Optimizing Data Plans: Usage Dynamics in Mobile Data Networks

Liang Zheng*, Carlee Joe-Wong†, Matthew Andrews‡, and Mung Chiang*§
*Princeton University, †Carnegie Mellon University, ‡Nokia Bell Labs, §Purdue University
*liangz@princeton.edu, †cjoewong@andrew.cmu.edu,‡matthew.andrews@nokia-bell-labs.com, §chiang@purdue.edu

***Abstract*—As the U.S. mobile data market matures, Internet service providers (ISPs) generally charge their users with some variation on a quota-based data plan with overage charges. Common variants include unlimited, prepaid, and usage-based data plans. However, despite a recent flurry of research on optimizing mobile data pricing, few works have considered how these data plans affect users' consumption behavior. In particular, while users with such plans have a strong incentive to plan their usage over the month, they also face uncertainty in their future data usage needs that would make such planning difficult. In this work, we develop a dynamic programming model of users' consumption decisions over the month that takes this uncertainty into account. We use this model to quantify which types of users would benefit from different types of data plans, using these conditions to extrapolate the optimal types of data plans that ISPs should offer. Our theoretical findings are complemented by numerical simulations on a dataset of user usage from a large U.S. ISP. The results help mobile users to choose data plans that maximize their utilities and ISPs to gain profit by understanding their user behavior while choosing what data plans to offer.**

I. Introduction

As the U.S. market for smartphones begins to saturate, with more than 80% penetration [1], Internet service providers (ISPs) appear to have converged on the types of mobile data plans that they offer. In particular, most data plans offered by major ISPs now enforce some version of a monthly data quota with overage charges for users who go over the quota. Common variants on this plan include unlimited data plans, in which users pay a flat fee for an unlimited amount of data usage, possibly with throttling after their usage reaches a monthly limit [2]; prepaid data plans, in which users do not have overage options but are cut off after reaching the quota; and usage-based plans in which users pay in proportion to their usage amounts. Yet, users are being offered more freedom to switch between data plans: for instance, two-year contracts were recently eliminated in the U.S. [3], with new ISPs joining the market [4], [5]. These developments create a need to understand how users should choose their data plans, and which plans ISPs should offer so as to make a profit.

Some qualitative insights into users' data plan choices are straightforward–for instance, while an unlimited plan may be attractive to heavy users with significant usage needs, lighter users might be better served by a usage-based data plan that allows them to pay only for the data they consume. It is not clear, however, how quota with overage plans fit into this picture, nor do such insights quantify the exact conditions under which users would prefer a certain type of data plan. Answering this question would have implications not just for users, but also for ISPs that aim to determine the specific combination of plans to offer that can attract more users.

The presence of monthly quotas in many mobile data plans makes determining users' utilities from these data plans particularly difficult: in general, *a user does not know in advance exactly what her data needs will be during an upcoming billing cycle.* As a result she will consume data throughout the month in a way that reflects this uncertainty, which will impact her resulting utility. Thus, determining users' utilities from different data plans requires developing a model of *quota dynamics* during the billing period in which the user makes usage decisions today solely based on distributional information regarding future needs for data.

In this work, we consider the problem of modeling user usage decisions based on the pricing of a mobile data plan and the utility that end users achieve by subscribing to it. At a high level, our results can be viewed as quantifying how much value a user assigns to the constraint imposed by a given data cap. This insight can in turn be used by an ISP to develop an optimal (i.e., profit-maximizing) set of data plans for its population of users; we can then evaluate whether users and ISPs would ever both prefer the same type of data plans. Our research contributions in these directions are as follows:

Models of user utility that capture users' usage decisions over the month (Section III): Though any model of user behavior is necessarily stylized, we can thus predict the actions of a user who may try to "save up" her quota for days in which her usage is more important. To do so, we suppose that a user has a weighted utility function on both her evaluation of the data consumption and cost on the data usage in each day. We account for uncertainty in how this weight can change from day to day by developing deterministic and stochastic models.

In Section III-A, we first consider the deterministic case in which her utility for each day is known in advance. For this case, we can obtain a solution in closed form, allowing us to rigorously derive qualitative insights into user utilities and ISP profit. In Section III-B we turn to the more realistic case in which the weights for the utility function are drawn from a distribution. We consider users with estimated utility making

The work in this paper was in part supported by the NSF under Grant No. CNS-1347234, the ONR under Grant No. N68335-17-C-0207, and DARPA DCOMP Program under Contract No. HR001117C0048.

decisions for the current day according to an estimate of how much utility can be gained in the future. The optimization for the stochastic case is complex and so we present a detailed analysis on a number of special cases in Section III-C, focusing particularly on comparing usage decisions in the deterministic and stochastic cases.

Quantification of user utility, offering implications of the optimal types of data plans that ISPs should offer (Section IV): We use our usage dynamics findings to find the optimal set of pricing plans offered by an ISP, given users' estimated usage under each type of plan. In doing so, we provide conditions under which users should choose each type of offered data plan, so as to maximize their utilities. We quantify the gap between the plans that maximize an ISP's revenue and those that maximize users' utilities, and show that under some conditions the same type of data plan maximizes both ISP revenue and user utility.

Analysis on a real-world dataset (Section V): To examine the findings from our usage dynamics models, we use a one-month daily usage trace from a U.S. ISP to infer the parameters that characterize each user's usage behavior. By estimating users' utilities under different data plans, we can infer their optimal data plan choices. We also show that our results can be used by ISPs to decide whether to offer a new data plan: on this dataset, the ISP should not offer a usage-based plan.

We present our conclusions in Section VI and match our findings to trends in the mobile data market, demonstrating the explanatory power of our model and suggesting possible future directions for mobile data plans. All proofs can be found in the appendix.

Our results can be applied in different ways. First, they allow both a user and an ISP to determine how the user will consume data over the billing cycle and also calculate the most appropriate data plan for the user. Second, our results could be utilized in the design of a smartphone app that makes recommendations to users regarding which data plan they should choose. After the plan is chosen the app can then make recommendations as to how much data the user should consume in the current day so as to maximize her eventual utility. Lastly, using our methodology an ISP can examine usage logs from their user population and use the results to optimize the menu of data plans that it offers to its customers.

II. Related Work

Previous works have studied the benefits of many variants on today's data plans that are currently deployed in a limited context. For instance, some ISPs in the U.S. offer sponsored data plans, where a third party content provider such as Netflix or Hulu can subsidize users' data consumption. Although users need not pay for this traffic, the ISP can still gain revenue by charging content providers instead [6], [7]. One can also consider the temporal impact of sponsored data, e.g., if content providers can sponsor different amounts of data at different times of the day, thus even-ing out network traffic [8]. Other works have examined the impact of data plans that combine access to multiple cellular and WiFi networks, as Google's Project Fi [4] does. An economic analysis shows that such plans can increase user utility and ISP profit [9], [10].

Little research, however, has been done on understanding quota-based data plans, which remain the dominant type of data pricing offered by U.S. ISPs today. Some works have focused on rollover [11] or shared [12] aspects of such plans, and others have considered the adoption of multiple service plans for various user requirements [13], [14]. Still other works have taken an empirical view: [15] models user dynamics for calls and text messages by fitting models to a large dataset, while [16] shows the influence of pricing on user usage and ISP profit. Other studies have shown that users do in fact make decisions based on their data quotas [17]. However, no work to date has modeled and explored users' data plan choices in the context of the dynamics of user usage throughout the month in the presence of data quotas and overage charges. We address this gap in our work and consider the effects on user usage, network traffic, and ISP profit.

III. Modeling Usage Dynamics

To model the types of data plans commonly offered by ISPs, we consider a monthly billing cycle with D days. The ISP charges users a flat fee P for a monthly quota A MB of data,[1] i.e., the data allowance within a month, with an overage fee π \$/MB for any data usage exceeding this monthly quota. The combination (A, P, π) defines an ISP's pricing policy; for instance, $(A = +\infty, P > 0, \pi = 0)$ represents an unlimited data plan, while data plans with $(A = 0, P = 0, \pi > 0)$ adopt usage-based pricing to charge users.

Given a pricing policy, we consider a utility-based model for a set $\mathcal{N}$ of N users in the mobile network who need to decide the amounts of data to consume in each day within a billing cycle that maximize their utilities from the current day until the end of the billing cycle. We define the *state* of each user on each day of the cycle in terms of the amount of data she has left in her quota and the number of days remaining in the billing cycle. We work backwards from the end of the billing cycle: at the beginning of the dth to last day of the current billing cycle, i.e., after $D-d$ days from the beginning of the billing cycle, user i has an amount $q_{i,d}$ MB of leftover data, leading to the state $(q_{i,d}, d)$.

To model the dynamics of users' state variables, we let $a_{i,d}$ denote the amount of data that user i consumes on the dth to last day of her billing cycle, so that $a_{i,d} = q_{i,d} - q_{i,d-1}$ and $q_{i,d} = A - \sum_{t=d+1}^{D} a_{i,t}$. We note that $q_{i,d} \leq 0$ means the user has used up her data quota and has consumed $-q_{i,d}$ overage data by the $(D-d)$th day. Before the billing cycle is over, each user i decides how much data to consume $(a_{i,d})$ based on her current state $(q_{i,d}, d)$. We use $\vec{\mathbf{a}}_i = \{a_{i,t}\}_{t=1}^{D}$ to denote a decision vector for each user i.

We suppose that users choose their usage so as to maximize their overall utility over a month. We model the usage utility from consuming $a_{i,d} \geq 0$ amount of data in the $(D-d)$th day using a nondecreasing and concave α-fairness function

[1] We assume that ISPs measures usage using 1MB minimum units.

$a_{i,d}^{1-\alpha_i}/(1-\alpha_i)$ with $\alpha_i \in [0,1)$ [7], [9].[2] Users can incur a monetary cost from consuming data: if the user consumes more than her leftover data in the day, i.e., $a_{i,d} > q_{i,d}$, she will need to pay the amount of $\pi(a_{i,d}-q_{i,d})$ overage fee to the ISP. In addition, we suppose that users incur a *time* cost from data consumption, as using mobile data generally requires them to pay attention to their smartphones. This cost can be thought of as an opportunity cost of using data instead of spending time on other activities. (Without this cost then a user would always consume their entire quota which is not realistic.) We suppose that the amount of time that a user invests in using data is proportional to the amount of data consumed, yielding a cost of $\phi_i a_{i,d}$, where $\phi_i > 0$ is a linear scale of user i's time value on consuming 1MB of data.[3] From the above discussion, we can find user i's utility function $v_{i,d}(a_{i,d})$ on the dth to last day:[4]

$$v_{i,d}=\begin{cases}\dfrac{\omega_{i,d}a_{i,d}^{1-\alpha_i}}{1-\alpha_i}-\phi_i a_{i,d},\, 0\le a_{i,d}\le q_{i,d}, & \text{(1a)}\\[2mm] \dfrac{\omega_{i,d}a_{i,d}^{1-\alpha_i}}{1-\alpha_i}-\pi(a_{i,d}-q_{i,d})-\phi_i a_{i,d},\, 0\le q_{i,d}< a_{i,d}, & \text{(1b)}\\[2mm] \dfrac{\omega_{i,d}a_{i,d}^{1-\alpha_i}}{1-\alpha_i}-\pi a_{i,d}-\phi_i a_{i,d},\, q_{i,d}<0\le a_{i,d}, & \text{(1c)}\end{cases}$$

where the utility weights $\omega_{i,d} > 0$ encode the relative importance of usage utility for different users on different days. In (1), α_i can also be viewed as an indicator of user i's *cost sensitivity*: cost-sensitive users with larger α_i would generally consume less data due to receiving a lower marginal utility from their data usage, compared to users with a smaller α_i. Formally, we note that (1a) models the case of user i having leftover data, while in (1b) and (1c), the user consumes overage data: (1b) models the case in which users begin incurring overage charges on the $(D-d)$th day, and (1c) models the case in which users have already exceeded their quotas before the $(D-d)$th day.

In the sections below, we find the usage patterns $a^\star_{i,d}$ that maximize (1) over the month. We first suppose that the weights $\omega_{i,d}$ are *deterministic* (Section III-A), i.e., that they are fixed and the user knows them in advance. This case would correspond to users who are good at planning their usage in advance, or whose usage is predictably consistent from month to month. In the more realistic *stochastic* model (which we consider in Section III-B) the weights $\omega_{i,d}$ for each day are drawn from a known distribution, but the realized values are not known to the user until the beginning of the day.

A. User Usage Dynamics - Deterministic Model

In the deterministic case, each user decides her usage on each day by maximizing the sum of utilities across the month subject to the sequence of state transitions:

$$\begin{aligned}\underset{\{a_{i,t}\}_{t=1}^D}{\text{maximize}} \quad & \textstyle\sum_{t=1}^D v_{i,t}(a_{i,t})\\ \text{subject to} \quad & q_{i,t-1}=q_{i,t}-a_{i,t},\ t=D,\ldots,1.\end{aligned} \tag{2}$$

Letting $V(q_{i,d},d)$ be the maximum utility that user i can achieve from day d to the end of the billing cycle if she has $q_{i,d}$ amount of leftover data at the beginning of the day, a transition from $(q_{i,d},d)$ to $(q_{i,d-1},d-1)$ brings user i a utility of $v_{i,d}(a_{i,d})$ as given in (1). By applying the principle of dynamic programming, (2) is equivalent to the following Bellman equations [15], [18]:

$$V^\star(q_{i,d},d)=\max_{a_{i,d}}\big\{v_{i,d}(a_{i,d})+V^\star(q_{i,d-1},d-1)\big\}, \tag{3}$$

with $q_{i,D}=A$. This model captures the dependency of $a_{i,d}$ on this user's current leftover and future usage for the remaining d days in this billing cycle. Intuitively, a user would wish to choose how much data to use by estimating the maximum utility that she would achieve in the future. We can find the analytical form for $V^\star(q_{i,d},d)$ by applying mathematical induction to (3), leading us to a user's optimal data usage:

Proposition 1: When all $\omega_{i,t}$ are deterministically known in advance, user i's optimal usage by solving (2) is given by:

$$a^\star_{i,d}=\begin{cases}(\omega_{i,d}/\phi_i)^{\frac{1}{\alpha_i}},\ q_{i,d}>\sum\limits_{t=1}^{d}(\omega_{i,t}/\phi_i)^{\frac{1}{\alpha_i}}, & \text{(4a)}\\[2mm] \big(\omega_{i,d}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}},\ q_{i,d}<\sum\limits_{t=1}^{d}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}}, & \text{(4b)}\\[2mm] q_{i,d}\Big/\sum\limits_{t=1}^{d}\Big(\dfrac{\omega_{i,t}}{\omega_{i,d}}\Big)^{\frac{1}{\alpha_i}},\ \text{otherwise}. & \text{(4c)}\end{cases}$$

We can build on Proposition 1's results to write users' decisions in terms of parameters known at the beginning of the month, allowing users to plan their optimal usage in advance:

Corollary 1: When all $\omega_{i,t}$ are deterministically known at the start of the month, user i's optimal usage on the dth day counting from last day of the billing cycle, i.e., $a^\star_{i,d}$ in (4), can be written in terms of A and π:

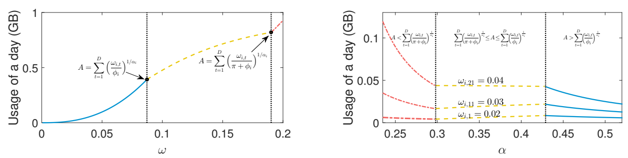
$$a^\star_{i,d}=\begin{cases}(\omega_{i,d}/\phi_i)^{\frac{1}{\alpha_i}},\ A>\sum\limits_{t=1}^{D}(\omega_{i,t}/\phi_i)^{\frac{1}{\alpha_i}}, & \text{(5a)}\\[2mm] \big(\omega_{i,d}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}},\ A<\sum\limits_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}}, & \text{(5b)}\\[2mm] \omega_{i,d}^{\frac{1}{\alpha_i}}A\Big/\sum\limits_{t=1}^{D}\omega_{i,t}^{\frac{1}{\alpha_i}},\ \text{otherwise}. & \text{(5c)}\end{cases}$$

Based on (5), Figure 1 illustrates user i's usage in the dth day counting from the last day of the billing cycle with varying $\omega_{i,d}$ and α_i. In Figure 1(a), we fix all $\omega_{i,t}$ for $t \neq d$ and calculate $a^\star_{i,d}$ for different $\omega_{i,d}$; and in Figure 1(b), we fix the $\omega_{i,t}$ at different values and plot the usage in terms of α_i for large, medium, and small $\omega_{i,d}$, respectively. We see that (5) is continuous but not continuously differentiable, with breaks

[2] As a special case, the usage utility function is $\log a_{i,d}$ when $\alpha_i = 1$.

[3] This term is meant to approximate users' overall time costs over a day, assuming that the time spent consuming data would be proportional to the amount of data usage. To model different users' different opportunity costs for their time, we allow ϕ_i to be user-specific.

[4] Since P is a fixed fee that users pay for the quota upfront and will not affect their usage, we need not consider it in maximizing users' utility.



(a) Increasing user usage with $\omega_{i,d}$. (b) Monotonic user usage with α_i.

Fig. 1. User usage dynamics for the deterministic model with varying $\omega_{i,d}$ and α_i in a billing cycle of $D = 30$ days, where the user is offered $A = 1000$MB data and any usage above that are charged at $\pi = 0.005$\$/MB. We suppose this user's time cost is $\phi_i = 0.008$\$/MB. (a) $\alpha = 0.4$, $\omega_{i,t} = 0.027, \forall t \neq d$; (b) $\{\omega_{i,d}\}_{d=1}^{D} = \{0.02, 0.021, \ldots, 0.049\}, \forall t$.

divided by the dotted lines in Figure 1 corresponding to the three usage cases below:

1) Data leftover, i.e., $A > \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$ in (5a)*:* These **light users** can fulfill their demands without incurring overage charges, as shown by the blue solid curves in Figure 1. This case corresponds to users that are more sensitive to their time costs, i.e., small $\omega_{i,t}$ and large α_i. If the ISP offers unlimited data plans with $A \to +\infty$, all users would fall under this case, consuming $(\omega_{i,t}/\phi_i)^{1/\alpha_i}$ of usage per day.

2) Data overage, i.e., $A < \sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i}$ in (5b)*:* These **heavy users** would reduce their daily usage from $\big(\omega_{i,t}/\pi\big)^{1/\alpha_i}$ to $\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i}$ due to overage costs, but still incur an overage fee due to small quotas A. As for light users, their usage increases with $\omega_{i,d}$ but decreases with α_i (the dash-dot red curves in Figure 1). If the ISP offers usage-based plans with $A = 0$, all users would fall into this case.

3) Data depletion, i.e., $\sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i} \leq A \leq \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$ in (5c)*:* These **moderate users** would incur overage charges if they consumed the maximum usage amounts $(\omega_{i,t}/\phi_i)^{1/\alpha_i}$ as in the data leftover case. However, unlike the data overage case, their desired usage is small enough ($\sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i} \leq A$) that they would prefer to limit their usage to the quota. Thus, they proportionally allocate their daily usage according to the factor $\omega_{i,d}^{1/\alpha_i}$, normalized such that they would use up their data quotas by the end of the billing cycle. Fixing α_i and $\omega_{i,t}, \forall t \neq d$, the user's optimal usage increases modestly with $\omega_{i,d}$, and can either increase or decrease with α_i, as shown by the dashed yellow curves in Figure 1.

Although we have derived the optimal decisions on user usage with $\omega_{i,d}$ known in advance, most users are not aware of how they will use the data in the future. We use a stochastic model to capture this uncertainty.

B. User Usage Dynamics - Stochastic Models

If the $\omega_{i,t}$ are not known in advance, we model their decisions as a stochastic process with weights $\omega_{i,t}$ drawn from known distribution. In Section VI, we will discuss how an ISP can estimate such a distribution. We first present the general case and then illustrate our findings with the tractable special case of prepaid data plans.

The user's objective in choosing her daily usage is to maximize her expected sum of utilities over the month:

$$\begin{aligned} &\underset{\{\tilde{a}_{i,t}\}_{t=1}^{D}}{\text{maximize}} && \textstyle\sum_{t=1}^{D}\mathbb{E}_t\left(v_{i,t}(\tilde{a}_{i,t})\right) \\ &\text{subject to} && \tilde{q}_{i,t-1} = \tilde{q}_{i,t} - \tilde{a}_{i,t},\ t = D, \ldots, 1. \end{aligned} \tag{6}$$

where we suppose that $\omega_{i,d}$ for each user i on the last dth day is independently drawn from a known distribution with probability density function (PDF) $f_{i,d}(\omega)$,[5] and $\mathbb{E}_t(\cdot)$ calculates the expectation over the stochastic $\omega_{i,d}$ for future days d, i.e., $t \geq d$. We use $\tilde{a}_{i,d}$ and $\tilde{q}_{i,d}$, respectively, to denote user usage and leftover quota under uncertainty, as opposed to the deterministic case. As in (3) for the deterministic case, we can write down the Bellman equations for (6) as:

$$V^{\star}(\tilde{q}_{i,d}, d) = \max_{\tilde{a}_{i,d}}\left\{v_{i,d}(\tilde{a}_{i,d}) + \mathbb{E}_{d-1}\big(V^{\star}(\tilde{q}_{i,d-1}, d-1)\big)\right\}, \tag{7}$$

where $V^{\star}(\tilde{q}_{i,d}, d)$ is the maximum utility that the user can obtain in state $(\tilde{q}_{i,d}, d)$ considering the future expected utility. In general, this dynamic programming under uncertainty as in (6) is not analytically tractable, as the optimization of the current state requires us to consider all possible combinations of future state transitions. However, (7) enables us to find the following condition on the optimal solution of (6):

Proposition 2: When the $\omega_{i,t}$ are random variables, user i's optimal usage $\tilde{a}_{i,d}^{\star}$ from solving (6) satisfies

$$\left.\frac{\partial v_{i,d}(\tilde{a}_{i,d})}{\partial \tilde{a}_{i,d}}\right|_{\tilde{a}_{i,d}^{\star}} = \mathbb{E}_{d-1}\left(\left.\frac{\partial v_{i,d-1}(\tilde{a}_{i,d-1})}{\partial \tilde{a}_{i,d-1}}\right|_{\tilde{a}_{i,d-1}^{\star}}\right). \tag{8}$$

Since $\tilde{q}_{i,d-1} = \tilde{q}_{i,d} - \tilde{a}_{i,d}$, (8) is derived from the first-order condition of (7). Proposition 2 implies that at the optimal solution, the user's marginal increase in her daily utility on the current day ($\partial v_{i,d}(\tilde{a}_{i,d})/\partial \tilde{a}_{i,d}$) equals the expected marginal utility increase in the next day, i.e., users should experience the same gain in utility every day. Leveraging the result in Proposition 2, we further characterize users' optimal usage:

Corollary 2: When $\{\omega_{i,t}, 1 \leq t \leq D\}$ are random variables, user i's optimal usage on the dth last day, i.e., $\tilde{a}_{i,d}^{\star}$ satisfying (8), can be extended as

$$\tilde{a}_{i,d}^{\star} = \begin{cases} \big(\omega_{i,d}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}}, \ \tilde{q}_{i,d} < \big(\omega_{i,d}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}}, & \text{(9a)} \\ \eta(\omega_{i,d}, \tilde{q}_{i,d}), \ \text{otherwise}, & \text{(9b)} \end{cases}$$

with $\eta(\omega_{i,d}, \tilde{q}_{i,d}) \in \big\{\tilde{a}_{i,d} \,\big|\, \partial v_{i,d}(\tilde{a}_{i,d})/\partial \tilde{a}_{i,d} = \mathbb{E}_{d-1}\big((\partial v_{i,d-1}(\tilde{a}_{i,d-1})/\partial \tilde{a}_{i,d-1})\big|_{\tilde{a}_{i,d-1}^{\star}}\big)\big\}$ satisfying

$$\omega_{i,d}^{\frac{1}{\alpha_i}}\tilde{q}_{i,d}/\big(\omega_{i,d}^{\frac{1}{\alpha_i}} + E_{d-1}^{\frac{1}{\alpha_i}}\big) \leq \eta(\omega_{i,d}, \tilde{q}_{i,d}) \leq (\omega_{i,d}/\phi_i)^{\frac{1}{\alpha_i}}, \tag{10}$$

where $E_d = \mathbb{E}_d\big((\omega_{i,d}^{1/\alpha_i} + E_{d-1}^{1/\alpha_i})^{\alpha_i}\big)$ with $E_1 = \mathbb{E}_1(\omega_{i,1})$, and $\big(\omega_{i,1}(\pi+\phi_i)\big)^{1/\alpha_i} \leq \eta(\omega_{i,1}, \tilde{q}_{i,1}) \leq (\omega_{i,1}/\phi_i)^{1/\alpha_i}$.

Specifically, a closed-form solution can be mathematically derived in the following special case.

[5] Users can learn $f_{i,d}(\omega)$ from the daily usage in previous months as it is reasonable to assume that each user's usage establishes a unique pattern.

Corollary 3: If $\phi_i = 0$ and $\pi = +\infty$, i.e., a prepaid data plan with no time cost, then the optimal usage satisfies

$$\tilde{a}^\star_{i,d} = \frac{\omega_{i,d}^{1/\alpha_i} \tilde{q}_{i,d}}{\omega_{i,d}^{1/\alpha_i} + E_{d-1}^{1/\alpha_i}}, \tag{11}$$

where E_d is defined in Corollary 2. If $\alpha_i = 1$, this user usage satisfies $\tilde{a}^\star_{i,d} \le \tilde{q}_{i,d}/d$.

C. Comparing the Deterministic and Stochastic Models

We can further elaborate Corollary 3's results by considering the special cases of different distributions for $\omega_{i,d}$, where $\mathbb{E}(\omega_{i,d}) = \mu$, $\phi_i = 0$, $\pi = +\infty$ and $\alpha_i = 1$. In particular we wish to compare the optimal usage amounts in (11) to those in the deterministic case (5). From the result in (5), we first state the optimal usage for the deterministic case, i.e., the distribution of $\omega_{i,d}$ is a delta distribution centered at μ:

Example 1: If $\Pr(\omega_{i,d} = \mu) = 1$, then $a^\star_{i,d} = a^\star_{i,d-1} = q_{i,d}/d$ and $V^\star(q_{i,d}) = d\mu \log(q_{i,d}/d)$.

In the stochastic case, the user would conserve her quota for the days with a high value of $\omega_{i,d}$.

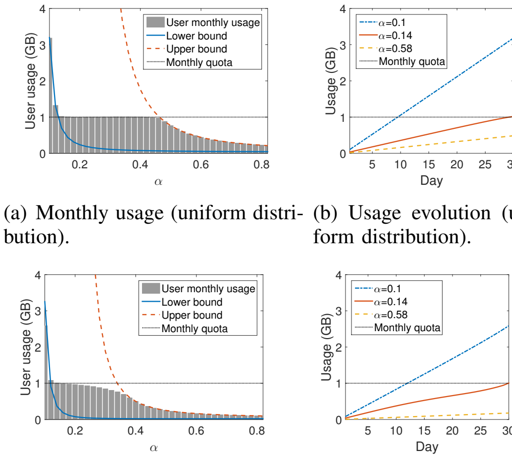
Example 2: Consider the stochastic case in which the $\omega_{i,d}$ are i.i.d. with $\mathbb{E}(\omega_d) = \mu$. Then $a^\star_{i,d} = \frac{\omega_{i,d} q_{i,d}}{\omega_{i,d} + (d-1)\mu}$ and,

$$V^\star(q_{i,d}, d) = d\mu \log q_{i,d} + \mathbb{E}\big((d-1)\omega_{i,d} \log \omega_{i,d}\big) + \textstyle\sum_{i=1}^{d-1} i\mu \log(i\mu) - \sum_{i=1}^{d-1} \mathbb{E}\big((\omega_{i,d} + i\mu)\log(\omega_{i,d} + i\mu)\big).$$

Consider the specific example in which $D = 30$ days, the quota $A = 1000MB$ and $\mu = 0.03$. For the deterministic case where $\omega_{i,d}$ is always μ the optimal utility is $V^\star = 3.16$. However if $\Pr(\omega_{i,d} = K) = \frac{1}{K}$ and $\Pr(\omega_{i,d} = 0) = \frac{K-1}{K}$, then for $K = 10$ the optimal utility is $V^\star = 4.86$. This additional utility arises from the fact that the user can conserve quota on low utility days until it is most needed on high utility days.

We conclude with a more general stochastic example. Figures 2(a) and 2(c) show users' expected monthly usage if their $\omega_{i,d}$ follow the same distribution and they are i.i.d. for all days. We consider two distribution types: a uniform distribution in Figure 2(a), e.g., if users' usage may vary significantly from day to day; and a Pareto distribution in Figure 2(c), corresponding to small $\omega_{i,d}$ for most days but occasional large values of $\omega_{i,d}$. For instance, some users may derive greater utility from consuming data while traveling.

We can observe from both figures that when α_i is sufficiently small, so that the data quota $A < D\mathbb{E}_d\big((\omega_{i,d}/(\pi + \phi_i))^{1/\alpha_i}\big)$, heavy users limit their usage to $\sum_{t=1}^{D} a^\star_{i,t} = D\mathbb{E}_d\big((\omega_{i,d}/(\pi + \phi_i))^{1/\alpha_i}\big)$, the lower bound in the figures, but still consume overage data. When α_i is large, so that $A > D\mathbb{E}_d\big((\omega_{i,d}/\phi_i)^{1/\alpha_i}\big)$, light users would consume the maximum usage $\sum_{t=1}^{D} a^\star_{i,t} = D\mathbb{E}_d\big((\omega_{i,d}/\phi_i)^{1/\alpha_i}\big)$, i.e., the upper bound in the figure, with leftover data at the end of the billing cycle. When $D\mathbb{E}_d\big((\omega_{i,d}/(\pi + \phi_i))^{1/\alpha_i}\big) \le A \le D\mathbb{E}_d\big((\omega_{i,d}/\phi_i)^{1/\alpha_i}\big)$, users' monthly usage remains very close to the quota as in the deterministic model, though this behavior diverges slightly when $\omega_{i,d}$ is drawn from a Pareto distribution. Since the Pareto distribution allows high values of $\omega_{i,d}$ to occur, users would limit their usage so as to hedge against the risk of a high utility weight in future days of the month.



(a) Monthly usage (uniform distribution). (b) Usage evolution (uniform distribution).

(c) Monthly usage (Pareto distribution). (d) Usage evolution (Pareto distribution).

Fig. 2. User usage dynamics for the stochastic model with varying α_i in a billing cycle of $D = 30$ days, where users are offered $A = 1000$MB data and overage fee is $\pi = 0.015$\$/MB. We suppose $\phi_i = 0.006$\$/MB. In (a) and (b), $\omega_{i,d}$ follows a uniform distribution with PDF $f(\omega) = \frac{1}{\bar{\omega} - \underline{\omega}}$ ($\bar{\omega} = 0.02$ and $\underline{\omega} = 0.04$). In (c) and (d), $\omega_{i,d}$ follows a Pareto distribution with PDF $f(\omega) = \frac{\gamma \omega_m^\gamma}{\omega^{\gamma+1}}$ ($\gamma = 3.2$ and $\omega_m = 0.01$). The upper bound and lower bound plot $D\mathbb{E}\big((\omega_{i,d}/\phi_i)^{1/\alpha_i}\big)$ and $D\mathbb{E}\big((\omega_{i,d}/(\pi + \phi_i))^{1/\alpha_i}\big)$ respectively.

In Figures 2(b) and 2(d), we plot the expected user usage in each day of the billing cycle, covering the cases of overage consumption ($\alpha_i = 0.1$), data depletion ($\alpha_i = 0.14$), and leftover ($\alpha_i = 0.58$). We observe a linear increase in the overage and leftover cases, due to $\omega_{i,d}$ following the same distribution everyday. We do observe from the middle curve in Figure 2(d), that user usage increase is slightly lower at the beginning of the month but more aggressive at the end until her usage reaches the given quota, showing that users reserve data by taking into account the future uncertainty.

Overall, we observe from Figure 2 that the three cases of data overage, depletion, and leftover derived in Corollary 1 for different quotas A with the deterministic model still approximately hold for the stochastic model. In the next section, we further discuss user utility and ISP profit with different pricing policies. For simplicity we focus on the deterministic model since it is more tractable. However, using the results from the previous section it is feasible to carry out a similar analysis for the stochastic model.

IV. Impacts of Pricing Policy

Given user usage under a pricing policy (P, A, π), we now turn to analyzing the user and ISP benefits. In doing so, we are able to derive conditions under which *both users and ISPs are better off under the same type of data plan.*

A. User Utility

Leveraging the results that we derived in Corollary 1, we first find the total monthly utility that a given user i receives under a pricing policy (P, A, π):

$$U(\vec{\mathbf{a}}_i^\star)=\begin{cases}\displaystyle\sum_{t=1}^{D}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}}-P,\ A>\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{\frac{1}{\alpha_i}}, & \text{(12a)}\\ \displaystyle\sum_{t=1}^{D}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}}+\pi A-P, & \\ \displaystyle\qquad\qquad A<\sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{\frac{1}{\alpha_i}}, & \text{(12b)}\\ \displaystyle\left(\sum_{t=1}^{D}\omega_{i,t}^{\frac{1}{\alpha_i}}\right)^{\alpha_i}\frac{A^{1-\alpha_i}}{1-\alpha_i}-\phi_i A-P,\ \text{otherwise.} & \text{(12c)}\end{cases}$$

For $(P \in \mathbb{R}_+, A \in \mathbb{R}_+, \pi \in \mathbb{R}_+)$, (12) gives the user utility under a quota with overage data plan. We can observe that (12) is continuously decreasing with A, meaning that capping user usage harms their utility; intuitively, this would be the case as a lower cap would make it more likely for users to incur overage charges. We next analyze specific variations of (12) corresponding to commonly offered data plans.

1) Unlimited data plans ($A = +\infty, P > 0, \pi = 0$)*:* All users with unlimited data plans fall into the conditions of (12a), i.e., users would always consume the maximum possible amount $a_{i,d}^\star = (\omega_{i,d}/\phi_i)^{1/\alpha_i}$. Thus, (12) can be reduced to $U_u(\vec{\mathbf{a}}_i^\star) = \sum_{t=1}^{D} \frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}} - P$ for user utility with unlimited data plans. Due to rapid increases in user demand for mobile data, many alternatives to unlimited data plans have been offered that attempt to penalize excessive amounts of user usage by either raising the price or capping users' data.

2) Usage-based data plans ($A = 0, P = 0, \pi > 0$)*:* Under these "pay-as-you-go" plans, users only pay for the usage volume at a unit price $\pi > 0$. By substituting $A = 0$ and $P = 0$ into (12), we find user utility with usage-based data plans: $U_s(\vec{\mathbf{a}}_i^\star) = \sum_{t=1}^{D} \frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}}$. Thus, *all* users' data usage are reduced by a factor of $\big(\phi_i/(\pi+\phi_i)\big)^{1/\alpha_i}$ as compared to unlimited data plans, though ideally, the ISP may only want to bring down the usage of heavy users.

3) Prepaid data plans ($A > 0, P > 0, \pi = +\infty$)*:* Under these plans, users' monthly data usage are forced below a given quota A. With no opportunity to purchase overage data, user utility in this case can be represented by (12a) or (12c): if $A > \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$, then $U_p(\vec{\mathbf{a}}_i^\star) = \sum_{t=1}^{D} \frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}} - P$; otherwise, $U_p(\vec{\mathbf{a}}_i^\star) = \big(\sum_{t=1}^{D}\omega_{i,t}^{1/\alpha_i}\big)^{\alpha_i}\frac{A^{1-\alpha_i}}{1-\alpha_i} - \phi_i A - P$. Thus, heavy users are primarily affected, who are more likely to exceed their quotas.

Given the above characterization of users' utilities under different types of data plans, we can now characterize which users benefit the most from each type of data plan:

Proposition 3: Suppose all data plans offer the same quota A and price P, with $A < +\infty$ and $P > 0$. Then users always derive more utility from an unlimited data plan compared to a quota-with-overage plan, which yields a higher utility than a prepaid plan: $U_u(\vec{\mathbf{a}}_i^\star) \geq U(\vec{\mathbf{a}}_i^\star) \geq U_p(\vec{\mathbf{a}}_i^\star)$. Depending on P, usage-based data plans may realize higher or lower utilities than any of the three other plans.

The details of these relationships under different conditions on P and A is listed in Table II of the appendix. As $U_u(\vec{\mathbf{a}}_i^\star)$ is always at least as large as $U(\vec{\mathbf{a}}_i^\star)$ and $U_p(\vec{\mathbf{a}}_i^\star)$, users prefer the freedom of not having overage charges with unlimited compared to prepaid plans, assuming the same flat-fee payments P. Users also benefit from the option of consuming overage data, compared to prepaid data plans without this option ($U(\vec{\mathbf{a}}_i^\star) \geq U_p(\vec{\mathbf{a}}_i^\star)$). Usage-based data plans yield higher utilities than unlimited or prepaid plans if the quota fee P is large: usage-based data plans are attractive to users when they are charged too much upfront.

B. ISP Profit

To analyze ISP profit, we suppose that the ISP incurs a linear operational cost $\sigma \in (0, \min\{P/A, \pi\})$ of handling data traffic [10], [19]. Since ISP profit is driven by user demands, we divide users into three subsets according to their usage volume, as in Sections III-A1–III-A3:

- **Light users** with small demands: $\mathcal{N}_S = \{i \in \mathcal{N} \mid A > \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}\}$. Light users would not exceed the monthly quota, paying no overage fees. These users each pay the quota fee P for a usage amount $\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$.
- **Moderate users** with medium demands: $\mathcal{N}_M = \{i \in \mathcal{N} \mid \sum_{t=1}^{D}(\omega_{i,t}/(\pi+\phi_i))^{1/\alpha_i} \leq A \leq \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}\}$. Like the light users, moderate users also consume no overage data, but they use up their entire monthly quota, so revenue from each of these users remains P with a traffic cost σA.
- **Heavy users** with large demands: $\mathcal{N}_L = \{i \in \mathcal{N} \mid A < \sum_{t=1}^{D}(\omega_{i,t}/(\pi+\phi_i))^{1/\alpha_i}\}$. Heavy users, whose monthly usage exceeds their quotas, not only need to pay overage fees of $\pi\Big(\sum_{t=1}^{D}(\omega_{i,t}/(\pi+\phi_i))^{1/\alpha_i} - A\Big)$ in total, but also yield a larger traffic cost $\sigma\sum_{t=1}^{D}(\omega_{i,t}/(\pi+\phi_i))^{1/\alpha_i}$ for their ISPs.

We note that the division of users is relative to the quota A set by the ISP. In general, the ISP receives the profit

$$R(A,P,\pi) = \sum_{i\in\mathcal{N}_S}\left(P-\sigma\sum_{t=1}^{D}\Big(\frac{\omega_{i,t}}{\phi_i}\Big)^{\frac{1}{\alpha_i}}\right) + N_M(P-\sigma A) + \sum_{i\in\mathcal{N}_L}\left(P+\pi\left(\sum_{t=1}^{D}\Big(\frac{\omega_{i,t}}{\pi+\phi_i}\Big)^{\frac{1}{\alpha_i}}-A\right)-\sigma\sum_{t=1}^{D}\Big(\frac{\omega_{i,t}}{\pi+\phi_i}\Big)^{\frac{1}{\alpha_i}}\right) \tag{13}$$

We now quantify this expression if the ISP only offers unlimited, usage-based, and prepaid data plans:

1) Unlimited data plans ($A = +\infty, P > 0, \pi = 0$)*:* In this case, all users are "light" users relative to A, so $\mathcal{N}_S = \mathcal{N}$. We find that the ISP profit is given by $R_u(A, P, \pi) = \sum_{i\in\mathcal{N}}(P - \sigma\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i})$. A possible reason for ISPs' elimination of unlimited data plans could be the increased network traffic shrinking their profit or even leading to negative profit, i.e., mathematically $NP < \sigma\sum_{i\in\mathcal{N}}\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$.

2) Usage-based data plans ($A = 0, P = 0, \pi > 0$)*:* In this case, all users would be "heavy" users relative to $A = 0$, so $\mathcal{N}_L = \mathcal{N}$. We then find ISP profit $R_s(A, P, \pi) =$

$(\pi-\sigma)\sum_{i\in\mathcal{N}}\sum_{t=1}^{D}(\omega_{i,t}/(\pi+\phi_i))^{1/\alpha_i}$. Since the overall user usage is lower compared to unlimited data plans, the ISP may earn more revenue from offering unlimited plans.

3) Prepaid data plans $(A > 0, P > 0, \pi = +\infty)$*:* In prepaid data plans, each user in $\mathcal{N}_M \cup \mathcal{N}_L$ contributes an amount $P - \sigma A$ to ISP profit with no overage consumption. Therefore, ISP profit by offering prepaid data plans is $R_p(A, P, \pi) = \sum_{i\in\mathcal{N}_S}(P - \sigma\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}) + (N_M + N_L)(P - \sigma A)$.

We also compare the profit that the ISP would gain from each type of data plan:

Proposition 4: Suppose all data plans offer the same quota A charged at the same P if $A < +\infty$ and $P > 0$. ISP profit with offering unlimited data plans $R_u(A, P, \pi)$, prepaid data plans $R_p(A, P, \pi)$, and quota with overage data plans $R(A, P, \pi)$ follows the relationship: $R(A, P, \pi) \geq R_p(A, P, \pi) \geq R_u(A, P, \pi)$. Depending on P, ISPs may gain higher or lower profit with usage-based data plans than any of the three other data plans.

We also elaborate the relationship of ISP profit under different data plans in Table III in the appendix. In general, users benefit the most when the ISPs make the least profit–for instance, users gain more utility from unlimited compared to prepaid data plans, but ISPs gain less profit from unlimited compared to prepaid plans (cf. Proposition 3). However, *both the users and the ISP benefit more from quota with overage data plans compared to prepaid data plans*. Moreover, as we show in Section V-A, in some cases quota with overage plans offer the same user utility as unlimited plans and ISP profit as prepaid plans, allowing both users and ISPs to benefit the most from the same type of data plan.

V. Experiments

A. Numerical Examples

We now use numerical examples in Figures 3 and 4 to compare the user utilities and ISP profit under different pricing policies, as analyzed in Propositions 3 and 4, respectively. In these figures, we vary the quota offered for the prepaid and quota with overage plans. Both user utilities and ISP profit under unlimited data plans and usage-based data plans are horizontal lines in all figures, since users are either offered infinite quota or no quota in these two types of data plans, i.e., their usage are not affected by the amount of quota. We find that at a large range of possible data quotas A, *many data plans give the same utilities or profits*, allowing us to find areas where a single data plan maximizes both ISP profit and user utility.

Figure 3 shows user utilities with the four data plans discussed in Section IV-A, illustrating Proposition 3. The received utilities with prepaid and quota with overage data plans are similar, becoming closer to each other as the quota increases and taking the same values when the quota exceeds $\sum_{t=1}^{D}\left(\omega_{i,t}/(\pi+\phi_i)\right)^{1/\alpha_i}$ (illustrated by the first vertical dotted line in the figures): at these quotas, having overage data does not affect user utility as the amount of quota can satisfy this user's desired monthly usage. As the offered quota approaches $\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$ (illustrated by the second vertical dotted line in the figures), i.e., the user is offered more than enough data, both plans reach the utility achieved with unlimited data plans. As the flat fee P increases from Figure 3(a), 3(b), 3(c) to 3(d), we can see that usage-based plans move from providing the least to the most utility. Matching this observation to real-life data plans, we notice that most overage fees are higher than the flat fee per unit of quota, i.e., $\pi > P/A$. If this relationship is reversed and P is too expensive, users may defect to a usage-based data plan to avoid this fee.

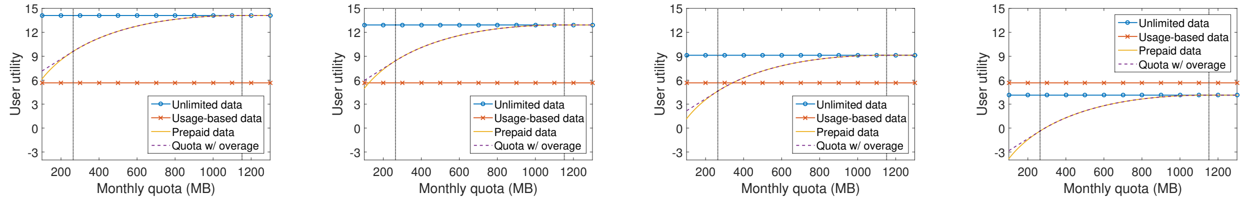
Figure 4 visualizes the ISP profit results in Proposition 4. Unlike the small gap between user utilities with prepaid data plans and the quota with overage data plans, the latter yields much higher profits for the ISP than the former. This gap is gradually reduced as the quota increases and eventually becomes zero after $N_L = 0$ (i.e., there are no more heavy users consuming overage data, illustrated by the first vertical dotted line in the figures). However, the profit gap between prepaid data plans and unlimited data plans is so small as to be invisible without zooming into Figure 4(c). Likewise, these two types of data plans yield the same profit when the ISP offers enough data to its user, i.e., $N_M = N_L = 0$ (illustrated by the second vertical dotted line in the figures). Opposite to user utility, ISP profit with usage-based data plans moves from the highest to lowest as the flat fee P increases from Figure 4(a), 4(b), 4(c) to 4(d).

Unlike user utility, ISP profit decreases with the quota amount. Thus, for most combinations of A and P, users' and their ISPs' benefits with different data plans conflict, reflected by the reversed orders of user utility and ISP profit. However, we do observe that if the ISP offers enough data and charges users at a reasonable price, they both prefer the same type of data plan, such as when the quota exceeds 12.44GB in Figures 3(c) and 4(d).

B. Choosing Which Plans to Offer

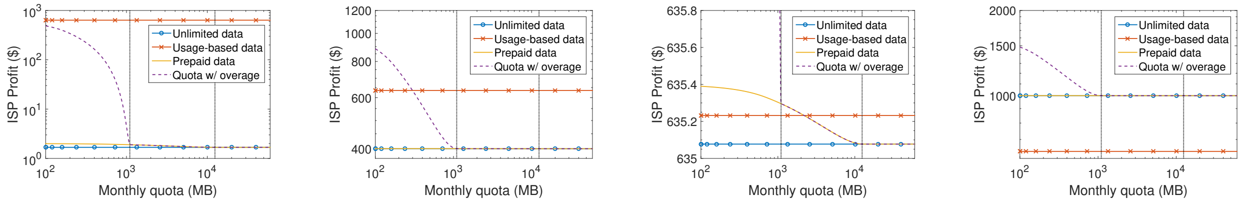
We now apply our analysis to a real-world usage trace from 13 mobile users of a U.S. ISP for one billing cycle. For each user in this dataset, we have the user's data plan information (monthly quota and flat fee for the quota), usage per day, and date of the usage from February 1 to March 2 in 2013. The users in the dataset subscribe to either an unlimited data plan at 30\$, or a quota with overage data plan at 20\$ for 1GB of mobile data. In 2013, this ISP charged users an overage fee $\pi = 20\$\text{GB}$, and it had not throttled user data when they run over their quota. We first show that *some users chose a suboptimal data plan*, and then evaluate *whether this ISP should offer a new, usage-based data plan*.

To compute user utility and ISP profit with different data plans from the current offering, we need to learn $\omega_{i,d}$ of each day d, ϕ_i, and α_i for each user. We refer to the appendix for the method of this parameter estimation. We show each user's estimated $\hat{\phi}_i$ and $\hat{\alpha}_i$ in Figure 5 and show their monthly usage by the sizes of the markers. Users with a 1GB quota who limit their usage below 1GB either have higher $\hat{\phi}_i$ or higher



(a) $P = 0.02\$/\text{GB}$. (b) $P = 1.2\$/\text{GB}$. (c) $P = 7\$/\text{GB}$. (d) $P = 10\$/\text{GB}$.

Fig. 3. Comparison of user utility under different pricing policies with varying flat fee P and monthly data quota A in a billing cycle of $D = 30$ days. We set $\pi = 0.015\$/\text{MB}$, and suppose this user has $\phi_i = 0.021\$/\text{MB}$ $\alpha_i = 0.38$, and $\omega_{i,d} = 0.08$ for all d.



(a) $P = 0.02\$/\text{GB}$. (b) $P = 4\$/\text{GB}$. (c) $P = 6.354\$/\text{GB}$. (d) $P = 7\$/\text{GB}$.

Fig. 4. Comparison of ISP profit under different pricing policies with varying flat fee P and monthly data quota A. All parameter settings are the same as in Figure 3. In addition, we suppose there are $N = 100$ users subscribing to the ISP, where $\{\alpha_i\}_{i=1}^{D} = \{0.4, 0.3983, \ldots, 0.23\}$, and $\{\omega_{i,d}\}_{i=1}^{D} = \{0.7, 0.701, \ldots, 0.8\}$ for each user i for all d. The traffic maintenance cost for the ISP is set to $\sigma = 10^{-6}\$/\text{MB}$.

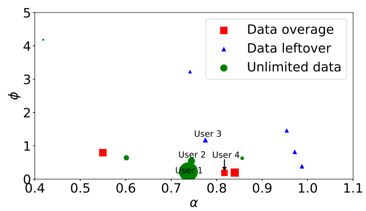


Fig. 5. Estimated $\hat{\alpha}_i$ and $\hat{\phi}_i$ for 13 users. Marker sizes are proportional to users' monthly usage.

$\hat{\alpha}_i$, meaning they either have a high time opportunity cost for mobile data usage or are very cost-sensitive. We shall pay more attention to the heterogeneity of users with unlimited data plans (marked with green circles): the largest data usage is among users with the unlimited data plan, but there is one unlimited user with high time cost ($\hat{\phi}_i = 1$) and thus low usage. Thus, we find that users may not fully understand their own usage, and may then choose the wrong data plans for themselves.

Depending on the $\hat{\phi}_i$ and $\hat{\alpha}_i$ that we obtained by learning users' usage for each of them, we estimate user utility and ISP profit from each user under three different types of data plans, including the unlimited data plan and quota with overage data plan that the ISP was offering, as well as an extra usage-based data plan with $\pi = 20\$/\text{GB}$. Table I lists four representative users' usage and utilities achieved under these data plans as well as ISP profit from each of them, assuming a traffic maintenance cost $\sigma = 0.1\$/\text{GB}$. We extend Table I to the same set of results for all 13 users in the appendix. User 1 is the heaviest in our dataset, i.e., the largest green circle in Figure 5: her usage is well aligned with her current data plan, since she would be charged much more in other data plans, dragging down her utility. Clearly, *unlimited data plans benefit heavy users more.* There are also users like user 2, whose usage is relatively higher than the quota but not ultra-high like user 1: switching to other data plans only slightly increases her utility. However, light users, like user 3, receive a higher utility by staying in the quota with overage data plan, as compared to the unlimited data plan. If user 3 could choose a usage-based data plan, she could receive an even higher utility. Thus, *usage-based data plans are suitable for light users, as they may waste too much leftover data under prepaid data plans.* Although user 4's usage could be higher than 1GB under the unlimited data plan, she can limit it to the quota and achieve a better utility under the quota with overage plan. Thus, *users with desired usage near the quota should choose the quota with overage data plan.*

We also calculate ISP profit if all 13 users subscribe to the unlimited data plan, quota with overage data plan, or usage-based data plans: they are 384.14\$, 454.73\$, and 357.10\$, respectively. We thus see that the ISP should not offer a usage-based data plan.

Finally, Figure 6 shows the average user utility and ISP profit while varying the quota A offered by the ISP, assuming the flat fee for the quota is set to $P = 0.02\$/\text{MB} \times A$ and users subscribe to the data plans that lead them to the highest utility. We find that the ISP's profit-maximizing data quota is 1.5GB. As we can expect, user utility decreases as ISP profit increases. When the quota exceeds 1.5GB, the flat fee P for the quota with overage data plan exceeds that for the unlimited data plan, users all choose the unlimited data plan. However, we note that this profit is based only on the current set of users. We do not model new users subscribing to the ISP if they are attracted by new data plans, or users leaving the ISP for another provider. Thus, we leave as an open question the effect of competition with other ISPs on ISPs' optimal data plans.

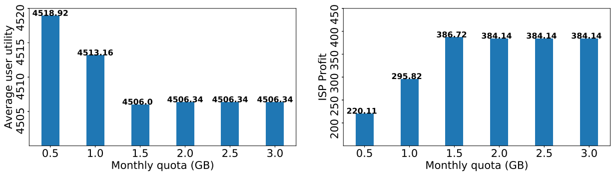
VI. Discussion and Conclusions

In this work, we use dynamic programming to introduce the first model of users' data consumption decisions in the presence of a monthly quota. Though our model allows us to rigorously derive conditions under which users benefit

TABLE I

COMPARISON OF USER UTILITY AND ISP PROFIT IF USERS SWITCH TO A DIFFERENT DATA PLAN.

	Current data plan	Monthly usage (MB)			User utility			ISP profit ($)		
		Unlimited	Quota	Usage-based	Unlimited	Quota	Usage-based	Unlimited	Quota	Usage-based
User 1	Unlimited	10628.97	9459.76	9459.76	**6632.52**	6462.03	6462.03	29.34	188.25	188.25
User 2	Unlimited	1379.85	1314.72	1314.72	2164.49	**2167.55**	**2167.55**	29.78	26.16	26.16
User 3	Quota	574.33	574.33	561.94	2292.20	2302.20	**2310.83**	29.77	19.94	11.18
User 4	Quota	1104.56	1000.00	976.03	897.80	907.00	**907.04**	29.91	19.90	19.42



(a) Average user utility.

(b) ISP profit.

Fig. 6. Estimated user utility and ISP profit with varying quota.

from different types of data plans, it–like any model of user behavior–is inherently stylized, and may not reflect the reality of user actions. Thus, to conclude the paper we first discuss some ways to extend these limitations before discussing the implications of our findings on mobile data markets and pointing towards promising directions of future work.

Modeling user behavior. In practice, users would not explicitly optimize their usage according to our dynamic program, but they might subconsciously trade-off data consumption today with saving data for the future. Empirical evidence suggests that users do engage in such planning [17], though in Section V-B we see that users do not always choose their optimal data plans. We argue, however, that dynamic programming provides a convenient approximation to this user behavior and allows us to provide qualitative insights into how ISPs should optimize their data plan offerings.

Parameter estimation. ISPs can use historical data on their users' data consumption amounts over several months to infer the parameters of users' utility models. In Section V, we utilize some practical methods for our numerical experiments on one month of user data. It is unlikely that ISPs would obtain exact estimates (indeed, users' true utilities may not always follow an α-fair utility function). However, the ISPs' decisions on which data plans to offer are robust to changes in model parameters, as the user behavior captured by different model parameters generally preserves the same pattern.

Implications of our work. Cellular data usage has greatly increased over the past years due to better network infrastructure and a wider range of data-consuming applications, leading to larger ω in user utility. We find that users prefer unlimited rather than usage-based data plans, which may partially explain why AT&T and Verizon have recently offered unlimited data plans again [2]. However, these plans throttle usage after a threshold quota, making them in effect a softer version of prepaid data plans. Since light users may prefer usage-based data plans, we conjecture that ISPs may start to offer such plans in an effort to attract light users. Google Fi [4] represents a step in this direction.

Future research. This work represents a first attempt to model users' data usage dynamics under a monthly quota. Future models could be made more accurate by taking into account the correlation in model parameters and known periodicity in usage data volumes, e.g., higher data consumption on weekends compared to weekdays. One could also expand the types of data plans considered to reason about how new types of pricing like sponsored data would affect ISPs' incentives for offering different data plans. Our usage dynamics models could then be used to derive optimal sponsoring algorithms.

REFERENCES

[1] A. Lella, "U.S. smartphone penetration surpassed 80 percent in 2016." ComScore, 2017. https://www.comscore.com/Insights/Blog/US-Smartphone-Penetration-Surpassed-80-Percent-in-2016.

[2] M. Reardon, "AT&T's new unlimited data plan: All you need to know (FAQ)," 2016. https://www.cnet.com/news/at-ts-new-unlimited-data-plan-everything-you-need-to-know-faq/.

[3] P. Michaels, "Two-year phone contracts are going away ł and that's good for you." Tom's Guide, 2015. https://www.tomsguide.com/us/sprint-ends-contracts,news-21478.html.

[4] Google, "Project Fi," 2017. https://fi.google.com/.

[5] K. Finley, "Comcast's new mobile service is a good deal, but maybe not good enough." Wired, 2017. https://www.wired.com/2017/04/comcasts-new-mobile-service-good-deal-maybe-not-good-enough/.

[6] P. Hande, M. Chiang, R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," in *Proc. of IEEE INFOCOM*, 2009.

[7] 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*, 2015.

[8] L. Zhang, W. Wu, and D. Wang, "TDS: Time-dependent sponsored data plan for wireless data traffic market," in *Proc. of IEEE INFOCOM*, 2016.

[9] 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.

[10] L. Zheng, J. Chen, C. Joe-Wong, C. W. Tan, and M. Chiang, "An economic analysis of wireless network infrastructure sharing," in *Proc. of IEEE WiOpt*, 2017.

[11] L. Zheng and C. Joe-Wong, "Understanding rollover data," in *Proc. of IEEE INFOCOM Workshop on Smart Data Pricing*, 2016.

[12] S. Sen, C. Joe-Wong, and S. Ha, "The economics of shared data plans," in *Proc. of WITS*, 2012.

[13] M. M. Lankhorst, H. Van Kranenburg, A. Salden, and A. J. Peddemors, "Enabling technology for personalizing mobile services," in *Proc. of HICSS*, 2002.

[14] X. Wang, R. T. Ma, and Y. Xu, "The role of data cap in two-part pricing under market competition," in *Proc. of IEEE INFOCOM Smart Data Pricing workshop*, 2015.

[15] 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.

[16] J. Blackburn, R. Stanojevic, V. Erramilli, A. Iamnitchi, and K. Papagiannaki, "Last call for the buffet: economics of cellular networks," in *Proc. of ACM MobiCom*, 2013.

[17] A. Mathur, B. Schlotfeldt, and M. Chetty, "A mixed-methods study of mobile users' data usage practices in south africa," in *Proceedings of ACM UbiComp*, 2015.

[18] R. E. Bellman, "An introduction to the theory of dynamic programming," 1953.

[19] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "A survey of smart data pricing: Past proposals, current plans, and future trends," *ACM Computing Surveys*, vol. 46, no. 2, p. 15, 2013.

Appendix

A. Parameter Inference for A Real-world Dataset

We introduce the parameter estimation for the dataset that we described in Sectoin V-B. This parameter estimation is based on our result in (4). We find that users in real-life either have leftover or consume overage, i.e., data depletion is not observed in our dataset. We also observe that most users do not exceed their quota, but overage consumption did happen for three of them. If we would have user usage for several months with some months of data leftover and some of data overage, we can divide the log of user i into two subsets: one subset of all users' daily usage in the months without overage usage, forming a cumulative density function (CDF) $F_i^{w/o}(\mathbf{a}_{w/o})$ of all observed daily usage $\mathbf{a}_{w/o}$ in these months, and the other subset of all observed daily usage $\mathbf{a}_w$ in the months with overage usage, forming the corresponding CDF $F_i^{w}(\mathbf{a}_w)$. Assuming $\omega_{i,d}$ follow a distribution with CDF $\hat{F}_i(\omega)$ that we can either assume a distribution that visually match the observation or learn from kernel density estimation, we first obtain $\hat{\phi}_i$ and $\hat{\alpha}_i$ by solving the following optimization problem:

$$\begin{aligned}\underset{\alpha_i,\phi_i,\boldsymbol{\gamma}}{\text{minimize}}\quad & \|\hat{F}_i(\phi_i {\mathbf{a}_{w/o}}^{\alpha_i}) - F_i^{w/o}(\mathbf{a}_{w/o})\|_2 \\ & + \|\hat{F}_i((\pi+\phi_i){\mathbf{a}_w}^{\alpha_i}) - F_i^{w}(\mathbf{a}_w)\|_2,\end{aligned} \tag{14}$$

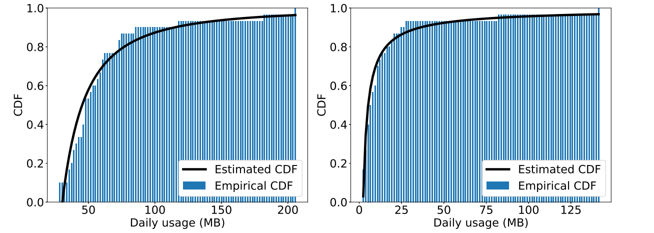
where $\boldsymbol{\gamma}$ is a group of parameters to shape the distribution $\hat{F}_i(\omega)$. Then, we can calculate $\hat{\omega}_{i,d} = \hat{\phi}_i a_{i,d}^{\hat{\alpha}_i}$ for the months without overage usage, and $\hat{\omega}_{i,d} = (\pi + \hat{\phi}_i) a_{i,d}^{\hat{\alpha}_i}$ for the months with overage usage. All these $\hat{\omega}_{i,d}$ constitutes to an estimated distribution for user i's weights on usage utility extracted from their empirical data usage.

Since our dataset only contains the usage of each user for a month, we consider either the first or second part of (14) depending on whether or not the user had leftover by the end of the billing cycle. Following our discussion in Section III, we use Pareto distribution $\hat{F}_i(\omega) = 1 - (m/\omega)^{\gamma}$, where we also need to learn the parameters for this distribution: γ and m. Thus, if the user did not consume overage, we solve for $\underset{\alpha_i,\phi_i,\gamma,m}{\text{minimize}} \left\|1 - \left(m/(\phi_i \mathbf{a}_{w/o}^{\alpha_i})\right)^{\alpha_i} - F_i^{w/o}(\mathbf{a}_{w/o})\right\|_2$; and otherwise, we solve for $\underset{\alpha_i,\phi_i,\gamma,m}{\text{minimize}} \left\|1 - \left(m/((\pi+\phi_i)\mathbf{a}_w^{\alpha_i})\right)^{\alpha_i} - F_i^{w}(\mathbf{a}_w)\right\|_2$.

Figure 7 shows the distribution fits: as we minimize the linear squares divergence between the estimated and empirical CDFs, we use R^2 values to illustrate the goodness of the fitting results, calculated by

$$R^2 = 1 - \frac{\sum(\hat{F}_i - F_i)^2}{\sum(\hat{F}_i - \langle \hat{F}_i \rangle)^2},$$

where $\langle \cdot \rangle$ is the average operation. We observe that the Pareto distribution fits the empirical data well, with both R^2 above 0.95.



(a) $R^2 = 0.9596$, $\gamma = 2.06$, $m = 3.78$, $\hat{\alpha}_i = 0.84$, $\hat{\phi}_i = 0.20$. (b) $R^2 = 0.9558$, $\gamma = 0.86$, $m = 1.84$, $\hat{\alpha}_i = 0.97$, $\hat{\phi}_i = 0.82$.

Fig. 7. Fitting the CDF of empirical user usage to our result in (4) by assuming Pareto distribution for $\omega_{i,d}$.

B. Proof of Proposition 1

Proof: We use mathematical induction to prove the hypothesis that user i's maximum utility for the remaining d days, i.e., $V^\star(q_{i,d}, d)$ is given by

$$\begin{cases} \displaystyle\sum_{t=1}^{d} \frac{\alpha_i}{1-\alpha_i} \omega_{i,t}^{\frac{1}{\alpha_i}} \phi_i^{1-\frac{1}{\alpha_i}}, & q_{i,d} > \displaystyle\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, \quad \text{(15a)} \\ \omega_{i,d}\left(\displaystyle\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d}}\right)^{\frac{1}{\alpha_i}}\right)^{\alpha_i} \dfrac{q_{i,d}^{1-\alpha_i}}{1-\alpha_i} - \phi_i q_{i,d}, & \displaystyle\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \le q_{i,d} \le \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, \quad \text{(15b)} \\ \displaystyle\sum_{t=1}^{d} \frac{\alpha_i}{1-\alpha_i} \omega_{i,t}^{\frac{1}{\alpha_i}} (\pi+\phi_i)^{1-\frac{1}{\alpha_i}} + \pi q_{i,d}, & 0 \le q_{i,d} < \displaystyle\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}, \quad \text{(15c)} \\ \displaystyle\sum_{t=1}^{d} \frac{\alpha_i}{1-\alpha_i} \omega_{i,t}^{\frac{1}{\alpha_i}} (\pi+\phi_i)^{1-\frac{1}{\alpha_i}}, & q_{i,d} < 0. \quad \text{(15d)} \end{cases}$$

We first show that (15) is true for the base case of $d = 1$. Since $V(q_{i,0}, 0) = 0$, we have $V^\star(q_{i,1}, 1) = \max\{v_{i,1}(a_{i,1})\}$, leading $V^\star(q_{i,1}, 1)$ to equal

$$\begin{cases} \dfrac{\alpha_i}{1-\alpha_i} \omega_{i,1}^{\frac{1}{\alpha_i}} \phi_i^{1-\frac{1}{\alpha_i}}, & q_{i,1} > \left(\dfrac{\omega_{i,1}}{\phi_1}\right)^{\frac{1}{\alpha_i}}, \\ \omega_{i,1} \dfrac{q_{i,1}^{1-\alpha_i}}{1-\alpha_i} - \phi_1 q_{i,1}, & \left(\dfrac{\omega_{i,1}}{\pi+\phi_1}\right)^{\frac{1}{\alpha_i}} \le q_{i,1} \le \left(\dfrac{\omega_{i,1}}{\phi_1}\right)^{\frac{1}{\alpha_i}}, \\ \dfrac{\alpha_i}{1-\alpha_i} \omega_{i,1}^{\frac{1}{\alpha_i}} (\pi+\phi_i)^{1-\frac{1}{\alpha_i}} + \pi q_{i,1}, & 0 \le q_{i,1} < \left(\dfrac{\omega_{i,1}}{\pi+\phi_1}\right)^{\frac{1}{\alpha_i}}, \\ \dfrac{\alpha_i}{1-\alpha_i} \omega_{i,1}^{\frac{1}{\alpha_i}} (\pi+\phi_i)^{1-\frac{1}{\alpha_i}}, & q_{i,1} < 0, \end{cases}$$

which holds for (15) when $d = 1$.

Assuming (15) is true, we then take the inductive step to prove that it is also true for $V^\star(q_{i,d+1}, d+1) = \underset{a_{i,d+1}}{\max}\left\{v_{i,d+1}(a_{i,d+1}) + V(q_{i,d}, d)\right\}$ by substituting $q_{i,d} = q_{i,d+1} - a_{i,d+1}$ so that $V^\star(q_{i,d+1}, d+1) = \underset{a_{i,d+1}}{\max}\left\{V(q_{i,d+1}, d+1)\right\} = \underset{a_{i,d+1}}{\max}\left\{v_{i,d+1}(a_{i,d+1}) + V(q_{i,d+1} - a_{i,d+1}, d)\right\}$. Although both (1) and (15) are piecewise functions, we can observe $a_{i,d+1} \le q_{i,d+1}$ for (1a), (15a), and

(15b) and $a_{i,d+1} > q_{i,d+!}$ for (1b), (1c), and (15d). We then discuss case by case below to calculate for $V^\star(q_{i,d+1}, d+1)$.

1) $q_{i,d} > \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}$ and $q_{i,d+1} > 0$.

In this case, (1a) plus (15a) results in

$$V(q_{i,d+1}, d{+}1) = \omega_{i,d+1}\frac{a_{i,d+1}^{1-\alpha_i}}{1-\alpha_i} - \phi_i a_{i,d+1} + \sum_{t=1}^{d}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}}.$$

By taking the first-order derivative of the above equation, we find

$$a_{i,d+1}^\star = \left(\frac{\omega_{i,d+1}}{\phi_i}\right)^{\frac{1}{\alpha_i}}. \tag{17}$$

By substituting $a_{i,d+1}^\star$ back to $V(q_{i,d+1}, d{+}1)$, we thus derive

$$V^\star(q_{i,d+1}, d+1) = \sum_{t=1}^{d+1}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}}$$

for

$$\begin{aligned} q_{i,d} = q_{i,d+1} - a_{i,d+1}^\star &> \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}} \\ \Rightarrow \qquad q_{i,d+1} &> \sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, \end{aligned} \tag{18}$$

which verifies (15a).

2) $\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \le q_{i,d} \le \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}$ and $q_{i,d+1} > 0$.

By adding (1a) to (15b), we find:

$$\begin{aligned} V(q_{i,d+1}, d+1) &= \omega_{i,d+1}\frac{a_{i,d+1}^{1-\alpha_i}}{1-\alpha_i} - \phi_i q_{i,d+1} \\ &+ \omega_{i,d}\left(\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d}}\right)^{\frac{1}{\alpha_i}}\right)^{\alpha_i}\frac{(q_{i,d+1} - a_{i,d+1})^{1-\alpha_i}}{1-\alpha_i}. \end{aligned}$$

By taking the first-order derivative of the above equation, we find

$$\begin{aligned} \omega_{i,d+1}a_{i,d+1}^{\star-\alpha_i} &= \omega_{i,d}\left(\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d}}\right)^{\frac{1}{\alpha_i}}\right)^{\alpha_i}(q_{i,d+1} - a_{i,d+1}^\star)^{-\alpha_i} \\ \Rightarrow \frac{q_{i,d+1}}{a_{i,d+1}^\star} - 1 &= \left(\frac{\omega_{i,d}}{\omega_{i,d+1}}\right)^{\frac{1}{\alpha_i}}\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d}}\right)^{\frac{1}{\alpha_i}} = \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d+1}}\right)^{\frac{1}{\alpha_i}} \\ \Rightarrow \qquad a_{i,d+1}^\star &= \frac{q_{i,d+1}}{\sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\omega_{i,d+1}}\right)^{\frac{1}{\alpha_i}}}. \end{aligned} \tag{19}$$

By substituting $a_{i,d+1}^\star$ back into $V(q_{i,d+1}, d+1)$, we thus derive

$$\begin{aligned} &V^\star(q_{i,d+1}, d+1) \\ &= \omega_{i,d+1}\frac{a_{i,d+1}^{\star 1-\alpha_i}}{1-\alpha_i} - \phi_i q_{i,d+1} \\ &\quad + \omega_{i,d}\left(\frac{\omega_{i,d}}{\omega_{i,d+1}}\right)^{\frac{1}{\alpha_i}-1}\sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\omega_{i,d}}\right)^{\frac{1}{\alpha_i}}\frac{a_{i,d+1}^{\star 1-\alpha_i}}{1-\alpha_i} \\ &= \omega_{i,d+1}\left(\sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\omega_{i,d+1}}\right)^{\frac{1}{\alpha_i}}\right)^{\alpha_i}\frac{q_{i,d+1}^{1-\alpha_i}}{1-\alpha_i} - \phi_i q_{i,d+1}, \end{aligned}$$

and $q_{i,d} = q_{i,d+1} - a_{i,d+1}^\star$ needs to satisfy the necessary condition for this case, implying that $q_{i,d+1}$ satisfies

$$\sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \le q_{i,d+1} \le \sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, \tag{20}$$

which verifies (15b).

3) $q_{i,d} < \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}$ and $q_{i,d+1} \ge 0$.

Both (1a) plus (15c) and (1b) plus (15d) result in

$$\begin{aligned} V(q_{i,d+1}, d{+}1) &= \omega_{i,d+1}\frac{a_{i,d+1}^{1-\alpha_i}}{1-\alpha_i} - \phi_i a_{i,d+1} + \pi(q_{i,d+1} - a_{i,d+1}) \\ &+ \sum_{t=1}^{d}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}} \end{aligned}$$

By taking the first-order derivative of the above equation, we find

$$a_{i,d+1}^\star = \left(\frac{\omega_{i,d+1}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}. \tag{21}$$

By substituting $a_{i,d+1}^\star$ back to $V(q_{i,d+1}, d{+}1)$, we thus derive

$$V^\star(q_{i,d+1}, d+1) = \sum_{t=1}^{d+1}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}} + \pi q_{i,d+1}$$

for

$$\begin{aligned} q_{i,d} = q_{i,d+1} - a_{i,d+1}^\star &< \sum_{t=1}^{d}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \text{ and } q_{i,d+1} \ge 0 \\ \Rightarrow \qquad 0 \le q_{i,d+1} &< \sum_{t=1}^{d+1}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}, \end{aligned}$$

which verifies (15c).

4) $q_{i,d+1} < 0$, meaning $q_{i,d} < 0$ as well.

Finally, (1c) plus (15d) results in

$$\begin{aligned} V(q_{i,d+1}, d+1) &= \omega_{i,d+1}\frac{a_{i,d+1}^{1-\alpha_i}}{1-\alpha_i} - \phi_i a_{i,d+1} - \pi a_{i,d+1}) \\ &+ \sum_{t=1}^{d}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}} \end{aligned}$$

Following the similar steps that we did for the previous cases, we also find $a_{i,d+1}^\star = \left(\frac{\omega_{i,d+1}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}$ for this case and can easily verifies (15d).

Therefore, we can conclude that $V^\star(q_{i,d}, d)$ formulated in (3) can be represented by (15).

Furthermore, case 1 (where (17) and (18) lead to (4a)) and case 2 (where (19) and (20) leads to (4c)) are both exclusive with other cases, since the condition satisfied for d also satisfies for $d+1$. Although a user's utility may switch from (15c) at $d{+}1$ to (15d) at d, meaning that case 3 and case 4 are inclusive, they together lead to (4b). Also, these conditions together include all real numbers. Thus, we conclude that the functions contained in (4) are exclusive with each other. ■

C. Proof of Corollary 1

Proof: Based on the result in Proposition 1, we further simplify $a^\star_{i,d}$. We have found that $a^\star_{i,d}$ can be expressed by (17) (case 1 above), (19) (case 2 above), and (21) (case 3 and 4 above) respectively under different conditions of $q_{i,d}$. Since only (19) is in terms of $q_{i,d}$, we simplify it by combining with $q_{i,d} = A - \sum_{t=d+1}^{D} a^\star_{i,t}$:

$$\begin{aligned} &\frac{a^\star_{i,d}}{\omega_{i,d}^{\frac{1}{\alpha_i}}} \sum_{t=1}^{d} \omega_{i,t}^{\frac{1}{\alpha_i}} = q_{i,d} = A - \sum_{t=d+1}^{D} a^\star_{i,t} \\ \Rightarrow \quad &\frac{a^\star_{i,d}}{\omega_{i,d}^{\frac{1}{\alpha_i}}} \sum_{t=1}^{d-1} \omega_{i,t}^{\frac{1}{\alpha_i}} = A - \sum_{t=d}^{D} a^\star_{i,t} \end{aligned} \tag{22}$$

Shifting one time slot, we have for $a^\star_{i,d-1}$ that

$$\frac{a^\star_{i,d-1}}{\omega_{i,d-1}^{\frac{1}{\alpha_i}}} \sum_{t=1}^{d-1} \omega_{i,t}^{\frac{1}{\alpha_i}} = A - \sum_{t=d}^{D} a^\star_{i,t} \tag{23}$$

Combining (22) and (23) leads to

$$\frac{a^\star_{i,d}}{a^\star_{i,d-1}} = \frac{\omega_{i,d}^{\frac{1}{\alpha_i}}}{\omega_{i,d-1}^{\frac{1}{\alpha_i}}} \quad \Rightarrow \quad \frac{a^\star_{i,D}}{a^\star_{i,d}} = \frac{\omega_{i,D}^{\frac{1}{\alpha_i}}}{\omega_{i,d}^{\frac{1}{\alpha_i}}} \tag{24}$$

If we solve (24) together with $a^\star_{i,D}$ in (23) such that $q_{i,D} = A$, we can derive the expression in (5c).

By substituting $a^\star_{i,D}$ in (5c) back into $\sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \leq q_{i,d} \leq \sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}$, we further simplified this inequality as follows:

$$\begin{aligned} &\sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \leq A - \sum_{t=d+1}^{D} a^\star_{i,t} \leq \sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}} \\ \Rightarrow \; &\sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \leq A \frac{\sum_{t=1}^{d} \omega_{i,t}^{\frac{1}{\alpha_i}}}{\sum_{t=1}^{D} \omega_{i,t}^{\frac{1}{\alpha_i}}} \leq \sum_{t=1}^{d} \left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}} \end{aligned}$$

which is equivalent to the domain in (5c). Since the functions contained in (4) are exclusive with each other (cf. Appendix B), we can conclude the result as in (5). ■

D. Proof of Proposition 2

Proof: By taking the first-order derivative of the right-hand side of (7) in terms of $\tilde{a}_{i,d}$, we have

$$\frac{\partial V^\star(\tilde{q}_{i,d}, d)}{\partial \tilde{a}_{i,d}} = \frac{\partial v_{i,d}(\tilde{a}_{i,d})}{\partial \tilde{a}_{i,d}} - \mathbb{E}\left(\frac{\partial V^\star(\tilde{q}_{i,d-1}, d-1)}{\partial \tilde{q}_{i,d-1}}\right) \tag{25}$$

and setting it to zero yields the first-order condition of (7):

$$\frac{\partial v_{i,d}(\tilde{a}^\star_{i,d})}{\partial \tilde{a}^\star_{i,d}} = \mathbb{E}\left(\frac{\partial V^\star(\tilde{q}_{i,d-1}, d-1)}{\partial \tilde{q}_{i,d-1}}\right). \tag{26}$$

Applying the Benveniste-Scheinkman formula to (7) gives us:

$$\begin{aligned} \frac{\partial V^\star(\tilde{q}_{i,d}, d)}{\partial \tilde{q}_{i,d}} &= \frac{\partial v_{i,d}(\tilde{a}^\star_{i,d})}{\partial \tilde{q}_{i,d}} + \mathbb{E}\left(\frac{\partial V^\star(\tilde{q}_{i,d-1}, d-1)}{\partial \tilde{q}_{i,d-1}} \frac{\partial \tilde{q}_{i,d-1}}{\partial \tilde{q}_{i,d}}\right) \\ &= \frac{\partial v_{i,d}(\tilde{a}_{i,d})}{\partial \tilde{q}_{i,d}} + \frac{\partial v_{i,d}(\tilde{a}^\star_{i,d})}{\partial \tilde{a}^\star_{i,d}}\left(1 - \frac{\partial \tilde{a}^\star_{i,d}}{\partial \tilde{q}_{i,d}}\right) = \frac{\partial v_{i,d}(\tilde{a}^\star_{i,d})}{\partial \tilde{a}^\star_{i,d}} \end{aligned}$$

where the second equality holds due to (26). Analogously, we can write for $a^\star_{i,d-1}$ that

$$\frac{\partial V^\star(\tilde{q}_{i,d-1}, d-1)}{\partial \tilde{q}_{i,d-1}} = \frac{\partial v_{i,d-1}(\tilde{a}^\star_{i,d-1})}{\partial \tilde{a}^\star_{i,d-1}}. \tag{27}$$

Combining (26) and (27) leads us to (8). ■

E. Proof of Corollary 2

Proof: First of all, for all $\tilde{a}_{i,d}$, we have

$$\frac{\partial v_{i,d}(\tilde{a}_{i,d})}{\partial \tilde{a}_{i,d}} = \begin{cases} \omega_{i,d}\tilde{a}_{i,d}^{-\alpha_i} - \phi_i, \; \tilde{a}_{i,d} \leq \tilde{q}_{i,d}, & \text{(28a)} \\ \omega_{i,d}\tilde{a}_{i,d}^{-\alpha_i} - \pi - \phi_i, \; \tilde{a}_{i,d} > \tilde{q}_{i,d}. & \text{(28b)} \end{cases}$$

Then, we apply mathematical induction again to prove Corollary 2.

Before deriving the base case $d = 2$, we start by looking into the first-order derivative of $v_{i,1}(\tilde{a}_{i,1})$:

$$\frac{\partial v_{i,1}(\tilde{a}^\star_{i,1})}{\partial \tilde{a}^\star_{i,1}} = \begin{cases} 0, \; \tilde{a}^\star = \left(\frac{\omega_{i,1}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, \; \tilde{q}_{i,1} > \left(\frac{\omega_{i,1}}{\phi_i}\right)^{\frac{1}{\alpha_i}}, & \text{(29a)} \\ 0, \; \tilde{a}^\star_{i,1} = \left(\frac{\omega_{i,1}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}, \; \tilde{q}_{i,1} < \left(\frac{\omega_{i,1}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}, & \text{(29b)} \\ \frac{\partial v_{i,1}(\tilde{a}_{i,1})}{\partial \tilde{a}_{i,1}}\Big|_{\tilde{a}^\star_{i,1} = \tilde{q}_{i,1}}, \; \text{otherwise.} & \text{(29c)} \end{cases}$$

We note that (29) is not continuous, i.e., the utility function in (1) is not continuously differentiable. This is reflected in (29) that for $\left(\frac{\omega_{i,1}}{\pi+\phi_i}\right)^{1/\alpha_i} \leq \tilde{q}_{i,1} \leq \left(\frac{\omega_{i,1}}{\phi_i}\right)^{1/\alpha_i}$, we have

$$\frac{\partial v_{i,1}(\tilde{a}^\star_{i,1})}{\partial \tilde{a}^\star_{i,1}}\Big|_{\tilde{a}^\star_{i,1} = \tilde{q}_{i,1}} = \begin{cases} \omega_{i,1}\tilde{q}_{i,1}^{-\alpha_i} - \phi_i, \; \lim\limits_{\tilde{a}^\star_{i,1} \to \tilde{q}^-_{i,1}} = \tilde{q}_{i,1}, & \text{(30a)} \\ \omega_{i,d}\tilde{q}_{i,1}^{-\alpha_i} - \pi - \phi_i, \; \lim\limits_{\tilde{a}^\star_{i,1} \to \tilde{q}^+_{i,1}} = \tilde{q}_{i,1}, & \text{(30b)} \end{cases}$$

where (30a) identifies the case that $\tilde{a}^\star_{i,1}$ approaches to $\tilde{q}_{i,1}$ from the left side, while (30b) shows the other direction. Though discontinuous, we observe from (29) and (30) that $\tilde{q}_{i,1} = \tilde{q}_{i,2} - \tilde{a}_{i,2} < 0$ falls only into the case in (29b), and thus $\frac{\partial v_{i,2}(\tilde{a}^\star_{i,2})}{\partial \tilde{a}^\star_{i,2}} = \mathbb{E}\big(\frac{\partial v_{i,1}(\tilde{a}^\star_{i,1})}{\partial \tilde{a}^\star_{i,1}}\big) = 0$ as integral over 0 yields 0. Therefore, $\tilde{a}^\star_{i,2} = \big(\frac{\omega_{i,1}}{\pi+\phi_i}\big)^{1/\alpha_i}$ when $\tilde{q}_{i,2} < \tilde{a}_{i,2} = \big(\frac{\omega_{i,1}}{\pi+\phi_i}\big)^{1/\alpha_i}$. On the other hand, $\tilde{q}_{i,1} = \tilde{q}_{i,2} - \tilde{a}_{i,2} \geq 0$ falls into all cases in (29a), (29b), and (29c), and thus $\frac{\partial v_{i,2}(\tilde{a}^\star_{i,2})}{\partial \tilde{a}^\star_{i,2}} = \mathbb{E}\big(\frac{\partial v_{i,1}(\tilde{a}^\star_{i,1})}{\partial \tilde{a}^\star_{i,1}}\big) \in [0, \mathbb{E}(\omega_{i,1})\tilde{q}_{i,1}^{-\alpha_i} - \phi_i]$: $\omega_{i,2}\tilde{a}_{i,2}^{\star -\alpha_i} - \phi_i \geq 0$ leads to $\tilde{a}^\star_{i,2} \leq$

$\left(\frac{\omega_{i,2}}{\phi_i}\right)^{1/\alpha_i}$, and $\omega_{i,2}\tilde{a}_{i,2}^{\star-\alpha_i} - \phi_i \le \mathbb{E}(\omega_{i,1})\tilde{q}_{i,1}^{-\alpha_i} - \phi_i$ leads to $\tilde{a}_{i,2}^{\star} \ge \frac{\omega_{i,2}^{1/\alpha_i}\tilde{q}_{i,2}}{\omega_{i,2}^{1/\alpha_i}+E_1^{1/\alpha_i}}$. Thus, (8) holds for $d = 2$.

Assuming (8) is true, we have $\mathbb{E}\big(\frac{\partial v_{i,d}(\tilde{a}_{i,d}^{\star})}{\partial \tilde{a}_{i,d}^{\star}}\big) = 0$ if $\tilde{q}_{i,d} < \big(\frac{\omega_{i,d}}{\pi+\phi_i}\big)^{1/\alpha_i}$, and otherwise, $0 \le \mathbb{E}\big(\frac{\partial v_{i,d}(\tilde{a}_{i,d}^{\star})}{\partial \tilde{a}_{i,d}^{\star}}\big) \le \mathbb{E}\Big(\omega_{i,d}\big(\frac{\omega_{i,d}^{1/\alpha_i}\tilde{q}_{i,d}}{\omega_{i,d}^{1/\alpha_i}+E_{d-1}^{1/\alpha_i}}\big)^{-\alpha_i} - \phi_i\Big) = E_d\tilde{q}_{i,d}^{-\alpha_i} - \phi_i$. To solve $\frac{\partial v_{i,d+1}(\tilde{a}_{i,d+1}^{\star})}{\partial \tilde{a}_{i,d+1}^{\star}} = \mathbb{E}\big(\frac{\partial v_{i,d+1}(\tilde{a}_{i,d}^{\star})}{\partial \tilde{a}_{i,d}^{\star}}\big)$, we shift (29) by one period for $\tilde{a}_{i,d+1}$. Similarly, $\tilde{a}_{i,d+1} > \tilde{q}_{i,d+1}$ in (28b) only requires to consider the case of $\tilde{q}_{i,d} < \big(\frac{\omega_{i,d}}{\pi+\phi_i}\big)^{1/\alpha_i}$, yielding $\frac{\partial v_{i,d+1}(\tilde{a}_{i,d+1}^{\star})}{\partial \tilde{a}_{i,d+1}^{\star}} = 0 \Rightarrow \tilde{a}_{i,d+1} = \big(\frac{\omega_{i,d+1}}{\pi+\phi_i}\big)^{1/\alpha_i}$ for $\tilde{q}_{i,d+1} < \tilde{a}_{i,d+1}^{\star} = \big(\frac{\omega_{i,d+1}}{\pi+\phi_i}\big)^{1/\alpha_i}$. If $\tilde{a}_{i,d+1} \le \tilde{q}_{i,d+1}$, $0 \le \frac{\partial v_{i,d+1}(\tilde{a}_{i,d+1}^{\star})}{\partial \tilde{a}_{i,d+1}^{\star}} \le E_d\tilde{q}_{i,d}^{-\alpha_i} - \phi_i$ leads us to $\frac{\omega_{i,d+1}^{1/\alpha_i}\tilde{q}_{i,d+1}}{\omega_{i,d+1}^{1/\alpha_i}+E_d^{1/\alpha_i}} \le \tilde{a}_{i,d+1}^{\star} \le \big(\frac{\omega_{i,d+1}}{\phi_i}\big)^{\frac{1}{\alpha_i}}$. Therefore, as we find that $\tilde{a}_{i,d+1}^{\star}$ also satisfies (8), we can conclude that (8) holds for all $\tilde{a}_{i,d}^{\star}$, $d = D, \ldots, 2$. ■

F. Proof of Corollary 3

Proof: When there is no time cost, i.e., $\phi = 0$ and overage fee is infinite (none of the users would consume overage data), i.e., $\pi = +\infty$, the utility function can be reduced to $v_{i,d}(a_{i,d}) = \omega_{i,d}\frac{a_{i,d}^{1-\alpha_i}}{1-\alpha_i}$. In this special case, we can again apply mathematical induction to solve (8) and obtain $\tilde{a}_{i,d}^{\star} = \frac{\omega_{i,d}^{1/\alpha_i}\tilde{q}_{i,d}}{\omega_{i,d}^{1/\alpha_i}+E_{d-1}^{1/\alpha_i}}$.

We now examine E_d^{1/α_i}. According to Jensen's inequality, the convexity of $g(\omega_{i,d}) = (\omega_{i,d}^{1/\alpha_i} + E_{d-1})^{\alpha_i}$ gives us

$$E_d = \mathbb{E}\left(\left(\omega_{i,d}^{\frac{1}{\alpha_i}} + E_{d-1}\right)^{\alpha_i}\right) \ge \left(\mathbb{E}(\omega_{i,d})^{\frac{1}{\alpha_i}} + E_{d-1}\right)^{\alpha_i}.$$

Applying this iteratively gives us

$$E_d^{\frac{1}{\alpha_i}} \ge \sum_{t=1}^{d} \mathbb{E}(\omega_{i,t})^{\frac{1}{\alpha_i}}. \tag{31}$$

Again, due to the concavity of $\tilde{a}_{i,d}^{\star} = \frac{\omega_{i,d}^{1/\alpha_i}\tilde{q}_{i,d}}{\omega_{i,d}^{1/\alpha_i}+E_{d-1}^{1/\alpha_i}}$ in terms of $\omega_{i,d}^{1/\alpha_i}$, we have the following derivations:

$$\mathbb{E}(\tilde{a}_{i,d}^{\star}) = \mathbb{E}\left(\frac{\omega_{i,d}^{\frac{1}{\alpha_i}}\tilde{q}_{i,d}}{\omega_{i,d}^{\frac{1}{\alpha_i}} + E_{d-1}^{\frac{1}{\alpha_i}}}\right) \le \frac{\mathbb{E}\big(\omega_{i,d}^{\frac{1}{\alpha_i}}\big)}{\mathbb{E}\big(\omega_{i,d}^{\frac{1}{\alpha_i}}\big) + E_{d-1}^{\frac{1}{\alpha_i}}}\tilde{q}_{i,d}$$
$$\overset{(a)}{\le} \frac{\mathbb{E}\big(\omega_{i,d}^{\frac{1}{\alpha_i}}\big)}{\mathbb{E}\big(\omega_{i,d}^{\frac{1}{\alpha_i}}\big) + \sum_{t=1}^{d-1}\mathbb{E}(\omega_{i,t})^{\frac{1}{\alpha_i}}}\tilde{q}_{i,d},$$

where (a) is due to (31). When setting $\alpha_i = 1$, we obtain $\tilde{a}_{i,d}^{\star} \le \tilde{q}_{i,d}/d$. ■

G. Proof of Example 2

Proof: We now use induction on d to show that:

$$\begin{aligned} &V^{\star}(q_{i,d}, d) \\ &= d\mu \log q_{i,d} + \mathbb{E}\big((d-1)\omega_{i,d}\log\omega_{i,d}\big) \\ &\quad + \sum_{i=1}^{d-1} i\mu\log(i\mu) - \sum_{i=1}^{d-1}\mathbb{E}\big((\omega_{i,d}+i\mu)\log(\omega_{i,d}+i\mu)\big). \end{aligned}$$

For $d = 1$, we have $a_{i,1}^{\star} = q_{i,1}$ and thus $V_1^{\star}(q_{i,1}, 1) = \mu\log q_{i,1}$. Now suppose that the result holds for day $d-1$. Since $a_{i,d}^{\star} = \frac{\omega_{i,d}q_{i,d}}{\omega_{i,d}+(d-1)\mu}$ and $q_{i,d} - a_{i,d}^{\star} = \frac{(d-1)\mu q_{i,d}}{\omega_{i,d}+(d-1)\mu}$, our inductive hypothesis implies:

$$\begin{aligned} &V^{\star}(q_d, d) \\ &= \mathbb{E}\left(\omega_{i,d}\log\frac{\omega_{i,d}q_{i,d}}{\omega_{i,d}+(d-1)\mu}\right) \\ &\quad + \mathbb{E}\left((d-1)\mu\log\frac{(d-1)\mu q_{i,d}}{\omega_{i,d}+(d-1)\mu}\right) \\ &\quad + \mathbb{E}\left((d-2)\omega_{i,d}\log\omega_{i,d}\right) + \sum_{i=1}^{d-2} i\mu\log(i\mu) \\ &\quad - \sum_{i=1}^{d-2}\mathbb{E}\left((\omega_{i,d}+i\mu)\log(\omega_{i,d}+i\mu)\right) \\ &= d\mu\log q_{i,d} + \mathbb{E}\left((d-1)\omega_{i,d}\log\omega_{i,d}\right) + \sum_{i=1}^{d-1} i\mu\log(i\mu) \\ &\quad - \sum_{i=1}^{d-1}\mathbb{E}\left((\omega_{i,d}+i\mu)\log(\omega_{i,d}+i\mu)\right). \end{aligned}$$

■

H. Proof of Proposition 3

Proof: Supposing all data plans offer the same amount A charged at the same P if $A < +\infty$ and $P > 0$, we first prove the relationship between $U_u(\vec{\mathbf{a}}_i^{\star})$, $U(\vec{\mathbf{a}}_i^{\star})$, and $U_p(\vec{\mathbf{a}}_i^{\star})$, as it is independent from P. It is easy to see that $g(A) = \big(\sum_{t=1}^{D}\omega_{i,t}^{1/\alpha_i}\big)^{\alpha_i}\frac{A^{1-\alpha_i}}{1-\alpha_i} - \phi_i A - P$ increases in A if $A \le \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$. We then have $g(A) \le g(A)\big|_{A=\sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}} = \sum_{t=1}^{D}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}\phi_i^{1-\frac{1}{\alpha_i}} - P$ if $A \le \sum_{t=1}^{D}(\omega_{i,t}/\phi_i)^{1/\alpha_i}$. Hence, if $A \le \sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i}$, we have $U_u(\vec{\mathbf{a}}_i^{\star}) \ge U(\vec{\mathbf{a}}_i^{\star})$; otherwise, $U_u(\vec{\mathbf{a}}_i^{\star}) = U(\vec{\mathbf{a}}_i^{\star})$ as they have the same expression. Similarly, we also have $g(A) \le g(A)\big|_{A=\sum_{t=1}^{D}\left(\omega_{i,t}/(\pi+\phi_i)\right)^{1/\alpha_i}} = \sum_{t=1}^{D}\frac{\alpha_i}{1-\alpha_i}\omega_{i,t}^{\frac{1}{\alpha_i}}(\pi+\phi_i)^{1-\frac{1}{\alpha_i}} - P$ if $A \le \sum_{t=1}^{D}\big(\omega_{i,t}/(\pi+\phi_i)\big)^{1/\alpha_i}$. It is obvious to see that $U_p(\vec{\mathbf{a}}_i^{\star}) \le U(\vec{\mathbf{a}}_i^{\star})$ in this case; and otherwise, they equal each other.

Since we have obtained the tendency of $U_u(\vec{\mathbf{a}}_i^{\star})$, $U(\vec{\mathbf{a}}_i^{\star})$, and $U_p(\vec{\mathbf{a}}_i^{\star})$, we next need to compare each of them to $U_s(\vec{\mathbf{a}}_i^{\star})$. After doing so, we have $U_s(\vec{\mathbf{a}}_i^{\star}) > U_u(\vec{\mathbf{a}}_i^{\star})$ if $P > P_L^v$, $U_s(\vec{\mathbf{a}}_i^{\star}) > U(\vec{\mathbf{a}}_i^{\star})$ if $P > P_M^v$, and $U_s(\vec{\mathbf{a}}_i^{\star}) > U_p(\vec{\mathbf{a}}_i^{\star})$ if $P > P_S^v$, where P_L^v, P_M^v, and P_S^v are given below Table II. Summarizing the above discussion leads us to Table II. ■

TABLE II
COMPARISON OF USER UTILITY WITH DIFFERENT DATA PLANS.

Conditions	$A < \sum_{t=1}^{D} \left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}}$	$\sum_{t=1}^{D} \left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} \leq A \leq \sum_{t=1}^{D} \left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}$	$A > \sum_{t=1}^{D} \left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}$
$P > P_L^v$	$U_s > U_u \geq U \geq U_p$	$U_s > U_u \geq U = U_p$	$U_s > U_u = U = U_p$
$P_M^v < P \leq P_L^v$	$U_u \geq U_s > U \geq U_p$	$U_u \geq U_s > U = U_p$	$U_u = U = U_p \geq U_s$
$P_S^v < P \leq P_M^v$	$U_u \geq U \geq U_s > U_p$	$U_u \geq U = U_p \geq U_s$	
$P \leq P_S^v$	$U_u \geq U \geq U_p \geq U_s$		

$P_L^v = \sum_{t=1}^{D} \frac{\alpha_i}{1-\alpha_i} \omega_{i,t}^{\frac{1}{\alpha_i}} \left(\phi_i^{1-\frac{1}{\alpha_i}} - (\pi+\phi_i)^{1-\frac{1}{\alpha_i}}\right)$
$P_M^v = \pi A$
$P_S^v = \left(\sum_{t=1}^{D} \omega_{i,t}^{\frac{1}{\alpha_i}}\right)^{\alpha_i} \frac{A^{1-\alpha_i}}{1-\alpha_i} - \phi_i A - \sum_{t=1}^{D} \frac{\alpha_i}{1-\alpha_i} \omega_{i,t}^{\frac{1}{\alpha_i}} (\pi+\phi_i)^{1-\frac{1}{\alpha_i}}$

TABLE III
COMPARISON OF ISP PROFIT WITH DIFFERENT DATA PLANS.

Conditions	$N_M \neq 0, N_L \neq 0$	$N_M \neq 0, N_L = 0$	$N_M = 0, N_L = 0$
$P > P_L^r$	$R > R_p > R_u > R_s$	$R = R_p > R_u > R_s$	$R = R_p = R_u > R_s$
$P_M^r < P \leq P_L^r$	$R > R_p > R_s \geq R_u$	$R = R_p > R_s \geq R_u$	$R_s \geq R = R_p = R_u$
$P_S^r < P \leq P_M^r$	$R > R_s \geq R_p > R_u$	$R_s \geq R = R_p > R_u$	
$P \leq P_S^r$	$R_s \geq R > R_p > R_u$		

$P_L^r = \frac{1}{N}\Big((\pi-\sigma)\sum_{i\in\mathcal{N}}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} + \sigma\sum_{i\in\mathcal{N}}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}}\Big)$
$P_M^r = \frac{1}{N}\Big((\pi-\sigma)\sum_{i\in\mathcal{N}}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} + \sigma\sum_{i\in\mathcal{N}_S}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}} + \sigma(N_M+N_L)A\Big)$
$P_S^r = \frac{1}{N}\Big((\pi-\sigma)\sum_{i\in\mathcal{N}_S\cup\mathcal{N}_M}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\pi+\phi_i}\right)^{\frac{1}{\alpha_i}} + \sigma\sum_{i\in\mathcal{N}_S}\sum_{t=1}^{D}\left(\frac{\omega_{i,t}}{\phi_i}\right)^{\frac{1}{\alpha_i}} + (\sigma N_M + \pi N_L)A\Big)$

TABLE IV
COMPARISON OF USER UTILITY AND ISP PROFIT IF USERS SWITCH TO A DIFFERENT DATA PLAN.

	Current data plan	Monthly usage (MB)			User utility			ISP profit ($)		
		Unlimited	Quota	Usage-based	Unlimited	Quota	Usage-based	Unlimited	Quota	Usage-based
User 1	Unlimited	10628.97	9459.76	9459.76	**6632.52**	6462.03	6462.03	29.34	188.25	188.25
User 2	Unlimited	1379.85	1314.72	1314.72	2164.49	**2167.55**	**2167.55**	29.78	26.16	26.16
User 3	Quota	574.33	574.33	561.94	2292.20	2302.20	**2310.83**	29.77	19.94	11.18
User 4	Quota	1104.56	1000.00	976.03	897.80	907.00	**907.04**	29.91	19.90	19.42
User 5	Unlimited	711.52	711.52	676.05	657.46	667.46	**673.59**	29.93	19.93	13.45
User 6	Unlimited	298.19	298.19	287.46	1082.66	**1092.66**	1106.80	29.89	19.97	5.72
User 7	Unlimited	63.93	63.93	63.20	162.48	172.48	**191.21**	29.98	19.99	1.2
User 8	Quota	393.75	393.75	388.16	11888.04	11898.04	**11910.22**	28.81	19.96	7.72
User 9	Quota	267.39	267.39	265.17	2440.52	2450.52	**2465.20**	29.76	19.97	5.2
User 10	Quota	462.02	462.02	450.64	12899.32	12909.32	**12920.19**	28.71	19.95	8.9
User 11	Quota	1828.41	1628.00	1628.00	**1839.91**	1835.41	1835.41	29.82	32.40	32.40
User 12	Quota	473.04	473.04	449.28	14213.93	14223.93	**14234.71**	28.58	19.95	8.94
User 13	Quota	1490.56	1424.46	1424.46	1411.16	**1412.02**	**1412.02**	29.86	28.35	28.35

I. Proof of Proposition 4

Proof: Similar to the proof of Proposition 3, we first show the relationship between $R(A,P,\pi)$, $R_p(A,P,\pi)$, and $R_u(A,P,\pi)$, as it is independent from P. We see that the profit that the ISP receives from a user is nondecreasing with user demand. When there is no modest and heavy users, we have $R(A,P,\pi) = R_p(A,P,\pi) = R_u(A,P,\pi)$ as they have the same expression. When there is no heavy users but some modest users, prepaid and quota with overage data plans lead to same usage for all users. Although all three data plans end up with the same revenue from users, the associated profit results in $R(A,P,\pi) = R_p(A,P,\pi) > R_u(A,P,\pi)$ due to more traffic cost for unlimited data plans. When there are some heavy users, profit of quota with overage data plans is more than that of prepaid data plans because of the overage charge.

From the relationship between $R(A,P,\pi)$, $R_p(A,P,\pi)$, and $R_u(A,P,\pi)$, we again compare each of them to $R_s(A,P,\pi)$ with the boundaries of P, P_L^r, P_M^r, and P_S^r, given below Table III. ■