To Accept or Not to Accept: The Question of Supplemental Discount Offers in Mobile Data Plans

Abstract—**As demand for Internet usage increases, Internet service providers (ISPs) have begun to explore pricing-based solutions to dampen data demand. However, while many of these solutions focus on reducing usage at times of network congestion, few explicitly consider the dual problem of monetizing idle network capacity at uncongested times. PopData is a recent initiative from Verizon that does so by offering supplemental discount offers (SDOs) at these times, in which users can pay a fixed fee in exchange for unlimited data in the next hour. This work is the first of its kind to assess the benefits and viability of SDOs by modeling user and ISP decisions as a game, considering both overall monthly decisions and hour-to-hour decisions throughout the month. We first use our monthly model to show that users are generally willing to accept some SDO offers, allowing the ISP to increase its revenue. We then show that users face a complex hourly decision problem as to which SDOs they should accept over their billing cycles. They must plan their decisions over the billing cycle, despite not knowing their future usage needs or when future SDOs will be made. The ISP faces a similarly challenging problem in deciding when to offer SDOs so as to maximize its revenue, subject to users' decisions. We develop optimal decision criteria for users and ISPs to decide whether to make or accept SDO offers. Our analysis shows that both users and ISPs can benefit from these offers, and we verify this through numerical experiments on a one-week trace of 20 cellular data users. We find that ISPs can exploit user uncertainty in when future SDOs will be made to optimize its revenue.**

I. Introduction

As mobile data usage continues to grow, with a 66% increase in 2016 [1] alone, Internet service providers (ISPs), mobile service providers in particular, are exploring ways to handle this rising demand. In the U.S., many ISPs have advocated changes to pricing plans; even "unlimited" data plans force users to submit to lower throughputs upon exceeding specified monthly data quotas [2], [3]. Internationally, most ISPs still offer quota-based plans with additional fees for exceeding the quota, e.g., Orange's EE in the U.K. [4]. Such pricing plans incentivize users to limit their overall mobile data demands so that they stay within ISPs' available capacity. However, they do not address the fact that congestion on ISP networks is concentrated at specific times of the day [5]. By reducing overall usage, they can thus have the unintended effect of increasing the amount of idle capacity, and its associated unrealized ISP revenue, at uncongested times.

Much recent research has proposed ways to reduce usage at congested times, e.g., by charging users more at these times [6] or incentivizing them to use WiFi instead [7]. However, few of these explicitly consider the dual problem of monetizing idle capacity [8], and many of them have proven complex for users to understand [9], [10]. *Supplemental discount offers* (SDOs) offer a solution to both problems. SDOs have recently been deployed by Verizon as PopData, a supplement to Verizon users' primary data plans [11]. Under PopData, a user pays an additional fee for unlimited data usage for a limited period of time, e.g., \$3 for one hour of unlimited usage. Over the month, the ISP occasionally makes these SDOs to subscribed users; by making offers in less congested times, it can incentivize users to consume more data at these times without fear of exceeding their data plan quotas. These SDOs may be particularly attractive for users who prefer to use the cellular network instead of public WiFi due to security concerns. Users can easily understand and react to such SDOs; they simply decide whether to accept offers when they are made.

Further inspection reveals, however, that fully understanding or even *optimizing* a user's acceptance of SDOs is quite complex. Such optimization requires a user to *plan their acceptance decisions over the month*. For instance, if a user knows she will not reach her data plan quota, it is better to ignore SDOs. In practice, however, users would not know their exact usage needs for the rest of the month, nor would they know when SDOs will be offered in the future. They thus need to optimize over both sources of uncertainty.

The uncertainty in user decision making leads to an equally challenging decision problem for the ISP. Namely, the ISP wants to offer SDOs at times and prices that maximize revenue, subject to network availability and the fact that user SDO acceptance is based on uncertain future data needs and future SDOs. Yet it is unclear what this optimal schedule would be. For instance, offering SDOs late in the billing cycle may or may not maximize ISP revenue: at that time, only users who know they will exceed their data plan quotas would accept the SDO to avoid overage fees. On the other hand, these users could be more likely to accept SDOs at the end of the month, when they know they will otherwise incur overage fees, than at the beginning of the month, which may increase ISP revenue.

In this work, we model user and ISP actions in accepting and making SDOs as a game in the presence of uncertainty, allowing us to assess SDO benefits for users and ISPs. By handling the uncertainty challenges discussed above, we address four fundamental questions:

- Which types of users would be most affected by SDOs?
- How should the ISP price its SDOs?
- When should ISPs offer and users accept SDOs?
- Are SDOs viable in practice?

To address the question of **which types of users would be most affected by SDOs**, we first consider a model that abstracts away the hour-to-hour SDOs by considering user utility and ISP revenue on a monthly basis. Under this model, we derive closed-form expressions for users' optimal decisions. In this study, we reach the two important conclusions that (1) *subscribers always accept a nonzero number of SDOs* and (2) *users who consume more data per accepted SDO also use more of their data plan*, so heavier users are more affected.

To address the question of **how the ISP should price SDOs**, we extend our model to include the ISP's ability to optimize the SDO price at the beginning of the month. We find that *when all users have limited data demands, the ISP should charge a high price*. However, in a more diverse mix of users, ISPs may reduce fees to incentivize users to accept offers.

To understand **when ISPs should offer SDOs and when users should accept them**, we model user and ISP hourly decisions with an iterative Stackelberg game. We then *derive conditions under which users would accept SDOs*. The ISP's decision problem in this model is NP-hard, so we *provide a near-optimal heuristic based on dynamic programming*. These user and ISP decision algorithms employ online learning to optimize over uncertainty in users' future data needs.

Finally, to assess **SDOs' practical viability**, we conduct extensive *trace-driven simulations* with real usage data to measure the effectiveness of our decision algorithms. We find that *ISPs can exploit user uncertainty in future SDO offers*, and can compute an optimal SDO schedule such that users, in their limited ability to be optimal without knowing the schedule, spend higher with SDOs to realize the same data needs than without SDOs using only overages.

We organize the paper as follows. In Section II, we discuss relevant related work in network management and pricing. We define our abstracted monthly model in Section III, and we develop a more precise model as an iterated game under uncertainty between the ISP and users in Section IV. In Section V, we describe our trace-driven simulation and highlight key results. We conclude the paper in Section VI with a discussion of our findings and potential future work. All proofs can be found in the appendix.

II. Related Work

To limit usage during congested times, some industry [12], [13] and academic [14]–[16] research has advocated for time-dependent pricing (TDP) for mobile data. Under TDP, users are charged higher rates when the network is congested and lower rates during times of low network utilization. These previous studies assessed the benefits of TDP compared to static pricing [14], [15], e.g., with game-theoretic models [16]. TDP has been shown to be effective in user trials for cellular networks [8], [9] and smart grids [17]. Users under TDP not only reduced their usage at congested, high-price times, but also increased their usage at uncongested, low-price times. We focus on this latter effect in our work. Variations on TDP include incorporating location into pricing models [18] and using lotteries to offer time-dependent rewards for reducing usage at congested times [19]. Many works show that ISPs can reduce congestion and increase revenue by offering different prices at different times of the day, but the apparent complexity for users has so far prevented deployment.

TABLE I: We summarize the notation used in the paper.

Symbol	Definition
(η, d, p)	ISP Data Plan
η	Fixed monthly charge
d	Data limit
p	Overage charge per GB beyond data limit
ρ	SDO price
n	Number of times ISP offers SDO over a month
β	Fraction of SDOs accepted
x	Monthly data usage by user
x_{max}	Maximum data consumed by a user during an SDO period
α	User price sensitivity
γ	Desired monthly maximum data consumption
$x_c(t)$	User's accrued consumption under their data plan until time t.

Complementary work has focused on offloading users' data traffic from cellular to WiFi [7], e.g., creating auctions for ISPs to dynamically purchase WiFi capacity at times of cellular network congestion [20], [21]. Yet while these measures can decrease congestion for ISPs, they may also decrease ISP profits, not only due to the cost of purchasing WiFi capacity, but also due to the reduction in usage on cellular networks. To model this loss in revenue in our discount offers scenario, we include the presence of WiFi in users' hourly decisions in Section IV-A. Other work has used large-scale usage datasets to model how users consume their data quotas over a month [22]. We leverage similar frameworks in developing user and ISP decision algorithms in Section IV.

III. Monthly SDO Decision Model

To assess the benefits of SDOs, we model the ISP and users respectively as the leader and followers in a game. The ISP offers and prices SDOs, and users decide whether to accept them. We assume a monopolistic ISP that offers a quota-based data plan to users, imposing a usage-based overage fee p per unit of data used over the monthly data quota d, with flat fee η. In addition, the ISP periodically makes SDOs; a user who accepts an SDO pays a fixed price ρ for unlimited data usage in the next time slot (e.g., one hour). Although a user's data use during this time slot is contractually unbounded, usage is still subject to network constraints and would in practice be finite. We assume that a user consumes a maximum of x_{max} data during an SDO session. We further assume there are N users in the system. Table I summarizes our notation.

In this section, we derive a monthly model of user and ISP behavior using a Stackelberg game. While this model is an approximation that abstracts away hourly dynamics, it provides qualitative insights into user benefits and SDO pricing. Under this model, the ISP sets the number of SDOs n offered during the month and chooses the optimal SDO price ρ in anticipation of user decisions. In the model developed in Section-IV, the ISP implicitly chooses n, or how many SDOs to offer over the billing cycle, by making hourly decisions on whether to offer SDOs. Given n and ρ at the start of the month, each user further chooses two parameters: the fraction β of accepted SDOs and their monthly data plan usage x.

A. Modeling User Utility

We model users' utilities as having two components: utility from data plan usage and utility from SDOs. We use the standard α-fair models for user utility from monthly data usage [23], [24] to obtain the utility function

$$u(x,\beta) = C_1 \frac{x^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] - \eta - p(x-d)^+, \quad (1)$$

where C_1 and C_2 are scaling factors capturing relative utility between data plan and SDO usage, and $\alpha \in [0,1)$ indicates the user's price sensitivity. The first terms in this utility function represents the overall utility from a user's regular monthly data plan and from SDOs, respectively. Since β represents the fraction of SDOs that the user accepts and n the number of offers that are made, we can interpret the utility from SDOs as the user receiving a utility of $C_2 x_{max}^{1-\alpha}/(1-\alpha) - \rho$ each time an offer is accepted. C_2 scales the utility from x_{max} usage depending on the received quality of service (QoS). If the average QoS during SDO periods is high, then the user will receive a higher utility from consuming data at that time. Similarly, C_1 can be scaled to represent the average QoS at non-discount times. By using different scaling factors for SDO and non-SDO times, we can model ISPs' choice of making SDOs only at uncongested hours of the day. The last term in (1) represents the cost of data plan usage with $(x-d)^+$ denoting the amount of users' overage data.

We further suppose that the user's overall data usage is constrained by a monthly maximum γ, imposing the constraint

$$x + \beta n x_{max} \leq \gamma. \quad (2)$$

For instance, we could take $\gamma = x_{max} T$, where T is the total number of time periods in a month. This maximum usage indicates the inherent limit on the amount of data that a user would consume even if not charged for this data usage. Since users in reality would limit their data consumption so as to avoid paying more for data, we assume that $\gamma \geq \max\left\{d, (C_1/p)^{1/\alpha}\right\}$, i.e., maximum usage γ without data costs is no less than the user's optimal data plan usage.

B. Optimizing User Utility

In maximizing the utility (1) subject to the constraint (2), the user jointly optimizes the data x consumed under the regular data plan and the fraction β of accepted SDOs for the month.

Optimizing Monthly Data Usage x. We initially consider β as given and identify the optimal values of x under different conditions, yielding the following.

Lemma 1. *The user's optimal data plan usage x^* is given by*

$$x^* = \begin{cases} d, & \text{if } (\frac{C_1}{p})^{1/\alpha} \leq d \\ (\frac{C_1}{p})^{1/\alpha}, & \text{if } d \leq (\frac{C_1}{p})^{1/\alpha} \leq \gamma - \beta n x_{max} \\ \gamma - \beta n x_{max}, & \text{if } (\frac{C_1}{p})^{1/\alpha} \geq \gamma - \beta n x_{max}. \end{cases}$$

Thus, if no SDOs are made ($n = 0$), the user would consume $x^* = \max\left\{d, (C_1/p)^{1/\alpha}\right\}$ amount of data.

TABLE II: Optimal x^* and β^* that maximize user utility (1) under different conditions on d (columns) and ρ (rows).

Conditions	$d \geq (C_1/p)^{1/\alpha}$	$d < (C_1/p)^{1/\alpha}$
$\rho \geq \frac{C_2 x_{max}^{1-\alpha}}{1-\alpha}$	$x^* = d$ $\beta^* = 0$	$x^* = (C_1/p)^{1/\alpha}$ $\beta^* = 0$
$\rho < \frac{C_2 x_{max}^{1-\alpha}}{1-\alpha}$	$x^* = d$ $\beta^* = \max\left\{\frac{\gamma - d}{n x_{max}}, 1\right\}$	$x^* = d'$ $\beta^* = \max\left\{\frac{\gamma - d'}{n x_{max}}, 1\right\}$
In the above, $d' = \left(\frac{C_1 x_{max}}{C_2 x_{max}^{1-\alpha}/(1-\alpha) - \rho + p x_{max}}\right)^{1/\alpha}$.		

From Lemma 1, we observe that if $(C_1/p)^{1/\alpha} \leq \gamma - \beta n x_{max}$, then the user's data plan usage would not change with SDOs. Thus, *heavy users' data plan consumption is most affected by SDOs*; light users would not change their usage behavior. These "light" users would have lower C_1 values, indicating that their marginal value from data consumption is low compared to the cost of their data plan.

Optimizing the Discount Acceptance Rate β. The above insight into lighter and heavier users is also reflected in the fraction β of accepted SDOs, as follows.

Proposition 1. *Table II gives the optimal (x^*, β^*) that maximize the utility (1) subject to the usage constraint (2).*

This table defines the different boundary conditions under which distinct utility-maximizing solutions emerge. We see that users with $(C_1/p)^{1/\alpha} \leq d$ would not change their data plan usage based on SDOs, rather supplementing their data plan with SDOs as needed. However, heavier users, as identified in Lemma 1, with $(C_1/p)^{1/\alpha} > d$, would change their data plan usage. Without SDOs, these users would consume $x^* = (C_1/p)^{1/\alpha}$ including overage usage. By inspection of Table II, we conclude that they always consume less than that when SDOs are made.

Corollary 1. *If $C_2 x_{max}^{1-\alpha}/(1-\alpha) > \rho$, i.e., the user gains positive utility from SDOs, then $\beta^* > 0$ and the user accepts at least some SDOs. However, data plan usage reduces with SDOs, as $x^* < \max\left\{d, (C_1/p)^{1/\alpha}\right\}$.*

We observe from this corollary that if users would have consumed overage data without SDOs, then *no matter how small their utility from the SDOs, they would replace some of their overage data consumption with SDO usage*. However, light users would still consume their data quota d (cf. Lemma 1), though they might accept SDOs on top of this usage. We next focus on how heavy users' data plan consumption with SDOs depends on their individual characteristics. In particular, we find that users' data plan usage x^* can increase with x_{max}.

Corollary 2. *If $(C_1/p)^{1/\alpha} > d$ and $\alpha < \rho(1-\alpha)/(C_2 x_{max}^{1-\alpha}) < 1$, usage x^* is minimized when $x_{max} = \rho(1/\alpha - 1)^{1/(1-\alpha)}$. When $x_{max} \geq \rho(1/\alpha - 1)^{1/(1-\alpha)}$, x^* increases with x_{max}.*

This result is somewhat surprising; we would expect larger x_{max} to lead to higher β, with less data plan usage. However, the opposite effect occurs when x_{max} is large. We can partially explain this latter result by noting that as x_{max} increases, users would approach their monthly data quota γ faster with each SDO. Thus, they would prefer to accept fewer offers, spreading their data more evenly throughout the month by consuming more of their data plan. This is particularly true for less price-sensitive users (with higher α), whose utility from an SDO session would increase slowly as x_{max} increases. They could then realize larger marginal utilities from usage on their data plans, compared to SDO usage.

We next examine the effect of the maximum usage γ in more detail. In particular, we observe that γ may be larger for users with a larger x_{max}, since both represent bounds on the user's desired data consumption.

Proposition 2. *If $\gamma = cx_{max}$ for a fixed $c > 0$ and a user has positive utility from SDOs, then both x^* and β^* increase as x_{max} increases, when $x_{max} \geq \rho(1/\alpha - 1)^{1/(1-\alpha)}$.*

In this scenario, *a larger x_{max} would lead to a larger maximum usage γ, allowing users to both accept more SDOs and consume more of their data plan.* Thus, even though users would consume more data per SDO as x_{max} increases, they would still increase both types of usage. However, users' data plan usage is still bounded by their usage without SDOs (Corollary 1); even as $x_{max} \to \infty$, $x^* \to \max\left\{d, (C_1/p)^{1/\alpha}\right\}$.

C. Maximizing ISP Revenue

Given the optimal user decisions in Proposition 1, we next find the optimum SDO price ρ to maximize ISP revenue. Since the ISP would set ρ at the beginning of the month, the monthly model guides this choice for a given number of SDO offers n. The ISP's choice of n is further considered in Section IV-B.

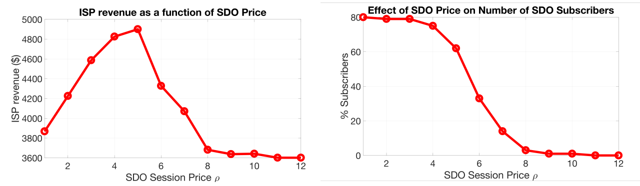
The ISP's revenue function is the sum of the revenue obtained from each user over the billing cycle, so the objective is to choose ρ to maximize this revenue, formulated as

$$\max_{\rho} \sum_{i \in U} \left(\eta + p(x_i^*(\rho) - d)^+ + \beta_i^*(\rho) n \rho\right), \quad \text{s.t. } \rho \geq 0, \tag{3}$$

where the subscript i is added to indicate user-specific values. We thus see that (3) is a complex optimization problem; the set of users whose x^* and β^* expressions fall into the different categories in Table II depend on ρ. We do not derive an analytical solution, since a line search will suffice to find the optimal ρ^*. We can, however, observe that when all users are light users, the ISP would charge them as much as possible.

Proposition 3. *When all users are homogeneous light users who do not consume overage data (i.e., $(C_1/p)^{1/\alpha} \leq d$, where C_1, C_2, α, and d are the same for all users), the optimal price ρ^* in (3) is $\rho = C_2 x_{max}^{1-\alpha}/(1-\alpha)$.*

From Table II, we see that as long as these users have positive utility from SDOs, they would accept as many offers as necessary to realize their maximum usage γ. Thus, the ISP



(a) ISP Revenue as a function of SDO Price

(b) Effect of SDO Price on Number of SDO Subscribers

Fig. 1: ISP revenue (a) fluctuates as the SDO price ρ increases, since (b) fewer users accept SDO offers for large ρ. These results correspond to a distribution of users with mean $\alpha = 0.5$ and $x_{max} = 0.5$GB.

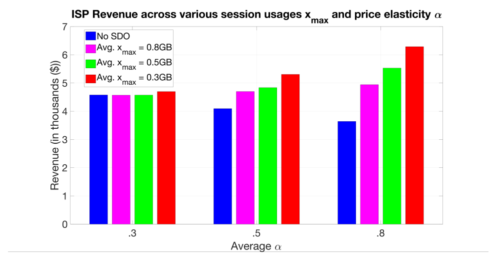


Fig. 2: The optimal ISP revenue always increases compared to revenue without SDOs, and is higher for less price-sensitive (higher α) users who consume more data with SDOs (lower x_{max}).

would have an incentive to charge as much as possible for these accepted offers. However, when there is a more diverse mix of users, the largest ρ may not be optimal. Figure 1(b) shows the ISP revenue as a function of ρ for a distribution of 100 light and heavy users. The optimal $\rho^* = \$5$ is lower than if all users were "light" users (Proposition 3), since the ISP can decrease ρ to encourage heavy users to accept more SDOs. Figure 1(a) also shows the decrease in the percentage of SDO subscribers (i.e., users who derive positive utility from an SDO) with SDO session price ρ. There is a steep drop-off in the subscription rate around $\rho = \$6$, indicating that many users no longer derive positive utility from SDOs $\left(C_2 x_{max}^{1-\alpha}/(1-\alpha) < \rho\right)$.

As in our analysis of SDOs' benefits in Section III-B, Figure 2 shows the optimal ISP revenue for user populations with different α and x_{max} values, compared to a scenario without SDOs. ISPs always increase their revenue by offering SDOs, especially for users with a higher price sensitivity; these users will accept more SDOs to avoid overage charges. As x_{max} increases, ISPs also earn more revenue, as indicated by Corollary 2, leading to more SDO revenue.

IV. Hourly Stackelberg Game

Building on the high-level insights provided by the monthly model, we develop a game between users and ISPs to model hour-by-hour SDO decisions. In what follows, we derive a decision criterion for users to accept SDOs and propose an algorithm to optimize ISP SDO schedules.

We break the monthly billing cycle into T time steps, e.g., $T = 720$ hours in a 30-day month. At the start of each time step t, the ISP notifies users if an SDO is offered ($y_t = 1$) or

TABLE III: We provide a list of additional symbols and definitions for the dynamic interaction model.

Symbol	Definition
t	indexes the time intervals that the billing cycle has been divided into
a_t	Binary variable indicating a user's SDO decision for t^{th} period
y_t	Binary variable indicating the ISP's SDO decision in the t-th period
x_t	User intended data consumption in t^{th} period
ϕ_t	QoS needs of a user's $x_t \in [0, 1]$
θ_t	Cellular network congestion measure for the t-th period $\in [0, 1]$
θ_W	Typical Public Wifi congestion measure in the region of interest $\in [0, 1]$
δ_W	User-specific Public WiFi preference metric $\in [0, 1]$

not ($y_t = 0$). An ISP's *SDO schedule* is the resulting set of decisions $\{y_t, t = 1, \ldots, T\}$. If an SDO is offered at time t, users respond by accepting or declining the SDO at the fixed price ρ. We model overage as an addition of d_O to the user's data quota at a cost p, as offered by most ISPs [25][1].

Suppose that at time t, a user has previously consumed $x_c(t)$ of their data plan quota and the total data quota currently sits at D_t, including any previously incurred overages. During time slot t, the user intends to consume x_t additional data at a desired QoS level $\phi_t \in [0, 1]$. For instance, a videoconference session may warrant a high ϕ_t, while accessing email may tolerate a low ϕ_t. To realize the desired x_t, the user can use their data plan, public WiFi if available, or an SDO if offered. We account for congestion and price sensitivity effects as follows. We define the congestion level θ_t of the cellular network, which is known to both the user and ISP, as well as the typical congestion θ_W for public Wifi networks. We also define $\delta_W \in [0, 1]$ as a user-specific parameter that captures the user's public WiFi preference, ranging from complete aversion ($\delta_W = 0$) to no aversion ($\delta_W = 1$), possibly based on security preferences as previously described. As ϕ_t increases and θ_t decreases, users experience more utility from their usage. A detailed list of this notation is presented in Table III.

A. User Decision Criteria

At each time t, the user must choose whether to accept the SDO if offered, use a data plan, or use WiFi if available. We assume that users are myopic, i.e., they do not plan their decisions over the month (since they do not know when SDOs will be offered, this would be prohibitively difficult). Instead, users make decisions based on perceived utility at the current time, with awareness of the risk of incurring future overages.

The user's utility at time t from SDOs, her data plan, and public WiFi are respectively given by

$$u_P(t) = (1 - \theta_t \phi_t) x_t - \rho(1 - \alpha), \tag{4}$$

$$u_D(t) = (1 - \theta_t \phi_t) x_t - R_t p(1 - \alpha) + N_t u_O(t), \tag{5}$$

$$u_W(t) = (1 - \theta_W \phi_t) x_t \delta_W, \tag{6}$$

respectively, with corresponding costs of access scaled by the user's price sensitivity[2]. R_t represents the risk of incurring a new overage in the remainder of the billing cycle (i.e., at time $\tau \geq t$), which depends on cumulative data plan usage up to time t, as well as x_t. Thus, usage decisions at time t affect the future risk of overage R_τ for $\tau \geq t$, as this risk evolves over the billing cycle. We account for user's utility from the extra data quota earned when incurring another overage charge by defining an overage utility $u_O(t)$. $N_t = 1$ indicates that the user incurs a new overage at time t (and 0 otherwise), so $u_O(t)$ is only realized if $N_t = 1$. We next discuss how a user would estimate the overage factors R_t and $u_O(t)$.

Modeling Risk R_t of New Overage. We define R_t as the probability that the user will incur a new overage charge in the remainder of the billing cycle. Computing this probability, however, is difficult, as the user would not know exactly how much data they would consume in the rest of the month. We thus propose to estimate this future usage by leveraging the user's historical usage patterns. We suppose the user has a typical pattern of data usage during the billing cycle, e.g., consistent usage throughout the cycle or gradually ramping up usage toward the end [22]. We model these consumption trends over the billing cycle as a random process $X(t) \sim \mathcal{F}_{\sigma(t)}(at^b)$ representing the user's cumulative (non-SDO) data consumption until time t. $\mathcal{F}_{\sigma(t)}$ represents a distribution around the mean cumulative usage at^b, parameterized by σ_t, e.g., a normal distribution with variance σ_t. We can learn the parameters a, b, and σ_t for each user from previous usage patterns[3]. In the appendix, we present a maximum-likelihood method for the user to learn the optimal a, b, and σ parameters assuming $\mathcal{F}$ is a normal distribution.

The probability R_t of incurring a new overage in the current cycle can then be written as

$$R_t = \mathbb{P}\left(X(T) > D_t | X(T) \geq x_c(t) + x_t\right), \tag{7}$$

where $X(T)$ is the total usage in the billing cycle.

Modeling Utility $u_O(t)$ from Overage. We define $u_O(t)$ as analogous to users' data plan utility in (5).

$$\begin{aligned} u_O(t) = H \min(&D_t + d_O - x_c(t) - x_t, \\ &\mathbb{E}[X(T) | X(T) \geq x_c(t) + x_t]), \end{aligned} \tag{8}$$

where $\mathbb{E}[\cdot]$ denotes expectation. The argument of the min function in (8) represents users' expected utility from the d_O data added to their quotas with an overage. The factor $H \in [0, 1]$ qualitatively captures any decrease in the actual utility realized in the future from the leftover data, e.g., due to future values of $(1 - \phi_t \theta_t)$. Predicting these exact values is likely impossible, as the user does not know their future data needs, but including H abstracts from the exact details.

Optimizing User Utility. At the start of time step t, the user chooses to consume data on an SDO, data plan, or WiFi to maximize utility. Given the utilities from each choice (4), (5), and (6), we derive the user's optimal decision criterion.

[1]Note that our monthly model in Section III-A uses continuous overage costs, but at the finer hourly timescale, our overage amounts are discrete.

[2]At a finer time scale, the concavity of a user's monthly utility as in Section III-A does not appear; thus, we assume a utility linear in x_t.

[3]The distribution $\mathcal{F}_{\sigma(t)}$ can be induced by an underlying random process on the parameters of users' utility functions, which will drive their demands x_t. However, these utility parameters are not directly observable by the user or ISP, so we model the directly observable usage itself as a random variable.

Proposition 4. *The user's optimal choice c^* of data access during t when overage is not expected ($N_t = 0$) is given by*

$$c^* = \begin{cases} \text{SDO}, & \text{if } \rho < R_t p \text{ and } v > \rho\alpha' \\ \text{Public WiFi}, & \text{if } v < \rho\alpha' \text{ and } v < R_t p\alpha' \\ \text{Data plan}, & \text{otherwise}, \end{cases}$$

while the optimal choice c^* during t when overage is expected ($N_t = 1$) is given by

$$c^* = \begin{cases} \text{SDO}, & \text{if } u_O(t) < (p-\rho)\alpha' \text{ and } v > \rho\alpha' \\ \text{Public WiFi}, & \text{if } v < \rho\alpha' \text{ and } v < p\alpha' - u_O(t) \\ \text{Data plan}, & \text{otherwise}. \end{cases}$$

where $v = x_t(1 - \phi_t(\theta_t - \theta_W\delta_W) - \delta_W)$ and $\alpha' = 1 - \alpha$.

From Proposition 4, we see that when the user is not expected to go into overage at time t ($N_t = 0$), SDO is the dominant choice over data plan if it costs less than the expected overage price $R_t p$. Between WiFi and SDO, we see that SDO is the dominant choice only when the congestion in the cellular network is lower WiFi's, subject to how important QoS is to the user (ϕ_t) and the user's affinity (or lack thereof) for WiFi δ_W. The overage case in Proposition 4 results in $R_t = 1$, and SDO is better than the data plan only if the estimated future utility from overage $u_O(t)$ is less than additional cost incurred by an overage over SDO, subject to the user's price sensitivity.

B. ISP Revenue Formulation

We next consider the ISP's decision of when to offer SDOs, given that users will respond according to Proposition 4. The ISP's revenue $r_{i,t}$ from user i in time period t is given by

$$r_{i,t} = a_{i,t} y_t \rho + (1 - a_{i,t} y_t)\omega_{i,t} N_{i,t} p \tag{9}$$

where $a_{i,t}$ is the user's binary decision to accept an SDO, depending on whether an SDO is offered at time t, and $\omega_{i,t} = 1$ if the user does not offload to WiFi. These can be found from each user's decision c^* in Proposition 4. Hence if $(1 - a_{i,t} y_t)\omega_{i,t} = 1$, the user does not accept an SDO but continues to use her data plan. If a new overage is incurred by i at t, then $N_{i,t} = 1$, and the ISP earns the overage price p.

While choosing the optimal y_t for (9) would maximize the ISP's revenue in time slot t, this could be sub-optimal in regard to the monthly billing cycle. The ISP must then account for the fact that its decision to offer an SDO at time t will affect users' risk of incurring an overage and hence the future acceptance of SDOs and future revenue. Hence, even though the ISP does not reveal the future SDO schedule to users, the current SDO decision is a function of the optimal schedule over the entire cycle. Therefore, this must be calculated at $t = 0$ for maximizing revenue over the entire billing cycle. The ISP thus aims to maximize the total revenue by optimizing the SDO schedule $\mathbf{y} = \{y_1, \ldots, y_T\}$ as

$$\mathbf{y}^* = \underset{y_1,\ldots,y_T}{\operatorname{argmax}}\, \mathbb{E}\left(\sum_{i\in U}\sum_{t=1}^{T} r_{i,t}\right) \tag{10}$$

In the revenue optimization in (10), note that the revenue terms $r_{i,t}$ are necessarily dependent on each other over time, seen by the inclusion of overage and conditional decision terms in (9). Most importantly, the expectation appears in (10) to capture the effects of the uncertainty in user decisions. In practice, the ISP could execute y_t^* at each time t and then re-compute its optimal schedule for the rest of a billing cycle given updated estimates of user parameters.

C. Optimizing ISP Revenue

To solve (10), the ISP must compute the distributions of $N_{i,t}$ and $a_{i,t}$ for each user so as to derive the expectation of $r_{i,t}$ in (9), noting that both depend on previous values of y_t. To do so, the ISP must estimate the parameters θ_t, α_i, and $\phi_{i,t}$ that influence users' SDO acceptance decisions in Proposition 4. While the ISP would know the cellular and WiFi congestion levels θ_t, it would need to use historical data from the user to estimate the user-specific $\phi_{i,t}$ and α_i parameters. The ISP must then estimate the distribution of users' future usage $x_{i,t}$. We suppose that it does so using the same method as the user in Section IV-A. Given this knowledge of user behavior, we can then recast (10) as a dynamic program and derive a heuristic algorithm to compute an approximate solution.

Dynamic Programming Formulation. The solution to (10) can be found by formulating the following Bellman equation for computing the optimal revenue V_t^* at t. It can be expressed as a function of the current time step decision y_t that allows for the maximum expected sum r_t (one summation over users in (10)) from the current time step revenue r_t and the optimal revenue from the next time step $V_t^*(\mathbf{D}_{t+1}, t+1)$, given by

$$V_t^*(\mathbf{D_t}, t) = \max_{y_t}\left(r_t + V_t^*(\mathbf{D_{t+1}}, t+1)\right), \tag{11}$$

where the boldface $\mathbf{D_t}$ is a vector of all users' data quotas and all of the terms depend on current and past values of y_t. The corresponding y_t value becomes the t^{th} entry in $\mathbf{y}^*$. As we see from (11), user quotas $\mathbf{D_{t+1}}$ at time $t+1$ are a function of the decision y_t from the current-step, according to the relationship

$$D_{i,t+1} = D_{i,t} + N_{i,t} d_O, \tag{12}$$

and as we know from (9), $N_{i,t}$ is a function of the SDO decision y_t. Thus, this data quota state update mechanism at every time step captures the tradeoff between overage and SDO revenue, dependent on both y_t and $\mathbf{D_t}$. The results of solving (11) are presented in Section V.

Fast Pruning Algorithm. Finding an optimal dynamic programming solution is known to be difficult. Since the ISP's decision variables y_t are binary, our problem is NP-hard. We thus develop an approximation algorithm for (11) that efficiently prunes the search space of possible SDO schedules. Our near-optimal numerical results are given in Section V.

Algorithm 1 presents the details of the algorithm. To facilitate our discussion, we define an *outcome state* $O_{t,\vec{y}}$ at time t as the vector of estimated accrued consumption for each user and accrued revenue for the ISP, given the y_τ decisions chosen at previous times $\tau \leq t$. At each time t, we consider both

Algorithm 1 Fast Pruning Algorithm for SDO Schedules.

```
procedure ComputeSDOSchedule(a, b, σ)
                                   ▷ Columns of all matrices are 0-indexed
    usageState[0, :] ← [0]
    revenue[0, :] ← 0
    for t ← 1, . . . , T do
        for y ← 0, 1 do
            for prevY ← 0, 1 do
                currUsage ← usageState[t − 1, prevY]
                currRev ← revenue[t − 1, prevY]
                incRev ← 0
                for each u ∈ {Users} do
                    [newUserUsage, newRev] ←
estIncUsageThisHour(currUsage[u], y, a[u], b[u], σ[u, t])
                    incRev ← incRev + newRev
                    newUsage[u] ← newUserUsage
                revVsYDecision[prevY] ← incRev +
currRev
                usageVsYDecisions[prevY] ← newUsage
            if revVsYDecision[0] > revVsYDecision[1]
then
                ySchedule[t, y] ← 0
                usageState[t, y] ← usageVsYDecisions[0]
                currRev[y] ← revVsYDecision[0]
            else
                ySchedule[t, y] ← 1
                usageState[t, y] ← usageVsYDecisions[1]
                currRev[y] ← revVsYDecision[1]
    if currRev[T, 0] > currRev[T, 1] then
        return ySchedule[0]
    else
        return ySchedule[1]
```

possible ISP decision outcomes: $y_t = 0$ (do not offer SDO) and $y_t = 1$ (offer SDO). For each option, we prune among the possible SDO schedules by retaining only one outcome of the option under consideration.

At $t = 1$, we start with one initial state of no usage or revenue. We then consider decisions $y_1 = 1$ and $y_1 = 0$ with resulting outcome states $O_{1,1}$ and $O_{1,0}$. At the next time step, $t = 2$, we again consider $y_2 \in \{0, 1\}$ and end up with two outcome states for each. For example, we could move to $y_2 = 1$ from either $O_{1,1}$ (ending up in $O_{1,(1,1)}$) or from $O_{1,0}$ (ending up in $O_{1,(0,1)}$. For each choice of y_2, we pick the outcome state that has higher aggregate revenue (hence implicitly choosing the associated parent state from $t = 1$). We continue until time T, when we chose the final outcome state $O_{1,(y_1,y_2...,1)}$ or $O_{1,(y_1,y_2...,0)}$ with higher accrued revenue. By *not* pruning between the two y_t options in each time step, but instead pruning between each previous outcome state, we account for the effect of accruing outcomes between the decision branches for $y_t = 1$ and $y_t = 0$.

V. Trace-Driven Evaluation

In this section, we illustrate user and ISP decisions in our hourly model. We use a cellular usage trace from 20 users to show that ISPs gain revenue from making SDOs and that the SDO schedule computed by our pruning heuristic (Algorithm 1) is close to the optimal. We then examine the effect of the SDO price ρ. We show that *ISPs can exploit user uncertainty to earn more overage revenue as* ρ *increases* and that *ISPs experience a tradeoff between maximizing their revenue and their network utilization* in making SDO offers. We also draw comparisons between our findings and Verizon's existing PopData deployment.

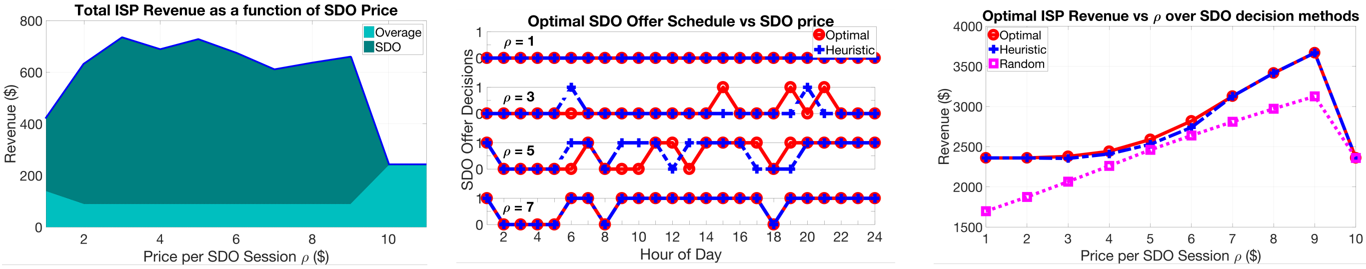
Simulation Setup. For illustration, we reduce the duration of the billing cycle to 24 hours, with an associated data overage threshold of 50MB (equivalent to a 1.5GB monthly quota). Our user-specific consumption patterns are taken from a one-week cellular usage trace of 20 users. The availability of public WiFi hotspots to users is drawn from a Rayleigh distribution with parameter 0.25 (where an availability below 0.5 is considered unavailable), as are users' price sensitivities α. We set $\theta_w = 0.5$ and draw θ_t and δ_W from a uniform distribution between 0 and 1. The ISP never offers SDO during hours 2-5 as typical network use is very low during these hours of the night; it also does not offer SDOs at 8AM and 6PM due to already high network congestion as done by Verizon with PopData [26]. These configurations apply to the following results unless noted otherwise.

SDO Schedules. Figure 3(b) compares the optimal ISP schedule for each value of ρ to the schedule generated by the fast pruning algorithm (Algorithm 1). Our pruning algorithm yields the optimal schedule when ρ is very low or high, and it closely trails the optimal schedule in other cases. When $\rho = 1$, the ISP does not offer any SDOs. Even though this SDO price is low enough to attract many users, the resulting SDO revenue does not compensate for the ISP's loss in overage revenue. As ρ increases, the ISP selectively makes SDOs in more hours. When ρ is sufficiently high, at \$7, the ISP makes an SDO in all hours, as its revenue from users' acceptance of an SDO exceeds any resulting loss in overage fees.

We next examine the revenues achieved by our pruning algorithm in Figure 3(c), with a low data overage threshold of 2MB. Our algorithm nearly achieves the revenue with the optimal schedule at all prices ρ. Both significantly improve the ISP revenue compared to a random schedule, with a 20% increase at the optimal $\rho^* = 9$, emphasizing the ISP's benefit from optimizing its SDO schedule. We next examine ISP benefits in more detail by comparing their overage and SDO revenues and considering the effect of SDOs on network utilization. These results use the optimal SDO schedule.

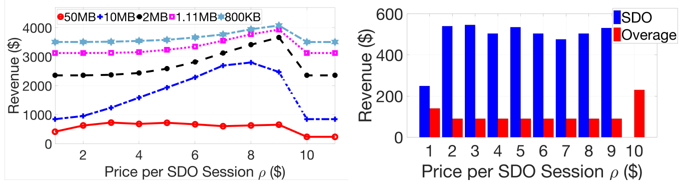
Overage vs. SDO revenue. We first examine the effect of users' overage thresholds on ISP revenue. Figure 4(a) shows that users incur more overage charges, increasing ISP revenue, as the overage threshold decreases from 50MB to 800KB. Moreover, the optimal SDO price ρ^* also increases as the ISP would discourage them from accepting SDOs and lowering its overage revenue. Hence, only higher values of ρ incentivize the ISP to offer SDOs as more users go into overage. In Verizon's PopData deployment, each PopData session costs \$2, indicating that few users would incur overage charges.

To confirm this intuition, we visualize user spending on overage and SDO fees in Figure 4(b) for an overage threshold of 1.5MB. Surprisingly, users spend more money overall under



(a) Total ISP Revenue as a function of SDO Price (b) Optimal SDO Schedule vs. SDO Price (c) ISP Revenue vs. ρ over SDO decision methods

Fig. 3: We illustrate the dependence of ISP revenue on the SDO price ρ. Our results indicate that (a) revenue from SDOs far exceeds that from overage when the ISP plans its SDO schedule optimally, (b) our heuristic SDO schedule closely matches the optimal one, with an exact match for very low or high fees ρ, and (c) our heuristic yields nearly the same revenue as the optimal SDO schedule, with significant improvement over a random schedule.



(a) Effect of Overage Threshold on Optimal ISP Revenue (b) Users' Overage vs SDO spending over SDO price

Fig. 4: As users' data quota decreases, (a) ISP revenue is maximized at higher SDO fees ρ. As ρ increases, (b) ISP's continue to make steady income from SDOs as in Figure 3(b). For each ρ, the ISP exploits user uncertainty in when SDOs will be offered, choosing its SDO schedule so as to induce users to myopically accept SDOs, even though the SDO fees incurred exceed users' future overage charges.

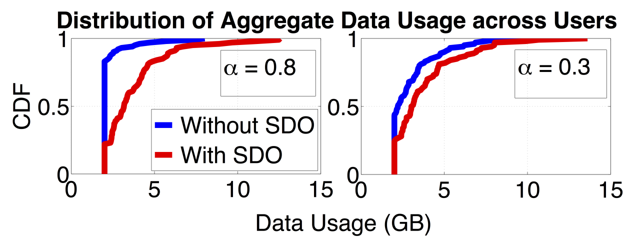


Fig. 5: Distribution of aggregate user usage across the population. The Distributions are representatives of two populations. One, with price sensitivity 0.8 and another with price sensitivity 0.3

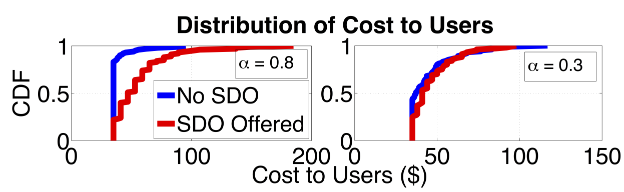


Fig. 6: Distribution of total user cost across the population. The Distributions are representatives of two populations. One, with price sensitivity 0.8 and another with price sensitivity 0.3

most regions of ρ with SDO than without. Without SDOs (at $\rho = \$10$ when no users accept SDOs), users spend approximately \$200 total on overage fees. At 20 users and \$10 for an overage, this implies 20 overages overall in the billing cycle. For the same data needs, users spend significantly more when offered SDOs. We show below that this substantial increase in revenue is not due to any significant shift from WiFi to SDOs. Instead, it is a direct consequence of users' inability to predict when future SDOs will be offered.

As users approach a new overage, i.e., R_t from (5) increases, they are more likely to accept SDOs. They do not, however, anticipate this increase in R_t in advance. As shown by users' myopic hourly utilities in (4–6), lack of information about future SDOs forces users to make bounded-rationality choices. Aversion to future overages then biases users towards accepting the SDO, allowing the ISP to plan its SDO schedule such that users' myopic decisions yield much higher revenue than the ISP could otherwise gain. While some users may avoid these charges, Figure 4(b) shows that most spend more under SDOs. If users, as in the monthly model, could plan their optimal usage up-front knowing the future SDO schedule, they could avoid these charges. If users, as in the monthly model, could plan their optimal usage up-front with the knowledge of the future SDO schedule, then they would properly balance SDO spending. In Figure 5, we show that in our monthly model, users consume more data with SDOs compared to without. Despite this increase in usage, however, they spend only slightly less with SDOs than without, indicating that they better balance their SDO spending with overage charges.

Network utilization vs. revenue. We finally examine

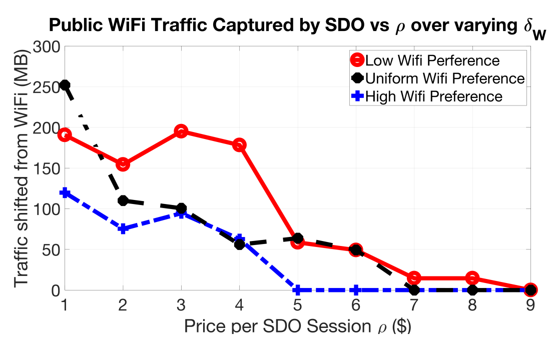


Fig. 7: The amount of public WiFi data captured by the ISP's network due to SDOs is non-monotonic in ρ, reflecting the ISP's strategic choices in computing the optimal SDO schedule.

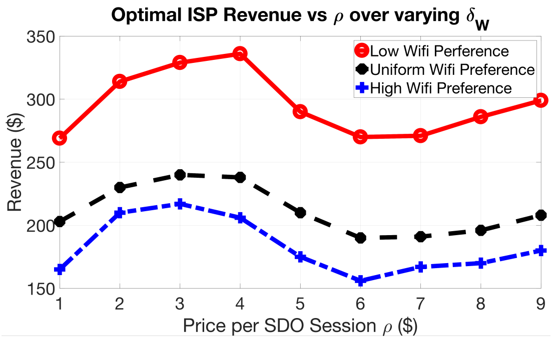


Fig. 8: The ISP can make more revenue from users with lower WiFi preferences, since these users would be more likely to accept SDOs. Comparing the revenue with the network utilization in Figure 7, the revenue maximizing ρ does not maximize network utilization.

the effect of WiFi availability on ISPs' revenue and network utilization. Though SDOs could incentivize users to consume cellular instead of WiFi data, thus allowing ISPs to monetize this otherwise "lost" usage, we find that there is a tradeoff between maximizing ISP revenue and the network utilization.

Figure 7 depicts the amount of data traffic onboarded onto the ISP's network from WiFi, as a function of ρ as well as the distribution of users' WiFi preference factor δ_W. Though the overall network utilization decreases as ρ increases, which we would expect since a higher SDO price ρ would lead to fewer users accepting SDOs instead of using WiFi, this decrease is non-monotonic. This is a direct effect of the ISP jointly optimizing ρ and the SDO schedule such that the optimal SDO offerings at each price are made strategically in hours that balance the ISP's predicted revenue from cellular onboarding and overage fees. Moreover, comparing Figures 7 and 8 shows that while network utilization is maximized at $\rho = 1$, ISP revenue is maximized at higher prices.

Our result realizes a key consequence of the dynamics of hourly SDO games. The ISP is able to learn user intentions from historical data and strategically choose the SDO schedule and price to maximize its revenue. Users are then at a disadvantage; even though they may increase their utility by switching from WiFi to SDOs, the ISP's offered SDO schedule and price does not maximize this utility increase. Thus, the ISP is able to control the information revealed about SDOs to profit from users' consequential myopic actions.

VI. Discussion and Conclusion

In this work, we analytically and empirically assess the viability of supplemental discount offers from ISPs to their users. We first abstract away from hour-to-hour dynamics to show that most users would accept some SDOs, and that those who consume the most data per SDO would also consume the most data on their cellular data plans. We then build on this framework by developing hourly decision algorithms for users to decide when to accept and ISPs to decide when to make SDOs. We simulate these algorithms over a two-week trace of data usage, empirically establishing that SDOs can increase ISPs' network utilization and revenue. Moreover, ISPs can exploit user uncertainty in when SDOs will be offered to further increase their revenue. Our work captures Verizon's claimed motivation of offering PopData in order to recover usage that would otherwise have been realized on WiFi networks [27], and indeed we find a tradeoff between the ISP maximizing its revenue and its network utilization.

Throughout this work, we assume a monopoly ISP. Though we do consider users' option to consume WiFi instead of cellular data, we do not model competition between ISPs. In a competitive setting, SDOs may attract new users to an ISP by allowing them to supplement their data plans; on the other hand, other ISPs could counter these offers by simply increasing their plans' monthly quotas. Thus, future work should consider these potential ISP competition effects, as well as users' ability to predict when SDOs will occur in the future. With such predictions, users could further optimize their SDO acceptances and undercut ISPs' SDO revenue. Our work can be viewed as a first step towards assessing SDOs' benefits and guiding users and ISPs as to how to use and react to them.

References

[1] Cisco, "Cisco Visual Networking Index: Global mobile data traffic forecast update, 2016–2021 white paper," 2017. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[2] Sprint, "Unlimited cell phone plan," 2017. https://www.sprint.com/en/shop/plans/unlimited-cell-phone-plan.html.

[3] AT&T, "Unlimited data plans," 2017. https://www.att.com/plans/unlimited-data-plans.html.

[4] Orange, "EE monthly price plans," 2017. http://ee.co.uk/content/dam/ee-help/help-pdfs/ee-pay-monthly-price-plans-may-2017.pdf.

[5] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl, "Anatomizing application performance differences on smartphones," in *Proc. of ACM MobiSys*, pp. 165–178, ACM, 2010.

[6] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: A survey of time-dependent pricing for internet access," *IEEE Communications Magazine*, vol. 50, no. 11, 2012.

[7] H. Park, Y. Jin, J. Yoon, and Y. Yi, "On the economic effects of user-oriented delayed wi-fi offloading," *IEEE Trans. on Wireless Communications*, vol. 15, no. 4, pp. 2684–2697, 2016.

[8] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "Tube: Time-dependent pricing for mobile data," *Proc. of ACM SIGCOMM*, 2012.

[9] S. Sen, C. Joe-Wong, S. Ha, J. Bawa, and M. Chiang, "When the price is right: enabling time-dependent pricing of broadband data," in *Proc. of ACM SIGCHI*, pp. 2477–2486, ACM, 2013.

[10] A. Odlyzko, "Will smart pricing finally take off?," in *Smart Data Pricing* (S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, eds.), pp. 1–33, Wiley Online Library.

[11] Verizon Wireless, "PopData FAQs," 2017. https://www.verizonwireless.com/support/popdata-faqs/.

[12] M. El-Sayed, A. Mukhopadhyay, C. Urrutia-Valdés, and Z. J. Zhao, "Mobile data explosion: Monetizing the opportunity through dynamic policies and qos pipes," *Bell Labs Technical Journal*, vol. 16, no. 2, pp. 79–99, 2011.

[13] J. M. Dyaberi, B. Parsons, V. S. Pai, K. Kannan, Y.-F. R. Chen, R. Jana, D. Stern, and A. Varshavsky, "Managing cellular congestion using incentives," *IEEE Communications Magazine*, vol. 50, no. 11, 2012.

[14] L. Zhang, W. Wu, and D. Wang, "Time dependent pricing in wireless data networks: Flat-rate vs. usage-based schemes," in *Proc. of IEEE INFOCOM*, pp. 700–708, IEEE, 2014.

[15] C. Zhang, B. Gu, S. Xu, K. Yamoriy, and Y. Tanaka, "Time-dependent pricing for revenue maximization of network service providers considering users preference," in *Proc. of IEEE APNOMS*, pp. 1–6, 2013.

[16] L. Jiang, S. Parekh, and J. Walrand, "Time-dependent network pricing and bandwidth trading," in *Proc. of IEEE Network Operations and Management Symposium Workshops*, pp. 193–200, 2008.

[17] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE transactions on industrial informatics*, vol. 7, no. 3, pp. 381–388, 2011.
[18] Q. Ma, Y.-F. Liu, and J. Huang, "Time and location aware mobile data pricing," *IEEE Trans. on Mobile Computing*, vol. 15, no. 10, pp. 2599–2613, 2016.
[19] P. Loiseau, G. Schwartz, J. Musacchio, S. Amin, and S. S. Sastry, "Incentive mechanisms for internet congestion management: Fixed-budget rebate versus time-of-day pricing," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 647–661, 2014.
[20] W. Dong, S. Rallapalli, R. Jana, L. Qiu, K. Ramakrishnan, L. Razoumov, Y. Zhang, and T. W. Cho, "ideal: Incentivized dynamic cellular offloading via auctions," *IEEE/ACM Transactions on Networking*, vol. 22, no. 4, pp. 1271–1284, 2014.
[21] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "A double-auction mechanism for mobile data-offloading markets," *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1634–1647, 2015.
[22] M. Andrews, G. Bruns, M. Doğru, and H. Lee, "Understanding quota dynamics in wireless networks," *ACM Transactions on Internet Technology*, vol. 14, no. 2-3, p. 14, 2014.
[23] C. Joe-Wong, S. Ha, and M. Chiang, "Sponsoring mobile data: An economic analysis of the impact on users and content providers," in *Proc. of IEEE INFOCOM*, pp. 1499–1507, IEEE, 2015.
[24] L. Zheng, C. Joe-Wong, J. Chen, C. G. Brinton, C. W. Tan, and M. Chiang, "Economic viability of a virtual ISP," in *Proc. of IEEE INFOCOM*, 2017.
[25] Verizon Wireless, "Verizon plan faqs," 2017. https://www.verizonwireless.com/support/the-verizon-plan-faqs/.
[26] M. Sorrentino, "Verizon may give you unlimited data, up to an hour at a time." CNet, 2017. https://www.cnet.com/news/verizon-popdata-let-you-have-unlimited-data-but-only-for-an-hour/.
[27] Kellen, "Verizon introduces popdata, 30 and 60 minute all-you-can-stream 4g lte sessions at $2 or $3," 2016. http://www.droid-life.com/2016/10/13/verizon-popdata-cost-available/.

APPENDIX

A. Proof of Lemma 1

Proof. In the event that $x \leq d$, the utility $u(x,\beta)$ can be written as

$$u(x,\beta) = C_1 \frac{x^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] - \eta, \tag{13}$$

for $x \leq d$, where u is strictly increasing in x. Hence, the user utility is maximized at $x^* = d$, at which point the user always consumes the data quota, as it is already paid for. Considering the case where $x \geq d$, the user utility expression is

$$u(x,\beta) = C_1 \frac{x^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] - \eta - p(x-d), \tag{14}$$

for $x \geq d$. In this region, $u(x,\beta)$ is convex and hence a maxima exists. However, this maxima is the optimal x for (14) only if it lies beyond d. Else, the function is strictly decreasing is this region and the optimal x is simply d. Assuming the maxima is beyond d, we find the utility-maximizing x by equating the derivative of the utility function to 0, yielding

$$\begin{aligned} \frac{\partial u}{\partial x} &= 0 \\ C_1 x^{-\alpha} - p &= 0 \\ x^* &= \left(\frac{C_1}{p}\right)^{\frac{1}{\alpha}} \end{aligned}$$

The optimal value obtained, $x^* = (C_1/p)^{1/\alpha}$, is subject to two constraints: it is lower bounded by d and upper bounded by $\gamma - \beta n x_{max} \geq d$ due to (2). By considering these bounds, we obtain the desired result. □

B. Proof of Proposition 1

Proof. We separately consider two cases. If $(C_1/p)^{1/\alpha} \leq d$, then $x^* = d$, and the user utility as a function of β is:

$$u(\beta) = C_1 \frac{d^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] - \eta, \tag{15}$$

for $d \geq (C_1/p)^{1/\alpha}$.

From (15), we see that the utility function is linear in β. If $C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho$ is negative, (15) decreases with β indicating that the satisfaction obtained from using PopData is less than the cost of PopData, and hence $\beta^* = 0$. A positive co-efficient for β, however, implies that the user utility increases linearly in β, and hence the difference between γ and x (which is, by definition, d in this region) in this region is accommodated by PopData. The optimal β in this region is hence:

$$\beta^* = \frac{\gamma - d}{n x_{max}} H \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right], \tag{16}$$

where H denotes the *unit step function* that equals one if the argument is greater than 0, and 0 otherwise.

We now consider the second case in which $(C_1/p)^{1/\alpha} > d$, for which $u(\beta)$ is

$$\begin{aligned} u(\beta) =& C_1 \frac{((\frac{C_1}{p})^{\frac{1}{\alpha}})^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] \\ & - \eta - p((\frac{C_1}{p})^{\frac{1}{\alpha}} - d), \end{aligned} \tag{17}$$

for $d \leq (C_1/p)^{1/\alpha} \leq \gamma - \beta n x_{max}$. As in the first case, we see that the optimal β is either 0 or the upper-bound from the constraint in (2),

$$\beta^* = \frac{\gamma - (\frac{C_1}{p})^{\frac{1}{\alpha}}}{n x_{max}} H[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho] \tag{18}$$

Finally, we jointly optimize β and x over the remaining region. If $(C_1/p)^{1/\alpha} \geq \gamma - \beta n x_{max}$, $u(\beta)$ is given by

$$\begin{aligned} u(\beta) =& C_1 \frac{(\gamma - \beta n x_{max})^{1-\alpha}}{1-\alpha} + \beta n \left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho \right] \\ & - \eta - p(\gamma - \beta n x_{max} - d), \end{aligned} \tag{19}$$

for $d < \gamma - \beta n x_{max} \leq (C_1/p)^{1/\alpha}$.

Upon equating the derivative of (19) to 0, we have:

$$\begin{aligned}
\frac{\partial u}{\partial \beta} &= 0 \\
\frac{C_1 n x_{max}}{(\gamma - \beta n x_{max})^\alpha} &= n\left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho\right] + pnx_{max} \\
\gamma - \beta n x_{max} &= \left(\frac{C_1 x_{max}}{\left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho\right] + px_{max}}\right)^{1/\alpha} \\
\beta^* &= \frac{\gamma - \left(\frac{C_1 x_{max}}{\left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho\right] + px_{max}}\right)^{1/\alpha}}{nx_max} \qquad (20) \\
x^* &= \left(\frac{C_1 x_{max}}{\left[C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho\right] + px_{max}}\right)^{1/\alpha} \qquad (21)
\end{aligned}$$

We now note that Region 2 is a special case of Region 3 when PopData has negative utility. To see this, substitute $[C_2 x_{max}^{1-\alpha}/(1-\alpha) - \rho] = 0$ in (20) and (21), thus setting utility from PopData to 0. Then, x^* and β^* take the values of x^* and β^* for Region 2. However, if the utility of PopData is 0, H in (18) would put β^* as 0, which is not the case as seen. Thus Region 2, in fact, does not apply when the Utility from PopData is non-negative, in which case, Region 3 accounts for the values of x^* and β^*. However, when the utility from PopData is negative, i.e, $[C_2 x_{max}^{1-\alpha}/(1-\alpha) - \rho] < 0$, then (18) correctly results in zero PopData usage and optimal x^* of $(C_1/p)^{1/\alpha}$.

We note as well that, by definition of $(C_1/p)^{1/\alpha}$ in Region 3, utility from PopData cannot be negative in Region 3. That is, if $[C_2 x_{max}^{1-\alpha}/(1-\alpha) - \rho] < 0$ in Region 3, then the optimal x^* given by (21) exceeds $(C_1/p)^{1/\alpha}$, in which case that x^* is infeasible as it violates usage constraint (2). This means that if utility from PopData is negative and $(C_1/p)^{1/\alpha} > d$ (i.e, we are not in Region 1), then the user must necessarily be in Region 2. On the other hand, if the utility from PopData is greater than 0 and $(C_1/p)^{1/\alpha} > d$ (i.e, we are not in Region 1), then the user must necessarily be in Region 3. These conditions yield the final result given in Table II. □

C. Proof of Corollary 2

Proof. Under the stated conditions, users' data plan usage is given by

$$x^* = \left(\frac{C_1 x_{max}}{C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho + px_{max}}\right)^{\frac{1}{\alpha}}.$$

Thus, it suffices to show that $x^{*\alpha}$ reaches its minimum value at $x_{max} = (\rho(1-1/\alpha))^{1/(1-\alpha)}$. We do so by taking the first derivative and setting it equal to zero, which is equivalent to

$$C_1\left(C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho + px_{max}\right) = C_1 x_{max}\left(p + C_2 x_{max}^{-\alpha}\right)$$

$$\frac{C_2 \alpha}{1-\alpha} x_{max}^{1-\alpha} = \rho,$$

from which the result follows directly. □

D. Proof of Proposition 2

Proof. Corollary 2 shows that x^* increases as x_{max} increases, for x_{max} above the given threshold, regardless of the value of γ. To show that β^* increases with γ, we consider two cases. First, if $(C_1/p)^{1/\alpha} \leq d$, then

$$\beta^* = \frac{\gamma}{n} - \frac{d}{nx_{max}},$$

which is increasing in x_{max} by inspection. Second, if $(C_1/p)^{1/\alpha} > d$, then we find that

$$\beta^* = \frac{\gamma}{n} - \frac{1}{n}\left(\frac{C_1 x_{max}^{1-\alpha}}{C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho + px_{max}}\right)^{\frac{1}{\alpha}}$$

thus, it suffices to show that

$$\frac{d}{dx_{max}}\left(\frac{C_1 x_{max}^{1-\alpha}}{C_2 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho + px_{max}}\right) < 0.$$

Taking this derivative, we find that it is proportional to

$$\begin{aligned}
&\left(C_1 \frac{x_{max}^{1-\alpha}}{1-\alpha} - \rho + px_{max}\right)(1-\alpha)C_1 x_{max}^{-\alpha} \\
&- C_1 x_{max}^{1-\alpha}\left(C_1 x_{max}^{1-\alpha} + p\right) \\
&= -pC_1 x_{max}^{-\alpha} - \alpha p C_1 x_{max}^{1-\alpha}
\end{aligned}$$

which is negative by inspection. □

E. Online Estimation of User Parameters

While the ISP calculates optimal SDO schedule at $t = 0$, it strategically does not reveal this to the users, hence gaining the advantage (amongst others detailed in Section V) to observe users' accrued consumption in the current billing cycle and measure any significant deviations from the learnt a, b and σ_T^2. This deviation from typical historic trends could be especially considerable when the ISP first introduces SDOs, as offloading to SDOs impacts the usage trend under the regular data plan. To accommodate such externalities, the ISP might use the following *online learning* procedure to recompute user characteristics and subsequently the SDO schedule for leftover timesteps.

Update Criteria. The ISP can periodically calculate the likelihood of the observed x_cs over the duration of the billing cycle and determine whether the user's consumption trend in the current month is in keeping with the learnt model. Given a vector of observed $\vec{x_c}$ and corresponding time-intervals $\vec{t}$, the update criteria is defined as:

$$p(\mathbf{x_c}(\vec{\mathbf{t}})|\mu(\vec{\mathbf{t}}), \mathbf{\Sigma}(\vec{\mathbf{t}})) = \prod_{i=1}^{t} \frac{1}{\sqrt{(2\pi)^t|\Sigma|}} \qquad (22)$$

$$exp\frac{-(x_c(i) - \mu(\vec{\mathbf{t}}))'\mathbf{\Sigma^{-1}}(x_c(i) - \mu(\vec{\mathbf{t}}))}{2} \qquad (23)$$

$$u(\mathbf{x_c}(\vec{\mathbf{t}})) = \begin{cases} 1, & \text{if } p(\mathbf{x_c}(\vec{\mathbf{t}})|a(t)^b, \sigma_t^2) \geq T_U \\ 0, & \text{otherwise} \end{cases} \qquad (24)$$

where u is the update decision, and T_U is a pre-defined empirical threshold for the likelihood of observations $\mathbf{x_c}(\vec{\mathbf{t}})$, below which the user is determined to be significantly deviant from their expected trend.

Update Algorithm. If the update decision u is affirmative, the ISP can use *weighted Maximum Likelihood* estimation to recalculate the learned parameters, where the observations of the current cycle $\vec{\mathbf{x_c}}$ are assigned a weight inversely proportional to the likelihood of the observations, and the rest of the historic observations are weighed equally. *i.e.,*

$$W_i(S) = \begin{cases} \frac{(1-p(\vec{\mathbf{x_c}(\mathbf{t})}|a(t)^b,\sigma_t^2))}{|S|}, & \text{if } S = \vec{\mathbf{x_c}(\mathbf{t})} \\ \frac{p(\vec{\mathbf{x_c}(\mathbf{t})}|a(t)^b,\sigma_t^2)}{|S|}, & \text{otherwise} \end{cases} \tag{25}$$