate has primarily focused on the issues of:

- Who should pay the price of congestion
- What form such pricing or network management policies should take

The former question, which has serious net-neutrality implications [4, 5], relates to the issue of whether content providers should pay a part of their revenues to finance ISPs' capital investments in the underlying network infrastructure, as discussed later. The latter question, which relates more directly to the theme of this article, centers on the fairness and appropriateness of current pricing policies, such as application- and usage-based pricing. App-based pricing, such as toll-free (zero-rated) access to specific applications as practiced by Mobistar[6] in Belgium, is arguably discriminatory and much more contentious from a net-neutrality standpoint [3]. However, usage-based pricing (or data caps with metering) is widely viewed as an acceptable way to "*match price to cost*," even by the U.S. Federal Communications Commission (FCC) [6]. Yet the effectiveness of usage-based pricing (UBP) has been questioned by both academic and industry veterans due to the fact that usage fees impose costs on users regardless of the actual network congestion in the network at any given time. For instance, Vinton Cerf [7] wrote on Google's Public Policy blog, "*Network Management also should be narrowly tailored, with bandwidth constraints aimed essentially at times of actual congestion. In the middle of the night, available capacity may be entirely sufficient, and thus moderating users' traffic may be unnecessary.*" Andrew Odlyzko *et al.* [3] echoed a similar view in stating that "*Unless UBP contains a time-of-day billing feature and some immediate feedback*

[1] *M. Reardon, "AT&T Wi-Fi Usage Skyrockets,"* Cnet News*, Oct. 24, 2011.*

[2] *E. Caggianni, "Growing Data Demands are Trouble for Verizon, LTE Capacity Nearing Limits,"* Talk Android*, Mar. 6, 2012.*

[3] *A. Lee, "AT&T to Begin Broadband Caps,"* Huffington Post*, May 2, 2011.*

[4] *L. Segall, "Verizon Ends Unlimited Data Plan,"* CNN Money*, July 6, 2011.*

[5] *D. Goldman, "Comcast Scraps Broadband Cap, Moves to Usage-based Billing,"* CNN Money*, May 17, 2012.*

[6] *J. Cryderman, "Leveraging Policy 2.0 and Analytics,"* Pipeline Magazine*, vol. 9, no. 2, July 2012, http://media.pipeline.pubspoke.com/files/issue/14/PDF/PipelineJuly2012_A4.pdf, pp. 3.*

Annual growth statistics				
Mobile video	Total traffic (2011)	62%	Streaming CAGR (2011 - 16)	34%
Source: Cisco VNI 2011 - 2016 (http://www.tinyurl.com/VNI2012)				
Cloud services	Consumer traffic (2011)	83%	Consumer traffic CAGR (2010 - 15)	67%
Source: Cisco global cloud index 2010 - 15 (http://www.tinyurl.com/CiscoCloud2010)				
Data-hungry apps	Facebook (2010)	267%	Skype (2010)	87%
Source: Allot global mobile trends report 2011 (http://money.cnn.com/2011/02/08/technology/smartphone_data_usage/index.htm)				
High-resolution devices	iPad2	1024 x 768 pixels	iPad LTE	2048 x 1536 pixels
Source: Apple (http://www.apple.com/ipad/compare/)				

Figure 1. *Some summary statistics on the annual growth of factors contributing to data explosion.*

The widespread adoption of handheld devices, equipped with powerful processors, high-resolution cameras, and larger displays, has made it convenient for users to stream high-quality videos and exchange large volumes of data.

on congestion it is hard to imagine how it can be used as a congestion management tool."

Cerf and Odlyzko essentially argued that UBP is not an appropriate pricing scheme for congestion alleviation, as it does not address the problem of several users simultaneously accessing the network resources, which in turn leads to large demand peaks and poor network performance at those times. What ISPs can benefit from are smart data pricing[7] (SDP) mechanisms, which can create the right level of *time-varying incentives* that induce consumers to shift their non-critical and bandwidth-heavy applications to periods of lower congestion. Time-dependent pricing (TDP) schemes focus on this aspect, and hence can significantly help ISPs manage their networks. In fact, TDP for voice calls has been widely practiced in the United States by AT&T and Verizon, with free calling on nights and weekends but caps on minutes used during the day. More dynamic versions of TDP have been implemented by MTN Uganda and Uninor India, which update the prices for voice calls hourly, depending on the congestion conditions at the call's originating location [8]. Similar TDP schemes for broadband data have been proposed recently [2, 9] with a full system implementation, and real-world deployment and trials [9, 10].

TDP leverages the trade-offs between consumers' price sensitivity and delay tolerance (time elasticity of demand) across different applications to benefit both consumers and ISPs: consumers benefit from more flexibility of choice, while ISPs benefit from reduced costs of capacity provisioning. Given the level of interest that TDP has received within both academia and industry [2, 9, 10], and its potential as a viable pricing policy for broadband data, this article reviews several TDP research works, with a particular focus on incentive-based models for time-shifting of data. To contextualize the problem, we first provide a brief background of the key factors that are contributing to network congestion, followed by an analysis of the trends in pricing policy changes that ISPs have adopted in response and their regulatory implications.

Key Drivers of Network Congestion

Technological advances as well as changing consumer behavior are driving the explosion in data demand. We discuss these factors next and provide some additional growth statistics in Fig. 1.

Bandwidth-Hungry Devices

The widespread adoption of handheld devices, equipped with powerful processors, high-resolution cameras, and larger displays, has made it convenient for users to stream high-quality videos and exchange large volumes of data. Data from laptops equipped with 3G dongles and netbooks with wireless high-speed data access contributes the most to wireless network congestion [2]. As for smartphones, Cisco projects that the average monthly data usage will rise from 150 MB in 2011 to 2.6 GB in 2016 [1]. New features like Siri on the iPhone 4S, which has doubled Apple users' data consumption, are driving this growth.[8]

Cloud Service and M2M Applications

Cloud-based services that synchronize data across multiple mobile devices, such as iCloud, Dropbox, and Amazon's Cloud Drive, can be a significant factor in traffic growth for ISPs and expensive monthly bills for consumers.[9] Similarly, machine-to-machine (M2M) applications that generate data intermittently (e.g., sensors and

[7] *Smart Data Pricing (SDP) Forum, http://scenic.princeton.edu/SDP2012/.*

[8] *J. Browning, "Apple's Siri Doubles iPhone Data Volumes," Bloomberg News, Jan. 6, 2012.*

[9] *D. Ngo, "iCloud: The Hidden Cost for the Magic, and How to Avoid It," Cnet News, Nov. 7, 2011.*

These developments have led ISPs to view pricing as their ultimate congestion management tool, paving the way for the adoption of Smart Data Pricing measures, such as time-dependent and usage-based pricing.

actuators, smart meters) or continuously (e.g., video surveillance) often load the network with large signaling overhead [2]. Yet the intrinsic time elasticity of such services also creates an opportunity for intelligently shifting them to low-congestion times by applying appropriate price incentives.

CAPACITY-HUNGRY APPLICATIONS

The popularity of handheld devices has also led to rapid growth in the development of bandwidth-hungry applications for social networking, file downloads, music and video streaming, personalized online magazines, and so on. For example, Allot Communications reported that in 2010, traffic from services like Skype grew about 87 percent, and Facebook by 267 percent.[10] Virgin Media Business reports that the average smartphone software uses 10.7 Mbytes/h, with the highest-usage app, Tap Zoo, consuming up to 115 Mbytes/h.[11]

These developments have led ISPs to view pricing as their ultimate congestion management tool, paving the way for the adoption of SDP measures, such as time-dependent and usage-based pricing.

TRENDS IN PRICING FOR CONGESTION MANAGEMENT

FLAT-RATE PRICING

AOL's historic switch[12] in 1996 from hourly rates to an unlimited flat rate data plan made the latter the dominant model for access pricing. However, as shown in Fig. 2, there has been a drastic shift in ISPs' pricing policies since 2008 in both wireline and wireless networks.[13] Comcast's data cap of 250 Gbytes was followed by similar measures for AT&T's U-Verse and digital subscriber line (DSL) services in 2011. Unlimited data plans for mobile services came to the same end: between 2010 and 2012, AT&T, Verizon, and T-Mobile, all eliminated their unlimited plans in favor of tiered data plans. While AT&T and Verizon embraced usage-based overages above data caps, T-Mobile adopted bandwidth throttling for users who exceeded their monthly cap.

USAGE-BASED PRICING

Since 2008, the slow demise of flat-rate plans has been widely covered in the media (Fig. 2). In August 2008, Comcast implemented a 250 Gbyte cap on its cable network per household, threatening repeat offenders (i.e., those who exceed the caps twice within six months) with termination for one year [3, 8]. Time Warner conducted UBP trials in Beaumont, Texas, which offered its customers a 768 kb/s line with 5 Gbyte cap for $29.95/mo or a 15 Mb/s line with 40 Gbyte cap for $54.90/mo, both with $1/Gbyte overage fees. Although these met with poor reception, a similar UBP trial by AT&T from November 2008 to April 2010, conducted in Beaumont and Reno, Nevada, fared better [3]. By May 2011, AT&T introduced a monthly data cap of 250 Gbytes on U-Verse and 150 Gbytes for DSL lines with $10 overage for additional blocks of 50 Gbytes.

For wireless networks, the transition to UBP has been faster. In 2010, AT&T introduced three tiered data plans of $15 for 200 Mbytes, $25 for 2 Gbytes, and $45 for 4 Gbytes; exceeding the caps resulted in $10/Gbyte overage fees for the 2 and 4 Gbyte plans, and $15 for each additional 200 Mbytes for the lowest tier plan [3]. Within a year, AT&T also started throttling the heaviest 5 percent of its users, even those with unlimited data plans [8]. As of July 2012, all major ISPs, including Comcast, AT&T, Verizon, and T-Mobile, have moved to some form of UBP; even Sprint, the only large carrier with an unlimited data plan, cancelled it for laptops and notebooks.[14] Next, we discuss some of the main criticisms and implications of replacing flat rate plans by UBP [3, 7].

APPLICATION-BASED PRICING

Although highly contentious from a network neutrality standpoint, as discussed next, app-based pricing is emerging as a new trend in some European and Asian markets [8]. App-based pricing can come in several forms, such as zero-rating (toll-free) or sponsored content. Under these plans, either the ISP or content provider pays for access to certain popular applications (e.g., Mobistar, a Belgian ISP, has a zero-rate plan for Facebook, Twitter, and Netlog). Another form of app-based pricing is content and access bundling; for example, the Danish operator TDC bundled its music streaming service, TDC Play, into its data plans [8]. In the United States, cable provider Comcast recently came under scrutiny for not counting its own Xfinity for Xbox video-on-demand service toward users' data caps.[15] We discuss the technological and regulatory perspectives of app-based bundling in greater detail later.

TIME-DEPENDENT PRICING

In the telecom industry, time-dependent pricing has long been practiced by ISPs, often as simple two-period (peak and off-peak) pricing for voice calls. For example, both AT&T and Verizon, the two largest wireless operators in the United States, offer free calling on nights and weekends, while capping minutes used during the day. But today, operators in price-sensitive regions such as India and Africa have begun to offer more innovative dynamic TDP for voice calls. MTN Uganda, for instance, has implemented a dynamic tariffing plan[16] in which the price to make a voice call changes hourly depending on the congestion conditions at the call's originating location. While these TDP innovations have been applied only to voice traffic, we discuss a recent system implementation and pilot trial of TDP for mobile data traffic [9, 10].

REGULATORY AND RESEARCH PERSPECTIVES ON PRICING TRENDS

As more innovative pricing practices are proposed by researchers and adopted by ISPs, they should be analyzed with regard to their efficacy in meeting today's challenges, as well as their implications on a free and open Internet. This section discusses academic research that seeks to

[10] *D. Goldman, "You are Using More Data Than You Think,"* CNN Money*, Feb. 8, 2011.*

[11] *K. C. Tofel, "Data Hungry Mobile Apps Eating into Bandwidth Use,"* Gigaom*, May 19, 2011.*

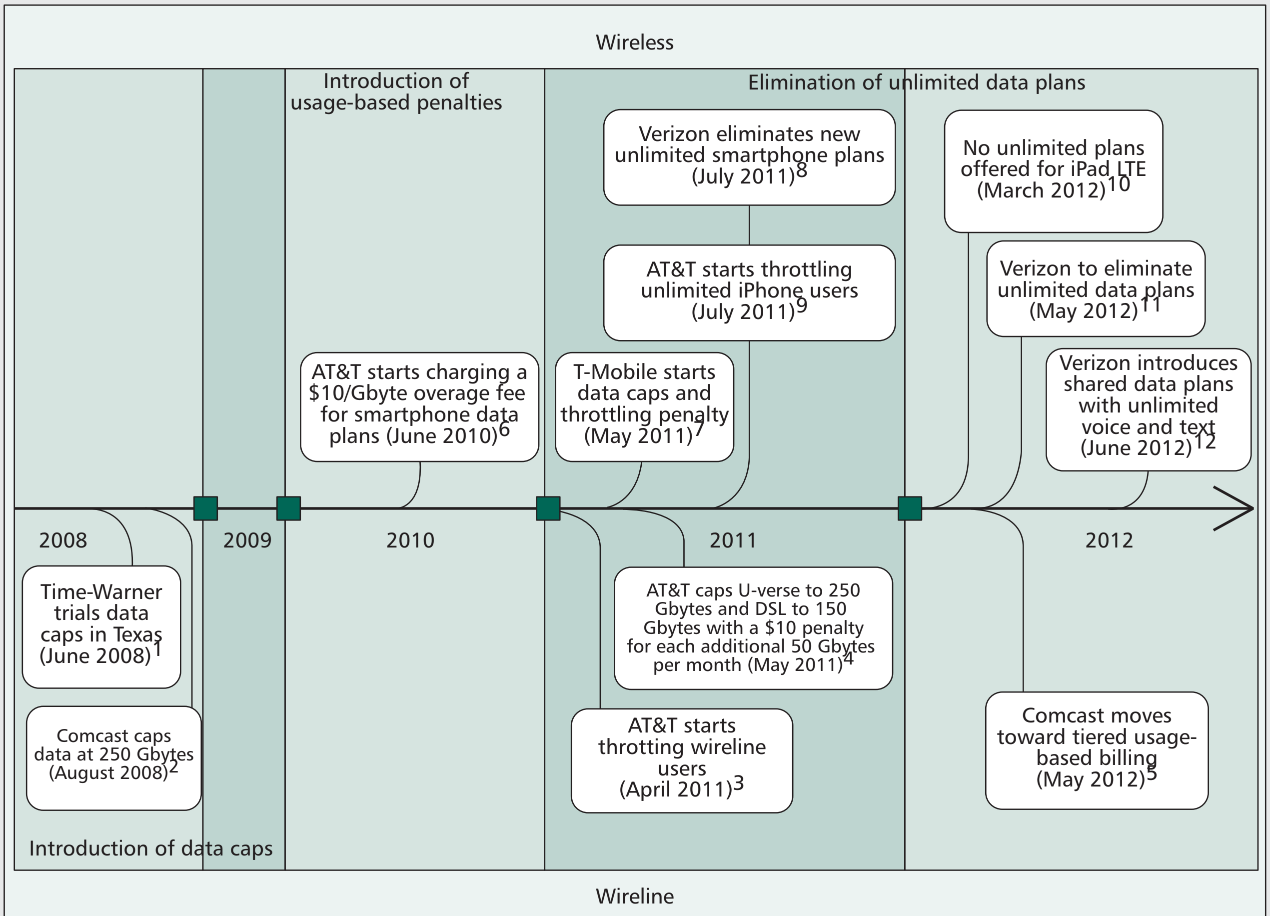
[12] *M. Grebb, "Washington State calls AOL on Flat-Rate Plan,"* Wired*, Nov. 22, 1996.*

[13] *References to news items discussed in this section can be found in Fig. 2.*

[14] *J. O. Gilbert, "Sprint Cancels Unlimited Data for Hotspots, Tablets and Netbooks,"* Huffington Post*, Oct. 21, 2011.*

[15] *D. Mitchell. "Is Comcast Violating Net-Neutrality Rules?"* CNN Money*, May 16, 2012.*

[16] *T. Standage, "The Mother of Invention," Special Report on Mobile Marvels, The Economist. Sept. 24, 2009.*



Wireline

1. R. Paul, "40GB for $55 per month: Time Warner bandwidth caps arrive," ArsTechnica, June 8, 2008.
2. Y. Adegoke, "Comcast to limit customer's broadband usage", Reuters, August 28, 2008.
3. B. Reed, "Bandwidth caps coming to AT&T wireline services," Network World, March 14, 2011.
4. A.Lee, "AT&T to begin broadband caps," Huffington Post, May 2, 2011.
5. D. Goldman, "Comcast scraps broadband cap, moves to usage-based billing," CNN Money, May 17, 2012.

Wireless

6. J. Cox, "AT&T shifts to usage-based wireless data plans," Network World, June 2, 2010.
7. B. Reed, "T-Mobile's data cap embrace leaves Sprint as lone 'unlimited' 4G carrier," Network World, May 23, 2011.
8. L. Segall, "Verizon ends unlimited data plan," CNN Money, July 6, 2011.
9. S. Lawson, "AT&T to throttle big users of unlimited data," Computer World, July 29, 2011.
10. K.C. Tofel, "Apple's iPad 4G LTE plan pricing detailed: Watch your speeds!" GigaOm, March 7, 2012.
11. J. Gilbert, "Verizon unlimited data plans for grandfathered customers to end soon," Huffington Post, May 16, 2012.
12. R. Cheng, "Why Verizon's shared data plan is a raw deal," CNet News, June 12, 2012.

Figure 2. *Key events in the evolution of ISP pricing practices between 2008–2012.*

address some of the key questions and concerns related to pricing policies. An overview of the key viewpoints on these different pricing plans is provided in Table 1.

Is Usage-Based Pricing Useful?

ISPs like Comcast, AT&T, and Verizon have argued to the FCC that UBP is necessary for investing in network infrastructure, so as to match their revenues to the cost incurred from the explosive growth in data usage [6]. The optimal UBP mechanism that maximizes ISP revenue in a resource-constrained network with incomplete user information was presented in [11]. While UBP can be beneficial as a congestion control mechanism when growth in demand exceeds the rate at which extra capacity can be added, its effectiveness depends on its implementation [3]. Odlyzko *et al.* argue in [3, 12] that due to the statistical multiplexing principle of the Internet, bandwidth should not be viewed as a "consumable" good that is unfairly appropriated by bandwidth hogs. Moreover, they argue that the fiber deployment initiatives of AT&T and Verizon were motivated more by competition with cable operators' triple-play services than by any real threat of congestion. In the case of wireless networks, where congestion at base stations is indeed an issue,[17] the effectiveness of UBP is also under question [3, 7]. The key criticism is that UBP cannot alleviate peak congestion unless it contains a time-dependent pricing component to provide dynamic feedback on the network congestion level [10].

[17] *G. Fleishman, "The State of 4G: It's All About Congestion, Not Speed,"* Ars Technica*, Mar. 29, 2010.*

Pricing Policy	Arguments in Favor	Arguments Against	Adoption Trends	Neutrality Perspectives
Flat-rate (unlimited)	• Simpler, preferred by consumers • Users are typically more willing to pay higher flat rate fees.	• Unsustainable because· of "bandwidth hogs." • Light users subsidize heavy users. • Encourages wastage.	• Traditional model • Mostly discarded by US wired and wireless ISPs like Comcast, AT&T, Verizon, T-Mobile	• Modeling flat-rates as a form of bundling -shows the benefit of heterogeneous user ' preferences, despite disproportionate usage volumes [12]. • Flat pricing can cause loss of user net utility, particularly when the consumers have low price sensitivity [16].
Usage-based (UBP: data caps, throttling, overages)	• Allows more efficient use of bandwidth. • Can lower cost for consumers as they pay in proportion to their consumption. • Helps ISPs to match their revenues to costs.	• Lacks the temporal dimension needed to solve ISPs' peak congestion Issue. • Can reduce consumer usage and spending.	• In 2008, Com cast introduced a data cap of 250GB/month per household. • T-Mobile and AT&T throttle heaviest users. • Verizon and AT&T both eliminated flat-rate in lieu of tiered data plans with $10/GB overages.	• Data caps are less efficient than transmission caps [7]. • UBP charges users regardless of network congestion at any given time [3, 7]. • Network maintenance costs are not directly relatable to traffic volume [3]. • ISPs' revenue losses due to regulations against price discrimination can be mitigated by offering prices nonlinear in data volume or data rate, with discounts on higher data-rate demand [16].
App-based (zero-rating, toll-free apps)	• Offers consumer, choice for personalized plans. • App bundling helps ISPs to gain consumers. • Content-provider subsidies (toll-free) benefits consumers.	• Discriminatory, non-neutral • Favors big content and app providers • Deep Packet Inpsection for app classification has privacy implications	• Bundled subscriptions in Europe by Tele-Danmark communications. • Zero-rated access to Facebook, Twitter, Netlog by Mobistar in Belgium. • Access and app bundling by Three in the UK.	• Last-mile termination fees imposed on third party content or application providers will lead to paid prioritization over competing services, harming innovation and creating market Inefficiencies [5]. • Inability to price discriminate across flows causes higher revenue losses when users are less price sensitive [16].
Time-dependent (TOP: static, dynamic)	• Exploits time-elasticity of demand by creating incentives for users to shift demand. • Can reduce peak demand and fill up valley periods.	• Can penalize light, low-income users who cannot shift their demand. • Needs improved user interface for price notifications.	• Exists as daytime/nighttime discount. • Already used in electricity, transport markets. • Dynamic TOP for voice calls is practiced in India, Africa.	• Empowers users with the choice of scheduling elastic traffic to save on monthly bills instead of flat-rate or volume caps [9, 10]. • Impact of TOP on application ecosystem needs further studies and trials [10].

Table 1. *Summary of features and opinions on some current pricing policies.*

Does App-Based Pricing Threaten Net-Neutrality?

Reducing congestion requires capital investment in the network infrastructure, but questions remain on *whom to charge* to subsidize this expense. Given that the consumers, even "bandwidth hogs," impose a relatively small marginal operational cost on the network, Odlyzko *et al.* [3] have argued that pricing mechanisms like UBP can often be misused as a tool to increase return on invested capital. On the other hand, former AT&T CEO Ed Whitacre generated a new debate on net-neutrality by suggesting that in addition to connectivity fees, ISPs should be allowed to charge content providers a part of ad revenues in exchange for the right to access end users.[18]

Recent research by Musacchio *et al.* [4] has shown that two-sided pricing (i.e., a "non-neutral" network) can sometimes increase social welfare, particularly when the ratio of certain model parameters characterizing advertising rates and end-user price sensitivity is either high or low. In contrast to this view, Economides [5] has argued that such ability would lead to paid prioritization arrangements in which app and content providers could pay ISPs to prioritize their packets over competitors'. This would create market inefficiencies and hurt innovation, and can even incentivize ISPs to inflate network congestion problems in order to extract higher revenues from content providers [5]. Similarly, agreements between ISPs and content providers on bundling applications and access or zero-rating specific applications are viewed as non-neutral, discriminatory practices [3].

Can Time-Dependent Pricing Hurt Demand?

One of the main arguments in favor of flat rate pricing as opposed to UBP or TDP has been that consumers are willing to pay more for flat-rate plans. Flat-rate plans encourage higher participation (i.e., a positive network externality), which benefits the Internet ecosystem [3]. Vinton Cerf suggests that time-dependent variants, like UBP with off-peak discounts, can also "*end up creating the wrong incentives for consumers to scale back their use of Internet applications over broadband networks*" [7]. But in contrast to the views expressed above, dynamic TDP for voice calls by MTN Uganda actually increased the demand of their price-sensitive demographic.[14] Broadband analytics firm Wireless Intelligence reports that every one US cent decrease in effective price per minute of voice calls leads to a 6.9 and 13.5 times increase in minutes used per month for developed and developing countries,

[18] *E. Whitacre, "At SBC, It's All About "Scale and Scope," edited by P. O'Connell,* Business Week Mag.*, Nov. 6, 2005.*

Incentive Schemes	Key Model Features	Contributions	Simulation/Systems	Disadvantages
Peak load pricing (PLP (Parris *et al.* [13])	• Higher prices and higher arrival rate of user requests in pre-classified peak hours. • Elasticity of user requests determines if the demand is deferred to off-peak hours.	PLP can flatten out daily demand and generates higher revenue than non-PLP, while also reducing call blocking probability.	Simulations are used to verify the effectiveness of PLP in reducing peak utilization in an integrated network.	Authors show that PLP can segment the user base, with low-budget and low-elasticity users denied service.
Off-peak discount (El-Sayed *et al.* [2])	• Studies the impact of off-peak window size, discounts, and usage shifted on the cost savings and revenues.	Provides several simulation results on capacity savings from peak load shifting.	The observations are based on numerical investigations for a range of scenarios.	Does not include a real world deployment or trial results to validate or quantify the benefits.
Dynamic day-ahead pricing (Ha *et al.* [9, 10])	Optimized "day-ahead" TOP for mobile data computed in a ISP-user feedback control loop.	In trial results, the maximum peak to average ratio of usage volume decreases by 30%.	Conducts the first system implementation and trial of TDP for mobile data.	Requires careful design of user interfaces and data usage notification system.
Real time pricing (MacKie-Mason and Varian [14])	• Dynamic adaptation of prices to congestion level. • ISP admits packet based on user bids for auction.	Introduces "Smart market" for real-time congestion pricing.	Provides analytical results, but does not include a real trial of the system.	Requires automation for end-users to bid on packets during congestion times.
Game-theoretic analysis (Jiang *et al.* [17])	• Introduces a parameterized model of users choosing their usage time based on congestion, price offers, and intrinsic preferences.	Studies Price of Anarchy, i.e., social cost of ISP offering profit-maximizing prices instead of maximizing social welfare.	Numerical examples show cases when ISPs' optimal prices may be extremely socially suboptimal.	Considers only a single service provider scenario.
(Shen and Basar [15])	• Uses non-cooperative games to study user incentives from a mechanism design perspective.	Shows that ISP profits can be larger with non-linear usage fees.	Numerical results show ISP profits increase more than 38% with non-linear pricing over linear fees	Marketing of non-linear pricing structure to consumers can be difficult.
Token bucket pricing (Lee *et al.* [18])	• Users receive a fixed number of tokens that can be used to obtain higher QoS during congested times.	Shows that token pricing gives higher utility to users than flat rate for both single and two service classes.	Numerical studies show that token bucket pricing increases social welfare relative to flat rate.	Requires automation to help users follow the optimal token usage policy.
Raffle-based pricing (Loiseau *et al.* [19])	• Lottery-based rewards in proportion to users' contribution in peak reduction.	Shows that the equilibrium prices maximize the total social welfare.	Numerical studies show robustness of social welfare to changes in user willingness to offload.	ISPs need to measure how much each user offloads, and user compensation depends on lottery.

Table 2. *Comparison of features, contributions, and limitations in representative research works.*

respectively.[19] With respect to mobile data, initial TDP trials (as discussed earlier) have found that TDP increases the overall demand, partly due to higher usage in discounted periods [9, 10].

Incentive Models for Time-Shifting of Data

Shifting demand from congested periods to off-peak times requires pricing policies with the right incentives to induce such a response from users.[20] Some of these policies (e.g., peak-load pricing, off-peak discounts, day-ahead pricing, and real-time pricing) have an explicit time-dependent pricing feature; others, like game-theoretic (auction) models and prioritization-based models (e.g., smart market, raffle-based, token bucket pricing), use more implicit time-varying incentives toward inducing time-shifting of data to less congested periods. An overview of the features, contributions, and limitations of the incentive models discussed below is provided in Table 2.

Peak Load Pricing

Parris *et al.* [13] considers a time-dependent pricing model for peak load pricing (PLP) in which specified periods of time are pre-classified as peak and off-peak hours. In their simulation-based study, the authors consider a higher arrival rate and a higher fee during the peak periods. Each user request has an associated elasticity that determines whether a request arriving during a peak period can be deferred to a subsequent off-peak period. The authors show that PLP can spread out demand over time periods and generate higher revenues than non-peak load pricing, but it can also segment the user base, with low-budget and low-elasticity users denied service. This potential adverse impact of PLP is similar to that of UBP (as observed by Odlyzko *et al.* in [3]) in that it hurts broadband adoption in low-income groups.

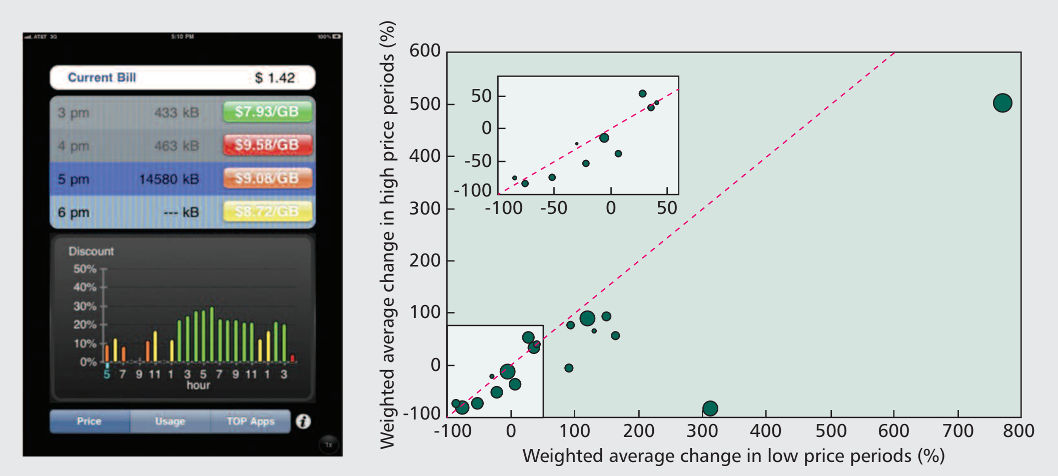
Off-Peak Price Discounts

Off-peak discounting from a predetermined baseline price is a logical dual of peak-load pric-

19 http://www.wirelessintelligence.com/analysis/2011/09/how-pricing-dynamics-affect-mobile-usage/

20 Economic incentives govern the actions of users and ISPs, and it forms a core of the "economic layer" of the network as described by J. Walrand in "Economic Models of Communication Networks," Ch. 3, Performance Modeling and Engineering, Ed. Z. Liu, Springer, 2008.

Auction-based mechanisms help users to explicitly specify their willingness to pay for handling packets from applications with different time elasticities. A limitation of this approach is that it requires intelligent automation on the client side to relieve the end users' burden of decision making.



(a) (b)

Figure 3. *TUBE Trial participants a) could view future prices (color-coded), along with price and usage history. They responded by b) increasing their usage more in low-price, relative to high-price periods.*

ing. El-Sayed *et al.* [2] numerically studied time-dependent price discounts for mobile data. Their model includes three parameters to quantify the impact of TDP on both network cost savings and revenue loss:

- The duration of the off-peak window in which off-peak incentives are offered
- A discount percentage per megabyte
- The percentage of busy hour load that the users shift due to off-peak incentives

The basic idea of offering TDP off-peak discounts is similar to later works on dynamic TDP [9, 10], though these latter works also include an analytical model of a control-feedback loop that adjusts the offered discounts and presents a real-world trial.

Real-Time Pricing

Time shifting of demand can also be realized by providing end users with an explicit pricing signal about the real-time network condition. MacKie-Mason and Varian [14] present the idea of a "Smart Market," which is a closed control-feedback loop in which the network adapts the prices to congestion levels. In this approach, each user places a "bid" on each packet that reflects her willingness to pay to send the packet onto the network at a given time. The gateway admits packets in descending order of their bids as long as the network performance remains above a desired threshold. Users are charged according to the minimum bid on a packet admitted into the network at the time, and thus pay only for the congestion cost at the market-clearing price. Auction-based mechanisms help users to explicitly specify their willingness to pay for handling packets from applications with different time elasticities. A limitation of this approach is that it requires intelligent automation on the client side to relieve the end users' burden of decision making.

Game-Theoretic Pricing

Game theory offers a natural, but somewhat stylized, model for time-dependent pricing, as it allows one to consider consumers' utility in the amount of usage shifted in response to offered prices. Shen and Basar [15] use non-cooperative games to study user incentives from the perspective of mechanism design. Although their work does not consider time-dependent pricing, they show that while most service providers use linear (i.e., usage-based) charging, nonlinear charging can yield large increases in ISP profit. Their results are relevant in the context of observations made by Hande *et al.* [16] that nonlinear pricing can also reduce ISPs' revenue losses, even in regulatory environments that prohibit explicit price discrimination across consumers.

Jiang *et al.* [17] study hourly time-dependent pricing offered by a selfish ISP, comparing the profit-maximizing time-dependent prices to the socially optimal ones. To compute these prices, the authors assume that users choose the time slot of their Internet usage based on the congestion condition, price offered, and intrinsic preference for that time slot. They then focus on the price of anarchy, defined in terms of the social welfare under profit-maximizing time-dependent prices when compared to the maximum social welfare. If several low-utility users are present in the network and the ISP cannot offer different prices to these users, the price of anarchy may be arbitrarily large (i.e., the ISP-optimal prices may be extremely socially suboptimal). This result suggests that some regulatory price restrictions can improve social welfare.

Token Bucket Pricing

Token bucket pricing [18] divides the day into peak and off-peak hours. Users can accumulate daily tokens, which may be exchanged for service during peak hours; without sufficient tokens, users must wait until off-peak hours to consume data. Viewing the tokens as currency, one sees an analogy with time-dependent pricing [2, 9, 14]; the principal difference is that users' budgets with token pricing are determined by the fixed rate at which tokens are received, instead of the variable amounts of money they are willing to spend on peak-hour usage. The presence of these token-based incentives is shown to increase social welfare relative to flat unlimited data plans. However, token pricing may require client-side automation for users to follow "optimal" policies for using their tokens. Moreover, token-based pricing may require several different plans, corresponding to different rates of token accumulation. For instance, a business user may be willing to pay a higher flat fee in exchange for more tokens that can be used during congested hours of the day.

Raffle-based Pricing

Raffle-based pricing also divides the day into two time periods: peak and off-peak. Users are then offered probabilistic incentives to shift their demand to off-peak periods, in the form of a raffle or lottery for a given monetary reward. The probability of winning the lottery is proportional to the user's contribution to the reduction in peak demand. This "all-or-nothing" lottery may instead be replaced by one in which the ISP divides the total reward by the total amount of traffic shifted, paying this amount to each user with a probability equal to the percentage of usage shifted to the off-peak period. In [19], Loiseau *et al.* derive expressions for the Nash equilibrium of this user-ISP interaction and show that in some cases the social welfare is more robust to price variations than a comparable time-of-day pricing plan with two periods. Yet the uncertainty inherent in raffle-based pricing may reduce its effectiveness; users may not shift their usage to off-peak periods without guaranteed rewards. This uncertainty is exacerbated by the reward's dependence on other users' behavior; hence, a field trial of such a policy would be necessary to understand its efficacy.

Dynamic Day-Ahead Pricing

Ha *et al.* [10] introduce Time-Dependent Usage-Based Broadband Price Engineering (TUBE), a fully integrated TDP system that uses "day-ahead pricing," a scheme often used in electricity markets.[21] In this dynamic TDP policy, the prices vary by the hour, but users are informed of the prices one day in advance. This advance notice alleviates the user uncertainty inherent in real-time pricing, allowing users to plan their bandwidth consumption over the next day if desired. TUBE models this user behavior as shifting mobile data usage from one period to another in response to future prices. Users have a certain probability of shifting their traffic from each period to each other period, as a function of the time difference between the periods and the price in the later period. Given functions describing these probabilities, the ISP can calculate the traffic pattern over the day as a function of the prices offered and use this calculation to optimize its profit. These prices are offered to users on their individual devices; the ISP then records the resulting usage and revises its estimates of users' responses to offered prices.

TUBE's principal advantage over PLP [13] and off-peak discounts [2] is its dynamic adjustment of future prices in response to changes in usage behavior; TUBE thus performs better in reducing peak congestion [10] than does static TDP. Additionally, to the best of our knowledge, TUBE is the first functional TDP system for *mobile data* that has undergone a real customer trial[22] [9, 10].

Feasibility Studies

As discussed in the previous sections, incentivizing time-shifting of data has been practiced in the form of static peak and off-peak discounts for voice calls. But more dynamic forms of TDP are emerging in growing economies; MTN Uganda's dynamic tariffing plan for voice calls resulted in a volume increase by 20–30 percent, indicating that price incentives for voice calls can strongly influence user behavior.

The potential benefits of TDP can be higher in the case of mobile data, since many applications today (e.g., online backups in the cloud, app and movie downloads, peer-to-peer) have an intrinsic time elasticity that can be effectively exploited [10].[23] TUBE successfully induced consumers to shift their data demand to lower-priced less congested periods [9, 10]. This nine-month trial consisted of 50 participants with iPads or iPhones and AT&T 3G subscriptions. Trial participants were charged for mobile 3G data usage according to TUBE's day-ahead pricing algorithms. As shown in Fig. 3a, participants could view their price and usage history, as well as future prices, on their iOS devices. In one phase of the trial, the hourly prices alternated between high and low, to test users' price sensitivity. Figure 3b shows the resulting change in usage, measured relative to usage with static UBP of $10/Gbyte. The reference line shows an equal percentage change in usage for low- and high-price periods, with each data point corresponding to the weighted average of one user's percentage changes in usage in all high-price and all low-price periods. Since most data points lie below the reference line, most users increased their usage in low-price than in high-price periods, indicating some price sensitivity. When optimized time-dependent prices were offered, this price sensitivity helped bring down the peak demand, measured in terms of the price-to average ratio (PAR) over each day. The maximum PAR decreased by 30 percent, while the overall usage increased by 130 percent due to higher demand in the valley periods with large discounts [10].

Conclusions

The growing fear of a data tsunami, particularly in wireless networks, has led ISPs to use pricing as a congestion management tool. However, understanding the potential benefits and risks associated with different pricing schemes is crucial in both choosing effective data plans and guiding regulatory decisions in maintaining an open Internet. For more than two decades, research in network economics has shown that providing the right incentives for consumers to time-shift their demand, as commonly done with time-dependent pricing, is the key to resolving the issue of network congestion. In this work, we provide an overview of various incentive mechanisms for time-shifting of demand, including some results from very recent works and consumer trials, with the aim of helping researchers

Raffle-based pricing divides the day into two time periods: peak and off-peak. Users are then offered probabilistic incentives to shift their demand to off-peak periods, in the form of a raffle or lottery for a given monetary reward.

[21] *CNET Energy Pricing. http://www.cntenergy.org/pricing/*

[22] *The TUBE project is similar in its motivation and aim to the Berkeley INDEX project of the 1990's (P. P. Varaiya, R. J. Edell, and H. Chand, INDEX Project Proposal, 1996).*

[23] *Dynamic day-ahead pricing has also been successfully practiced for some years in the electricity industry, e.g., Ontario IESO (http://www.ieso.ca/imoweb/marketData/marketData.asp). Viewing different mobile applications as analogous to appliances with different delay tolerances, one sees a clear analogy between the demand-response models of time-dependent pricing for mobile data and electricity market.*

The growing fear of a data tsunami, particularly in wireless networks, has led ISPs to use pricing as a congestion management tool. However, understanding the potential benefits and risks associated with different pricing schemes is crucial.

and practitioners become aware of the various challenges and opportunities arising from today's network congestion problems. The evolution of new technologies like 4G/LTE, smart grids, and cloud computing will greatly depend on the pricing policy choices that we make today.

ACKNOWLEDGEMENTS

We would like to specially thank Andrew Odlyzko, Roch Guerin, Nicholas Economides, Rob Calderbank, Krishan Sabnani, Shyam Parekh, Sundeep Rangan, Victor Glass, Prashanth Hande, Raj Savoor, Steve Sposato, Patrick Loiseau, and all the participants of the 2012 Smart Data Pricing (SDP) workshop in Princeton, for sharing their valuable comments and insights. We also acknowledge the support of our industrial collaborators, in particular, Reliance Communications, MTA Alaska, AT&T, Qualcomm, Comcast, Mediacom, and the National Exchange Carrier Association.

REFERENCES

[1] Cisco Visual Networking Index, White Paper, Feb. 2012, available online at http://www.cisco.com/ en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf, 2012.
[2] El-Sayed *et al.*, "Mobile Data Explosion: Monetizing the Opportunity through Dynamic Policies and QoS Pipes," *Bell Labs Tech. J.*, vol. 16, no. 2, 2011, pp. 79–100.
[3] A. Odlyzko *et al.*, "Know Your Limits: Considering the Role of Data Caps and Usage Based Billing in Internet Access Service," Public Knowledge White Paper, http://publicknowledge.org/files/UBP%20paper%20FINAL.pdf, 2012.
[4] J. Musacchio, G. Schwartz, and J. Walrand, "A Two-Sided Market Analysis of Provider Investment Incentives With an Application to the Net-Neutrality Issue," *Rev. Network Economics*, vol. 8, no. 1, Berkeley Electronic Press, 2009.
[5] N. Economides, " Why Imposing New Tolls on Third-Party Content and Applications Threatens Innovation and Will Not Improve Broadband Providers' Investment," NET Institute Working Paper No. 10-01; NYU Law and Economics Research Paper No. 10–32, 2010.
[6] S. Higginbotham, "FCC Opens the Door for Metered Broadband," Dec. 1, 2010, http://gigaom.com/2010/12/01/fcc-opens-the-door-for-metered-web-access/, 2010.
[7] C. Vinton, "What's A Reasonable Approach for Managing Broadband Networks?," *Google Public Policy Blog*, August 4, 2008, http://googlepublicpolicy.blogspot.com/2008/08/whats-reasonable-approach-for-managing.html, 2008.
[8] S. Sen *et al.*, Pricing Data: Past Proposals, Current Plans, and Future Trends, http://arxiv.org/pdf/1201.4197.pdf, 2012.
[9] S. Ha *et al.*, "Demo: Enabling Mobile Time-Dependent Pricing," *Proc. ACM Mobisys*, Lake District, U.K., 2012.
[10] S. Ha *et al.*, "TUBE: Time Dependent Pricing for Mobile Data," *Proc. ACM SIGCOMM*, Helsinki, Finland, 2012.
[11] S. Li, J. Huang, and S.-Y. R. Li, "Revenue Maximization for Communication Networks with Usage-Based Pricing," *Proc. Globecom*, Honolulu, HI, 2009.
[12] P. Nabipay, A. Odlyzko, and Z.-L. Zhang, "Flat Versus Metered Rates, Bundling," and "Bandwidth Hogs," *Proc. NetEcon*, San Jose, CA, 2011.
[13] C. Parris, S. Keshav, and D. Ferrari, A Framework for the Study of Pricing in Integrated Networks,Tech. Rep., Tenet Group, ISCI, UC Berkeley, 1992.
[14] J. MacKie-Maso and H. Varian, *Pricing the Internet,* Public Access to the Internet, B. Kahin and J. Keller, Eds., Prentice-Hall, 1995.
[15] H. Shen and T. Basar, "Optimal Nonlinear Pricing for a Monopolistic Network Service Provider with Complete and Incomplete Information," *IEEE JSAC*, vol. 25, no. 6, 2007, pp. 1216–23.
[16] P. Hande *et al.*, "Pricing under Constraints in Access Networks: Revenue Maximization and Congestion Management," *Proc. IEEE INFOCOM*, San Diego, CA, 2010.
[17] L. Jiang, S. Parekh, and J. Walrand, "Time-Dependent Network Pricing and Bandwidth Trading," *Proc. IEEE NOMS Wksp.*, Salvador, Brazil, 2008.
[18] D. Lee *et al.*, "A Token Pricing Scheme for Internet Services," *Proc. 7th ICQT*, Paris, France, 2011.
[19] P. Loiseau *et al.*, "Incentive Schemes for Internet Congestion Management: Raffle versus Time-of-Day Pricing," *2011 Annual Allerton Conf.*, Urbana-Champaign, IL, 2011.

BIOGRAPHIES

SOUMYA SEN [S'08, M'11] received a B.E. (Hons.) in electronics and instrumentation engineering from BITS-Pilani, India, in 2005, and both M.S. and Ph.D. degrees in electrical and systems engineering from the University of Pennsylvania in 2008 and 2011, respectively. He is currently a postdoctoral research associate at Princeton University. He has been actively involved in forging various industrial and academic partnerships, and serves as a reviewer for several ACM and IEEE journals and conferences. He is also a co-organizer of the Smart Data Pricing (SDP) Forum 2012 and GLOBECOM 2012 SDP Industry Forum, which promotes industry-academic interaction on pricing policy research. His research interests are in Internet economics, communication systems, social networks, and network security. He has published several research articles in these areas and won the Best Paper Award at IEEE INFOCOM 2012.

CARLEE JOE-WONG [S'11] is a graduate student at Princeton University's Program in Applied and Computational Mathematics. She received her A.B. (magna cum laude) in mathematics from Princeton University in 2011. Her research interests include network economics and optimal control theory, and her work on the fairness of multi-resource allocations received the Best Paper Award at IEEE INFOCOM 2012. In 2011, she received the NSF Graduate Research Fellowship and the National Defense Science and Engineering Graduate Fellowship.

SANGTAE HA [S'07, M'09, SM'12] is an associate research scholar in the Department of Electrical Engineering at Princeton University, leading the establishment of the Princeton EDGE Laboratory as its associate director. He received his Ph.D. in computer science from North Carolina State University in 2009 and has been an active contributor to the Linux kernel. During his Ph.D. years, he participated in inventing CUBIC, a new TCP congestion control algorithm. Since 2006, CUBIC has been the default TCP algorithm for Linux and is currently being used by more than 40 percent of Internet servers around the world and by several tens of million Linux users for daily Internet communication. His research interests include pricing, greening, cloud storage, congestion control, peer-to-peer networking, and wireless networks.

MUNG CHIANG [S'00, M'03, SM'08, F'12] is a professor of electrical engineering at Princeton University, and an affiliated faculty member in applied and computational mathematics, and computer science. He received his B.S. (Hons.), M.S., and Ph.D. degrees from Stanford University in 1999, 2000, and 2003, respectively, and was an assistant professor 2003–2008 and Associate Professor 2008–2011 at Princeton University. His research on networking has received several awards, including the 2012 IEEE Kiyo Tomiyasu Award, a 2008 U.S. Presidential Early Career Award for Scientists and Engineers, several young investigator awards, and several paper awards. His inventions resulted in a few technology transfers to commercial adoption, and he received a Technology Review TR35 Award in 2007 and founded the Princeton EDGE Laboratory in 2009. He serves as an IEEE Communications Society Distinguished Lecturer in 2012–2013, and wrote an undergraduate textbook, *Networked Life: 20 Questions and Answers*, for a new undergraduate course, Networks: Friends, Money, and Bytes.