Title
The Effect of Data Caps upon ISP Service Tier Design and Users

Journal
ACM Transactions on Internet Technology, 15(2)

Author
Jordan, Scott

Publication Date
2015-06-01

Peer reviewed

The Effect of Data Caps upon ISP Service Tier Design and Users

WEI DAI and SCOTT JORDAN, University of California, Irvine

We model the design and impact of Internet pricing plans with data caps. We consider a monopoly ISP that maximizes its profit by setting tier prices, tier rates, network capacity, data caps, and overage charges. We show that when data caps are used to maximize profit, a monopoly ISP will keep the basic tier price the same, increase the premium tier rate, and decrease the premium tier price and the basic tier rate. We give analytical and numerical results to illustrate the increase in ISP profit, and the corresponding changes in user tier choices, user surplus, and social welfare.

Categories and Subject Descriptors: C.2.2 [**Computer-Communication Networks**]: Network Protocols

General Terms: Economics, Management, Performance

Additional Key Words and Phrases: Charging, pricing, business models

ACM Reference Format:
Dai, W. and Jordan, S. 2015. The effect of data caps upon ISP tier design and users. *ACM Trans. Internet Technol.* 15, 2, Article 8 (June 2015), 28 pages.
DOI:http://dx.doi.org/10.1145/2774973

8

1. INTRODUCTION

In recent years, it has become common for wireless Internet Service Providers (ISPs) in the United States to place caps on the monthly usage of cellular data plans. Some wireline ISPs have also started placing caps on monthly usage of their broadband service offerings. The data caps often differ by the tier of the plan and are often in the range of 50GB to 500GB per month [Higginbotham 2012]. The consequences of exceeding the cap differ by ISP; some charge an overage charge per unit volume over the cap, some reduce the throughput of violators, and some issue warnings and/or upgrade subscribers to a higher tier.

ISPs commonly claim that caps benefit most users. They cite statistics [Sandvine 2012] that show that a small percentage of users consume a high percentage of network capacity, typically because these subscribers are heavy users of video streaming or file sharing [ACLP NYU 2012]. The ISPs claim that flat-rate pricing, in which all subscribers to a tier pay the same amount independent of usage, is unfair to the majority of users [Ford 2012]. They further claim that caps affect only a small percentage of heavy users [Bennett 2012] and that caps result in lower tier prices than would be offered without caps. Finally, ISPs claim that caps increase the incentive for ISPs to add capacity to the network, since the incremental capacity would benefit a broader set of users [Yu 2012].

This work is supported by the National Science Foundation. Portions of this work appear in Dai and Jordan [2013a, 2013b].
Authors' addresses: W. Dai and S. Jordan (corresponding author), Department of Computer Science, University of California, Irvine; email: sjordan@uci.edu.

1533-5399/2015/06-ART8 $15.00
DOI:http://dx.doi.org/10.1145/2774973

In contrast, many public interest groups claim that caps hurt most users. They claim that caps discourage the use of certain applications, including video streaming, and that this is often intended to protect the ISP's other services from competition [Kelsey 2012]. They further claim that caps encourage a climate of scarcity and that ISPs can increase their profits through the use of caps principally because of a lack of consumer choice in broadband providers [Hussain et al. 2012a]. Finally, public interest groups often claim that caps and their corresponding overage charges do not correspond to the cost for network capacity and that the use of caps may decrease an ISP's incentive to add capacity [Hussain et al. 2012b].

There is a vigorous debate over the use of caps. Some public interest groups have called for government oversight [Odlyzko et al. 2012]; in particular, some have asked the U.S. Federal Communications Commission (FCC) to investigate AT&T's broadband data caps [PK and NAF 2011]. A U.S. Senate bill, the Data Cap Integrity Act, would require the FCC to evaluate data caps to determine whether they reasonably limit network congestion without unnecessarily restricting Internet use [U.S. Congress 2012].

However, there is little related academic literature. A first set of papers seek to address the impact of data caps. Sen et al. [2012a] create an analytical framework to investigate user choices between shared data plans and separate data plans for their devices. User utility is modeled as a logarithmic function of data consumption. However, the impact of data caps upon ISPs, users, and social welfare is not the focus of this paper. Light users and heavy users are not differentiated in the model. Nor do the authors consider how data caps can affect network congestion or application performance, which is the reason ISPs give for adopting data caps. Minne [2012] explores ISP motivations for using data caps. It argues that heavy users are often profitable for ISPs and that data caps may be a method for ISPs to price gouge and to protect an ISP's video business. However, no mathematical model of usage is proposed. Lyons [2012] evaluates the merits of data caps and other usage-based pricing strategies. He argues that data caps can shift more network costs onto heavier Internet users and reduce network congestion. However, no models are proposed to quantitatively analyze the impact of data caps on the tiered pricing plans and the resulting impact on users. Waterman et al. [2012] express similar concerns that data caps may be anticompetitive behaviors in the online television market. However, caps are not the focus of the paper, and no model is proposed. Chetty et al. [2012] focus on the impact of data caps on subscribers. In a study of 12 households in South Africa, they find that uncertainties related to caps pose substantial challenges.

A second set of papers investigate user decisions between multiple Internet technologies. Joseph et al. [2007] and Sen et al. [2010] both propose models to study the adoption of two network technologies. In both papers, user utility is modeled as the sum of a standalone benefit (which depends on the values that individual users place on each network technology) and a network externality (which depends on the number of the subscribers to each technology). Joe-Wong et al. [2013] propose a similar model to study user adoption of a base wireless technology and a supplementary technology, in which an ISP can benefit from offloading traffic from the base technology to the supplementary technology. User utility is modeled as the sum of a standalone benefit and a congestion externality (which is a decreasing function of the number of the subscribers). It may be possible to use these approaches to consider data caps by modeling a basic service tier with a lower data cap as a standalone benefit, a premium tier with a higher data cap as a supplementary benefit, and network performance as a congestion externality. However, the utility models used in these papers are too general to capture many of the tradeoffs considered in our article. First, it would be difficult to model data consumption, which should be related to the standalone benefit. Second, it would be difficult to understand the relationship between applications used, time devoted, and data consumed, which should be related to user utility. Finally, explicit modeling of network performance for each set of applications provides more insight to usage and data caps than would be possible using a single characterization of a network or congestion externality.

A third set of papers investigate user choices between differentiated information goods by using existing decision models, statistical models, or pricing models. Eikebrokk and Sorebo [1999] propose a modified form of a Technology Acceptance Model (TAM) to predict user acceptance and use behaviors in a multiple-choice situation. Chaudhuri et al. [2005] use statistical models to analyze the impact of a variety of socioeconomic influences (e.g., income, education) on households' decisions to pay for basic Internet access. These generic decision models may be applied for user choices among multiple Internet service tiers with or without data caps. However, decision models cannot easily capture the interaction between an ISP and Internet users, compared with the traditional pricing models. For example, Bhargava and Choudhary [2008] use pricing models to answer when versioning (a form of second-degree price discrimination) is optimal for information goods, where one monopoly firm can segment the market by introducing additional lower-quality versions. Tiered Internet pricing plans can be viewed as a special case of versioning, and we adopt similar ideas in our ISP profit maximization problem. However, the focus of our model is on the impact of data caps rather than ISP profit maximization.

A fourth set of papers investigate flat-rate, usage-based (including data caps), and time-dependent pricing plans. Nabipay et al. [2011] propose an economic model of flat-rate pricing as a form of bundling, where some bandwidth hogs exist. Their analysis shows, for a monopoly service provider with negligible marginal costs, that flat-rate pricing almost always maximizes profit, even when there are some buyers with disproportionately large usage. While the authors suggest that the model might be used to explore the effects of data caps, the model presented does not incorporate data caps, nor does it model user willingness-to-pay or network congestion. Sen et al. provide a survey on time-dependent Internet pricing plans [2012b, 2013]. Some insights and findings related to data caps are collected from the survey results. However, again no mathematical model of usage is proposed. Jiang et al. [2008] investigate the design of time-dependent pricing plans. Although the method used to analyze the impact of time-dependent pricing on user behavior is useful for data caps analysis, data caps are not the focus of that paper.

In this article, we propose models to evaluate the impact of data caps upon subscribers and ISPs. The model includes the critical elements of both Internet architecture and economic motivation. First, rather than solely modeling user's subscription choices, we also model the decision by users of how much time to devote to each Internet application based in part on network performance. This additional detail is a critical factor to consider in analysis of data caps, since it affects both user data consumption and user willingness-to-pay. Second, rather than differentiating users either by their data consumption or their technology quality valuation, we differentiate users by the value they place on leisure time and the value they place on each application. This allows us to capture the relationship between user consumption and user willingness-to-pay. Third, rather than either assuming that network performance is fixed or modeling network congestion using a single externality variable, we explicitly consider the impact of network throughput and network delay on user willingness-to-pay for each set of Internet applications. Finally, rather than assuming users will consume a fixed fraction of the demand over the cap, we derive data consumption from the value that users place on Internet applications, the opportunity cost of their leisure time, the data caps, overage charges, and network performance.

This article is organized as follows. In Section 2, user utility is represented as a function of the time devoted per month to Internet applications, performance, and a user's relative utility for high bandwidth applications. User willingness-to-pay is expressed as utility minus the opportunity cost of the time devoted, which depends on the income of the user. We model ISP profit maximization by considering a monopolist that sets tier prices, tier rates, network capacity, data caps, and overage charges. In Section 3, we propose a simpler model of user decisions on tier choices by making some additional assumptions. In Section 4, we examine which users are affected by data caps based on the model. We show how users fall into five categories: those who do not subscribe to the Internet, those who subscribe to the basic tier, those who subscribe to the premium tier and are unaffected by a cap, those who subscribe to the premium tier and are capped but do not pay overage charges, and those who subscribe to the premium tier and pay overage charges. In Section 5, we analyze the impact of data caps on the tiered pricing plans. We examine a monopolist's use of caps and compare the optimal tier rates, tier prices, and network capacity without caps to the same quantities when caps are added. We show that when an ISP sets caps and overage charges to maximize its profit, it will increase the premium tier rate, decrease the premium tier price, decrease the basic tier rate, and approximately maintain the same basic tier price. In Section 6, we analyze the impact of caps, overage charges, and the corresponding tiered pricing plans on various subscribers. Users with low to medium valuations on video streaming and high incomes benefit from the data caps, while users with medium to high valuations on video streaming and low incomes are hurt by the data caps. Users with high valuations on video streaming and high income may benefit from or be hurt by data caps, depending on the shape of their utility functions. Finally, in Section 7, we give numerical results to illustrate how the tier rate, tier price, cap, and overage charges vary with the standard deviation in Internet usage amongst subscribers. We also illustrate the increase in ISP profit when caps are used, the corresponding change in user surplus, and the change in social welfare. The major contributions of this article are as follows.

— *Novel user utility models*. In contrast to previous literature that models user utility solely as a function of bandwidth, we propose novel utility functions that incorporate the time users devote to Internet applications and the opportunity cost of users' free time, thus differentiating light and heavy users on an economic basis.
— *First model of ISP cap design*. We present what we believe is the first model in the academic literature of how an ISP may set data caps and overage charges.
— *First mathematical model of the impact of caps on pricing plans and subscribers*. We characterize how tier rates and tier prices change in the presence of data caps, and which users benefit from or are hurt by caps.

2. CAP MODEL FORMULATION

In this section, we introduce utility functions for subscriber usage, and we introduce ISP tier and cap models for profit maximization. We consider two interconnected problems separated by timescale. On a timescale of days, broadband Internet subscribers choose how much time to devote to Internet applications. On a timescale of months, subscribers choose what tier to subscribe to, and ISPs choose tier rates, tier prices, data caps, overage charges, and network capacity.

2.1. Short-Term Model

The dominant applications on North American fixed-access broadband Internet access networks are video streaming, Web browsing, and peer-to-peer (p2p) file sharing, which together account for approximately 85% of download traffic volume

[Sandvine 2012]. Real-time entertainment traffic consists almost exclusively of video streaming. For purposes of analysis, we split p2p into two subsets: p2p streaming, which we aggregate with other video streaming, and p2p file sharing, which we aggregate with Web browsing. Although email is an important component of users' willingness-to-pay, it is an insignificant burden upon the network, and we similarly aggregate it with other file-sharing applications into Web browsing. We thus focus in the remainder of this article on two applications: Web browsing and video streaming.

Web browsing utility is commonly modeled as an increasing concave function of throughput. However, users' utilities also depend on how much Web browsing they do. Define t_i^b as the time (in seconds per month) that user i devotes to Web browsing, consisting of t_i^r time per month reading webpages and the time spent on waiting for them to download. We posit that the perceived utility by user i for Web browsing should be a function U_i^b of the number of hours devoted to Web browsing per month, the performance of Web browsing, and a user's relative utility for Web browsing. Utility is an increasing concave function $V^b(t_i^r)$ of the time devoted to it, independent of the user. With respect to performance, Web browsing is an elastic application, and thus performance is often measured by throughput. However, a user's observation of Web browsing performance consists of the download times of webpages, rather than direct observation of throughput, and thus the ratio $r_i^b = t_i^r/t_i^b$ is a more direct measurement of the Web browsing performance; it will be incorporated into a user's willingness-to-pay when we consider a user's valuation of time. User i's utility for Web browsing relative to other users is modeled using a scale factor v_i^b. The interaction between these factors has not been studied; we model user i's utility for Web browsing (in dollars per month) as

$$U_i^b = v_i^b V^b\left(t_i^b r_i^b\right).$$

Similarly, we posit that the perceived utility by user i for video streaming should be a function U_i^s of the time devoted to video streaming per month, the performance of video streaming, and a user's relative utility for video streaming. Denote t_i^s as the time (in seconds per month) that user i devotes to video streaming; normal economic assumptions are that a user's utility is an increasing concave function $V^s\left(t_i^s\right)$ of time devoted [Gerber and Pafumi 1998]. With respect to performance, video streaming is commonly classified as a semi-elastic application; we thus model a component of user utility by a sigmoid function $Q^s\left(x_i^s\right)$ of the throughput x_i^s (in bits per second) experienced by video streaming applications [Weber and Veeraraghavan 2007], normalized so that $Q^s(\infty) = 1$. User i's utility for video streaming relative to other users is modeled using a scale factor v_i^s.The interaction between these three factors has not been studied; we model user i's utility for video streaming (in dollars per month) as

$$U_i^s = v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right).$$

User i's willingness-to-pay for Web browsing and video streaming also depends on how the user values leisure time. The scale factors v_i^b and v_i^s should be increasing with this value. However, the time devoted to video streaming is also likely to be viewed as an opportunity cost. Denote p_i^t as the opportunity cost (in dollars per second) of user i's time, which is usually estimated to be between 20 to 50 percent of user i's income [Cesario 1976]. We model user i's willingness-to-pay for Web browsing and video steaming (in dollars per month) as

$$W_i^b = U_i^b - p_i^t t_i^b, W_i^s = U_i^s - p_i^t t_i^s. \tag{1}$$

Most ISPs have started designing and marketing tiers on the basis of the applications for which they are intended. We focus here on the decision by a user whether to subscribe to the tier designed for video streaming (hereafter referred to as the *premium tier*) or to a lower tier (hereafter referred to as the *basic tier*). Denote the prices of the basic tier and the premium tier by P_1 and P_2, respectively. Denote the caps (i.e., the maximum allowed number of bytes downloaded per month without incurring an overage charge) of the basic tier and the premium tier by C_1 and C_2, respectively. Denote the price per byte charged for usage above the cap by p^o. Denote the average throughput of Web browsing by $x_i^b = r_i^b L/M$, where L is the average size (in bits) of a webpage and M is the average time (in seconds) spent on reading a webpage. User i's overage charge is $p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right)$, where T_i denotes user i's tier choice. User i is assumed to choose the times devoted to Web browsing and video streaming so as to maximize surplus, S_i, defined as the difference between willingness-to-pay and payment:

$$\max_{t_i^b, t_i^s} S_i = W_i^b + W_i^s - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right) - P_{T_i}. \tag{2}$$

If user i is not capped, the marginal utility from Web browsing or video streaming is equal to the user's valuation of time, p_i^t, or equivalently,

$$\partial v_i^b V^b\left(r_i^b t_i^b\right)/\partial t_i^b = p_i^t, \quad \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right)/\partial t_i^s = p_i^t. \tag{3}$$

If user i is capped but not paying an overage charge, the marginal utility is larger than the user's valuation of time but smaller than the sum of this and the overage charge (per unit time), or equivalently,

$$p_i^t < \partial v_i^b V^b\left(r_i^b t_i^b\right)/\partial t_i^b < p_i^t + p^o x_i^b, \quad p_i^t < \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right)/\partial t_i^s < p_i^t + p^o x_i^s. \tag{4}$$

Finally, if user i is capped and paying an overage charge, the marginal willingness-to-pay is equal to the overage charge:

$$\partial v_i^b V^b\left(r_i^b t_i^b\right)/\partial t_i^b = p_i^t + p^o x_i^b, \quad \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right)/\partial t_i^s = p_i^t + p^o x_i^s. \tag{5}$$

We now turn to the relationship between traffic and performance. Denote the total downstream Web browsing and video streaming traffic (in bits per month) on the bottleneck link within the access network by $\lambda = \Sigma_i\left(x_i^s t_i^s + x_i^b t_i^b\right)$. As is common, we model the bottleneck link using an M/M/1/K queue to estimate the average delay d and loss l as a function of the traffic λ and the capacity μ.

It remains to express the dependence of application performance upon delay and loss. Denote the rates of the basic tier and the premium tier by X_1 and X_2, respectively. Suppose that user i has subscribed to tier j and thereby obtains a tier rate X_j. For Web browsing, utility depends on performance through a function $V^b\left(t_i^r\right)$ that measures the relative value of time devoted to reading webpages. The ratio of time spent reading webpages to time spent Web browsing, $r_i^b = t_i^r/t_i^b$, can be derived from a TCP latency model [Cardwell et al. 2000]; we denote it as a function TCP^b of the access network delay d, access network loss l, and the user's tier rate X_j:

$$r_i^b = t_i^r/t_i^b = TCP^b\left(d, l, X_{T_i}\right).$$

Since Web browsing performance is constrained by the minimum of the user's tier rate and the throughput obtained using TCP, the function $TCP^b(d, l, X_j)$ is independent of tier rate X_j, when X_j is larger than a threshold X_0 [Cardwell et al. 2000].

For video streaming, utility depends on performance through a sigmoid function $Q^s\left(x_i^s\right)$ of the throughput x_i^s experienced by video streaming applications. Most video streaming uses TCP or TCP-friendly protocols. The throughput of video streaming can be expressed as the minimum of the tier rate and the limit that TCP places on the flow, denoted by $TCP^s(d,l)$ from the TCP throughput model [Padhye et al. 1998].

$$x_i^s = \min\left(X_{T_i}, TCP^s(d,l)\right).$$

2.2. Long-Term Model

In the long term, say on a timescale of months, users seek to maximize their surplus by making the optimal Internet subscription decision. If there is competition between multiple ISPs, then a user would also have to choose between different tiers offered by multiple ISPs. Denote user i's tier choice by $T_i = 0, 1, 2$, where $T_i = 0$ means that user i chooses not to subscribe. User i will choose tier T_i if and only if

$$T_i = \arg\max_j \left[W_i\left(t_i^b, t_i^s \middle| j\right) - P_j - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_j\right)\right], \tag{6}$$

where $W_i\left(t_i^b, t_i^s \middle| j\right) = W_i^b + W_i^s$ is user i's willingness-to-pay for tier j; the amount of time user i devotes to Web browsing $\left(\text{i.e., } t_i^b\right)$ and videos streaming (i.e., t_i^s) can be derived from Equation (2).

The total number of subscribers in tier j is $N_j = \left|i : T_i = j\right|$.

In the United States and many other countries, it is common that only one or two ISPs offer wireline broadband services [Economides 2008]. The proposed user utility functions are general enough to be applied to models with or without competition between ISPs. In the remainder of the article, we consider one ISP that monopolizes the market, which is a reasonable starting point given that there is no academic literature on how an ISP may set data caps and overage charges. A monopoly ISP is presumed to maximize its profit by controlling the parameters in the tiered pricing plan and network capacity:

$$\max_{P_1,P_2,X_1,X_2,C_1,C_2,p^o,\mu} P_1N_1 + P_2N_2 + p^oO - K(\mu) - k\left(N_1 + N_2\right), \tag{7}$$

where $O = \Sigma_i \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right)$ is the total amount of data above the cap; $K(\mu)$ is the ISP's variable cost per month for bandwidth capacity required to accommodate the total user demand; k is the ISP's fixed cost per subscriber per month.

3. MODEL SIMPLIFICATION

To simply model analysis, we make the following four assumptions.

Assumption A. The ISP will set network capacity so that the network load remains at a threshold ρ^{th} $\left(\text{i.e., } \mu = \lambda/\rho^{th}\right)$. The ISP will set tier rates X_1 and X_2 higher than the achievable Web browsing throughput over TCP $\left(\text{i.e., } X_2 > X_1 > X_0\right)$, but no higher than the achievable video streaming throughput over TCP $\left(\text{i.e., } x_i^s = X_{T_i}\right)$.

The assumption regarding network capacity seems to be common practice amongst ISPs. It is further justified by numerical results that show that a monopoly ISP can achieve near-optimal profit using that dimensioning rule [Dai and Jordan 2013c]. The assumptions regarding tier rates hold for almost all current pricing plans by major ISPs in the United States: the performance of Web browsing in most tiers are similar, whereas the performance of video streaming is typically constrained by either the tier rate or the video server rate.

Denote user i's willingness to pay for Web browsing and video streaming in tier j by $W_i^{b,j}$ and $W_i^{s,j}$, respectively. Denote the throughput of Web browsing and video streaming in tier j by $x^{b,j}$ and $x^{s,j}$, respectively.

Assumption B. Users who subscribe to the premium tier prefer the basic tier to no subscription: $W_i^{b,1} + W_i^{s,1} - P_1 - p^o \max\left(0, x^{b,1}t_i^b + x^{s,1}t_i^s - C_1\right) > 0, \quad \forall i : T_i \neq 0.$

This is based on the observation that almost all Internet users who spend considerable time on video streaming also spend considerable time on Web browsing.

Assumption C. The presence of a data cap affects a user's video streaming but not a user's Web browsing, that is, a user's choice of time devoted to Web browsing and video streaming so as to maximize his surplus changes from Equation (2) to

$$\max_{t_i^b, t_i^s} S_i \Rightarrow \begin{cases} \max_{t_i^b} W_i^b, \\ \max_{t_i^s} W_i^s - p^o \max\left(0, x_i^b t_i^b + x_i^s t_i^s - C_{T_i}\right). \end{cases} \tag{8}$$

For uncapped users, the marginal utility from Web browsing or video streaming still satisfies Equation (3). However, for users who are capped but not paying an overage charge, the marginal utility changes from Equation (4) to

$$\partial v_i^b V^b\left(r_i^b t_i^b\right) / \partial t_i^b = p_i^t, \quad p_i^t < \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right) / \partial t_i^s < p_i^t + p^o x_i^s. \tag{9}$$

For users who are capped and paying an overage charge, the marginal utility changes from Equation (5) to

$$\partial v_i^b V^b\left(r_i^b t_i^b\right) / \partial t_i^b = p_i^t, \quad \partial v_i^s V^s\left(t_i^s\right) Q^s\left(x_i^s\right) / \partial t_i^s = p_i^t + p^o x_i^s. \tag{10}$$

Assumption D. No users in the basic tier are capped: $x^{b,1}t_i^b + x^{s,1}t_i^s < C_1, \forall i : T_i = 1.$

Major ISPs in the United States that use data caps do place caps on the lower service tiers. However, the caps placed on the lower service tiers, although lower than the caps placed on the higher service tiers, are high enough so that they do not affect a user's Web browsing, which is consistent with Assumption C. Users in the basic tier do not spend much time on video streaming due to a combination of low interest and poor video streaming performance [Higginbotham 2012; Sandvine 2012]. Thus, the vast majority of basic tier subscribers are not limited by basic tier data caps. Although a few users who do a lot of file sharing can be capped in the basic tier, we conjecture it does not constitute a significant portion of users' willingness-to-pay. We thus do not consider such users. Assumption D simplifies the model by removing one degree of freedom. The model thus predicts that users will upgrade from the basic tier to the premium tier principally to receive increased rates, not increased data caps. However, we acknowledge that it is unknown what contribution differentiated data caps may have in user subscription decisions.

These four simplifying assumptions enable a simplification of the characterization of user tier choice.

THEOREM 3.1. *If Assumptions A–D hold, then user i's tier choice in Equation (6) simplifies to*

$$T_i = \begin{cases} 2 \ \textit{if}\ W_i^{s,2} - W_i^{s,1} > P_{21} + p^o \max\left(0, x_i^b t_i^b + X_2 t_i^s - C_2\right), \\ 0 \ \textit{if}\ W_i^{b,1} + W_i^{s,1} - P_1 < 0, \\ 1 \ \textit{otherwise}, \end{cases} \tag{11}$$

where $P_{21} = P_2 - P_1$*, and* t_i^b *and* t_i^s *can be calculated from Equation (8).*

4. AFFECTED USERS

User i may or may not be capped depending on the user's value placed on Web browsing, v_i^b, on video streaming, v_i^s, and on time, p_i^t. We partition potential Internet subscribers into five groups: those not subscribing to the Internet, basic tier subscribers, premium tier subscribers who are not capped, premium tier subscribers who are capped but not paying an overage charge, and premium tier subscribers who are capped and paying overage charges.

We first focus on users who are indifferent between the basic tier and the premium tier, henceforth referred to as *marginal premium subscribers*. According to Equation (11) in Theorem (3.1), if user i is a marginal premium subscriber, the value placed on Web browsing, on video streaming, and on time satisfy

$$W_i^{s,2} - W_i^{s,1} = P_{21} + p^o \max\left(0, x_i^b t_i^b + X_2 t_i^s - C_2\right). \tag{12}$$

Marginal premium subscribers who are not capped satisfy Equation (3) and Equation (12). Denote $v^{s1,u}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b)$ as the solution to the fixed-point equation in v_i^s resulting from Equation (3) and Equation (12). There is a unique solution for v_i^s, because by Equation (1) and Equation (3), $d(W_i^{s,2} - W_i^{s,1})/dv_i^s > 0, \forall v_i^s > 0$, which makes the left side of Equation (12) an increasing function of v_i^s. Similarly, according to Assumption C, marginal premium subscribers who are capped but not paying an overage charge satisfy Equation (9) and Equation (12); denote $v^{s1,c}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b)$ as the solution to the fixed point from these equations. Finally, marginal premium subscribers who are capped and paying overage charges satisfy Equation (10) and Equation (12); denote $v^{s1,o}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b)$ as the solution to the fixed point from these equations.

Thus the marginal premium subscribers lie on the following curve.

$$v_i^s = v^{s1}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b) = \begin{cases} v^{s1,u}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b) & \text{if } t_i^s < C/X_2 \\ v^{s1,c}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b) & \text{if } t_i^s = C/X_2 \\ v^{s1,o}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b) & \text{if } t_i^s > C/X_2 \end{cases}.$$

We can use this curve to partition all premium tier subscribers. According to Equation (3), premium tier subscribers who are not capped satisfy

$$v_i^s < \frac{p_i^t}{V^{s\prime}\left((C_2 - x_i^b t_i^b)/X_2\right) Q^s(X_2)} \triangleq v^{s2}(C_2, X_2, p_i^t, v_i^b).$$

Denote the set of such subscribers:

$$G_u = \left\{i : v^{s1}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b) < v_i^s < v^{s2}(C_2, X_2, p_i^t, v_i^b)\right\}.$$

According to Equation (9), premium tier subscribers who are capped but not paying an overage charge satisfy

$$v^{s2}(C_2, X_2, p_i^t, v_i^b) < v_i^s < \frac{p_i^t + p^o X_2}{V^{s\prime}\left((C_2 - x_i^b t_i^b)/X_2\right) Q^s(X_2)} \triangleq v^{s3}(C_2, X_2, p^o, p_i^t, v_i^b).$$

Denote the set of such subscribers:

$$G_c = \left\{i : \max\left(v^{s1}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b), v^{s2}(C_2, X_2, p_i^t, v_i^b)\right) < v_i^s < v^{s3}(C_2, X_2, p^o, p_i^t, v_i^b)\right\}.$$

Finally, according to Equation (10), premium tier subscribers who are capped and paying overage charges satisfy

$$v_i^s > v^{s3}(C_2, X_2, p^o, p_i^t, v_i^b).$$

Denote the set of such subscribers:

$$G_o = \left\{i : v_i^s > \max\left(v^{s1}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b), v^{s3}(C_2, X_2, p^o, p_i^t, v_i^b)\right)\right\}.$$

We then focus on users who are indifferent between the basic tier and no Internet subscription, henceforth referred to as *marginal basic subscribers*. According to Equation (11) in Theorem 3.1, if user i is a marginal basic subscriber, the value placed on Web browsing, on video streaming, and on time, p_i^t satisfy

$$W_i^{b,1} + W_i^{s,1} - P_1 = 0. \tag{13}$$

Denote $v^{s0}(P_1, X_1, p_i^t, v_i^b)$ as the solution to the fixed-point equation resulting from Equation (3) and Equation (13). Nonsubscribers, denoted by G_n, place a smaller value on video streaming than do marginal basic subscribers:

$$G_n = \{i : v_i^s < v^{s0}(P_1, X_1, p_i^t, v_i^b)\}.$$

Basic tier subscribers, denoted by G_b, place a smaller value on video streaming than do marginal premium subscribers but larger than do marginal basic subscribers:

$$G_b = \{i : v^{s0}(P_1, X_1, p_i^t, v_i^b) < v_i^s < v^{s1}(C_2, P_{21}, X_1, X_2, p^o, p_i^t, v_i^b)\}.$$

These five sets define a partition of Internet subscribers on the basis of (v_i^s, v_i^b, p_i^t). However, it is more revealing to use $(v_i^s/p_i^t, v_i^b/p_i^t, p_i^t)$ as the basis, as relative values placed on video streaming (i.e., v_i^s/p_i^t) and Web browsing (i.e., v_i^b/p_i^t) determine the amount of time that user i devotes to video streaming and Web browsing absent a data cap. As illustrated in Figure 1, the functions v^{s0}, v^{s1}, v^{s2}, and v^{s3} form the boundaries of the five sets. Users with a very small relative value on Web browsing, v_i^b/p_i^t, and/or a small income (and hence a small p_i^t) do not subscribe to Internet access. Users with a larger relative value on Web browsing but still small relative value on streaming, v_i^s/p_i^t, and/or a small income (and hence a small p_i^t) subscribe to the basic tier. Users with a small relative value on streaming, v_i^s/p_i^t, but a larger income (and hence a larger p_i^t) subscribe to the premium tier but are not capped due to their low interest in streaming. Users with a moderate relative value on streaming, v_i^s/p_i^t, and moderate or high incomes subscribe to the premium tier and are capped. Users with a high relative value on streaming, v_i^s/p_i^t, and/or high incomes subscribe to the premium tier and are willing to pay overage charges.

5. IMPACT OF CAP UPON PRICING PLAN

In this section, we analyze the impact of data caps on the ISP pricing plan. We wish to compare the optimal tier rates, tier prices, and network capacity without caps to the same quantities when caps are added.

5.1. Optimal Pricing Plan without Data Caps

We start by characterizing the rates, prices, and network capacity, without data caps, that maximize an ISP's profit. If Assumptions A and B hold, the ISP profit maximization problem in Equation (7) without data caps (i.e., $p^o = 0$ or $C_1 = C_2 = \infty$) can be reformulated as

$$\max_{P_1, P_{21}, X_1, X_2} Profit_0 = (P_1 - k)(N_1 + N_2) + P_{21}N_2 - K(\lambda/\rho^{th}), \tag{14}$$

where N_1 and N_2 are the number of users subscribing to the basic tier and the premium tier, respectively.

THEOREM 5.1. *If Assumptions A and B hold, the first-order optimality conditions for the profit maximization problem without data caps in Equation (14) satisfy the following.*

$$\partial Profit_0/\partial P_1 = N_1 + N_2 + (P_1 - k)\partial(N_1 + N_2)/\partial P_1 - (p^\mu/\rho^{th})\partial\lambda/\partial P_1 = 0,$$
$$\partial Profit_0/\partial X_1 = (P_1 - k)\partial(N_1 + N_2)/\partial X_1 + P_{21}\partial N_2/\partial X_1 - (p^\mu/\rho^{th})\partial\lambda/\partial X_1 = 0,$$
$$\partial Profit_0/\partial P_{21} = N_2 + P_{21}\partial N_2/\partial P_{21} - (p^\mu/\rho^{th})\partial\lambda/\partial P_{21} = 0,$$
$$\partial Profit_0/\partial X_2 = P_{21}\partial N_2/\partial X_2 - (p^\mu/\rho^{th})\partial\lambda/\partial X_2 = 0,$$

where $p^\mu = dK(\mu)/d\mu$ is the marginal cost for network capacity μ.

5.2. Data Caps that Ensure Heavy Users Pay for Their Usage

We now turn to the effect of adding a data cap into the premium tier, according to Assumption D. We do so in two steps. First, we consider the case in which an ISP institutes caps merely in order to ensure that heavy users pay an amount equal to the cost of their usage. This case is interesting in its own right, as some ISPs claim this is the purpose of their data caps [Yu 2012]. In the next section, we consider the case in which an ISP uses caps to maximize its profit.

Suppose that an ISP imputes a cost to user i equal to $p^\mu(t_i^s X_2 + t_i^b x_i^b)/\rho^{th}$, on the basis that user i's usage is $t_i^s X_2 + t_i^b x_i^b$, and that this requires incremental capacity $(t_i^s X_2 + t_i^b x_i^b)/\rho^{th}$ at an incremental cost per unit capacity p^μ. Then given the optimal prices P_1, $P_2 = P_1 + P_{21}$ and rates X_1, X_2 as calculated in Theorem 5.1, we presume in this section that the goal of the ISP is to set a data cap C_2 and overage charge p^o so that

$$P_1 + P_{21} + p^o(t_i^s X_2 + t_i^b x_i^b - C_2) - p^\mu(t_i^s X_2 + t_i^b x_i^b)/\rho^{th} - k \geq 0, \forall i : T_i = 2.$$

Denote by $t_{max}^{b,1}$ and $t_{max}^{s,1}$ the maximum amount of time users in the basic tier spent on Web browsing and video streaming, respectively. Denote by x^b the throughput of Web browsing of all users, since $X_2 > X_1 > X_0$ by Assumption A. Considering $x^{b,1}t_i^b + x^{s,1}t_i^s < C_1, \forall i : T_i = 1$ from Assumption D, we examine a simple method of achieving this goal: $p^o = p^\mu/\rho^{th}$, $C_1 = (P_1 - k)/p^o$ and $C_2 = P_{21}/p^o + x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}$. We henceforth refer to this choice of cap C_2 and overage charge p^o as the heavy users cap. Under this policy, a premium tier subscriber i with usage greater than or equal to C_2 will pay an amount $P_2 + p^o(t_i^s X_2 + t_i^b x^b - C_2)$, which is larger than or equal to the user's imputed cost (i.e., $p^\mu(t_i^s X_2 + t_i^b x^b)/\rho^{th} + k$).

The impact of such a cap on premium tier subscribers is illustrated in Figure 1. Under the optimal pricing plan without data caps from Equation (14), users above and to the right of the green region subscribe to the premium tier. Under the heavy users cap, users in the brown, blue, and white regions subscribe to the premium tier. The red region corresponds to users who downgrade from the premium tier to the basic tier when data caps are added, because they have high valuations on video streaming but low incomes, denote this set of users by G_d. Subscribers with moderate valuations on video streaming and high incomes (the brown region) are unaffected by the cap. Users with high valuations on video streaming and high incomes (the blue and white regions) have lower surplus after a heavy users cap is added.

5.3. Data Caps to Maximize ISP Profit

We now consider the case in which an ISP sets caps and overage charges to maximize its profit. The pricing plan derived in the previous section does not maximize profit,

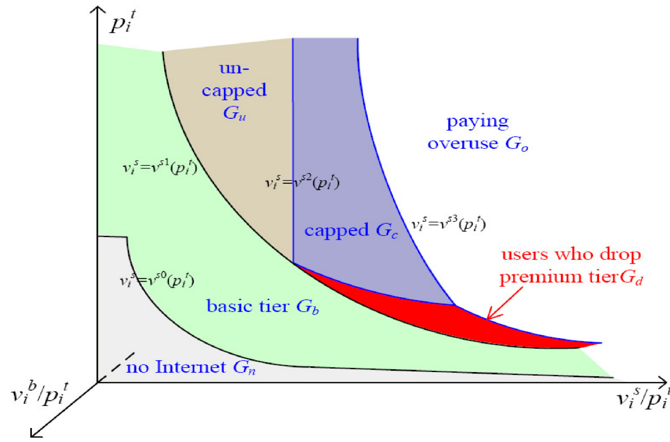


Fig. 1. The impact of a cap designed to ensure that heavy users pay an amount equal to the cost of their usage.

since the cap and overage charge were only intended to ensure that heavy users pay for their usage.

According to Assumption A, $\mu = \lambda/\rho^{th}$. If Assumption D holds, then C_1 can be ignored in Equation (7). Thus, the ISP profit maximization problem with cap in Equation (7) can be reduced to

$$\max_{P_1,P_{21},X_1,X_2,C_2,p^o} \mathit{Profit} = (P_1-k)(N_1+N_2)+P_{21}N_2+p^oO-K(\lambda/\rho^{th}). \tag{15}$$

We henceforth refer to the optimal tier rate, tier price, cap, and overage charge in Equation (15) as the profit-maximizing data cap.

THEOREM 5.2. *If assumptions A–D hold, then the partial derivatives of Profit defined in Equation (15) with respect to P_1, X_1, P_{21}, X_2, C_2, and p^o can be expressed as*

$$\begin{aligned}
&\partial \mathit{Profit}/\partial P_1 = N_1+N_2+(P_1-k)\partial(N_1+N_2)/\partial P_1 - (p^\mu/\rho^{th})\partial\lambda/\partial P_1,\\
&\partial \mathit{Profit}/\partial X_1 = (P_1-k)\partial(N_1+N_2)/\partial X_1 + P_{21}\partial N_2/\partial X_1 + p^o\partial O/\partial X_1 - (p^\mu/\rho^{th})\partial\lambda/\partial X_1,\\
&\partial \mathit{Profit}/\partial P_{21} = N_2 + P_{21}\partial N_2/\partial P_{21} + p^o\partial O/\partial P_{21} - (p^\mu/\rho^{th})\partial\lambda/\partial P_{21},\\
&\partial \mathit{Profit}/\partial X_2 = P_{21}\partial N_2/\partial X_2 + p^o\partial O/\partial X_2 - (p^\mu/\rho^{th})\partial\lambda/\partial X_2,\\
&\partial \mathit{Profit}/\partial C_2 = P_{21}\partial N_2/\partial C_2 + p^o\partial O/\partial C_2 - (p^\mu/\rho^{th})\partial\lambda/\partial C_2,\\
&\partial \mathit{Profit}/\partial p^o = P_{21}\partial N_2/\partial p^o + O + p^o\partial O/\partial p^o - (p^\mu/\rho^{th})\partial\lambda/\partial p^o.
\end{aligned} \tag{16}$$

The optimal pricing plans with and without caps can be numerically calculated from Equation (15) and Equation (14), respectively. Unfortunately, a closed-form characterization of the optimal tier rate, tier price, cap, and overage charge is difficult to obtain from the first-order optimality conditions. We can, however, compare the cap and overage charge from Equation (15) to those in the heavy users cap.

THEOREM 5.3. *If assumptions A–D hold, tier rates and prices are set to maximize profit without caps (i.e., P_1, P_{21}, X_1, and X_2 are set using Theorem 5.1), and data caps* *and overage charges are set using the heavy users cap* $\left(i.e., C_2 = P_{21}/p^o + x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}\right.$ *and* $\left.p^o = p^\mu/\rho^{th}\right)$*, then,* $\partial Profit/\partial C_2 \le 0, \partial Profit/\partial p^o \ge 0$.

Based on Theorem 5.3, we can summarize how ISPs might change the cap parameters starting from the heavy users cap presented in the previous section: the ISP has the incentive to reduce the premium tier cap C_2 and increase the overage charge p^o above p^μ/ρ^{th}. Thus, an ISP that uses caps to maximize profit will have smaller caps and higher overage charges than one that uses caps only to ensure that heavy users pay for their usage.

The ISP also has the incentive to change tier rates and tier prices to further maximize its profit. However, the rates and prices depend on market demand. In a given market, denote by $f(v^b/p^t, v^s/p^t, p^t)$ the joint density of users' relative value placed on Web browsing, on video steaming, and on time. To analyze the changes in tier rates and tier prices when profit-maximizing data caps are adopted, we first focus on the set of users who switch from the premium tier to the basic tier when a heavy users data cap is adopted (i.e., those in the red region in Figure 1). Consider two users in this set, denoted i and i', who place different values on their time but the same relative value on Web browsing and the same relative value on video streaming. As previously noted, these users have high valuations on video streaming but low income (and hence a small p_i^t). It is helpful to understand the variation of the user density function $f(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t)$ with the value placed on time p_i^t. Household income in the United States can be approximated by a lognormal distribution [U.S. Census Bureau 2009]. Users with low income fall into the increasing portion of the lognormal distribution, that is, the user density function $f(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t)$ is an increasing function of value placed on time p_i^t in this set.

Assumption E. $\forall i$ and $i' \in G_d$, if $v_i^b/p_i^t = v_{i'}^b/p_{i'}^t$, $v_i^s/p_i^t = v_{i'}^s/p_{i'}^t$, and $p_i^t > p_{i'}^t$, then $f(v_i^b/p_i^t, v_i^s/p_i^t, p_i^t) > f(v_{i'}^b/p_{i'}^t, v_{i'}^s/p_{i'}^t, p_{i'}^t)$.

We now compare the tier rates and tier prices for an ISP that uses data caps to maximize profit to those for an ISP that does not use data caps. We already know from Theorem 5.3 that a profit-maximizing ISP will set $p^o \ge p^\mu/\rho^{th}$ and $C_2 \le P_{21}/p^o + x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}$ regardless of the user density function. In the case that the data caps are sufficiently low, we can prove a relationship between these tier rates and prices.

THEOREM 5.4. *If Assumptions A–E hold, tier rates and prices are set to maximize profit without caps (i.e.,* P_1*,* P_{21}*,* X_1*, and* X_2 *are set using Theorem 5.1),* $C_2 \le P_{21}/p^o$ *and* $p^o \ge p^\mu/\rho^{th}$*), then,* $\partial Profit/\partial P_1 = 0, \partial Profit/\partial X_1 \le 0, \partial Profit/\partial P_{21} \le 0, \partial Profit/\partial X_2 \ge 0$.

The theorem only applies when $C_2 \le P_{21}/p^o$; however, this case is very likely to occur given that $P_{21}/p^o \gg x^b t_{max}^{b,1} + X_1 t_{max}^{s,1}$. Thus, an ISP that uses caps to maximize profit is likely to have a higher premium tier rate, a lower premium tier price, a lower basic tier rate, and approximately the same basic tier price as an ISP that does not use caps or that uses caps only to ensure that heavy users pay for their usage.

6. IMPACT OF CAP UPON USERS

In this section, we analyze the impact of data caps on users. We wish to determine for which users surplus S_i increases when data caps are adopted. We use Assumptions A–E throughout this section.

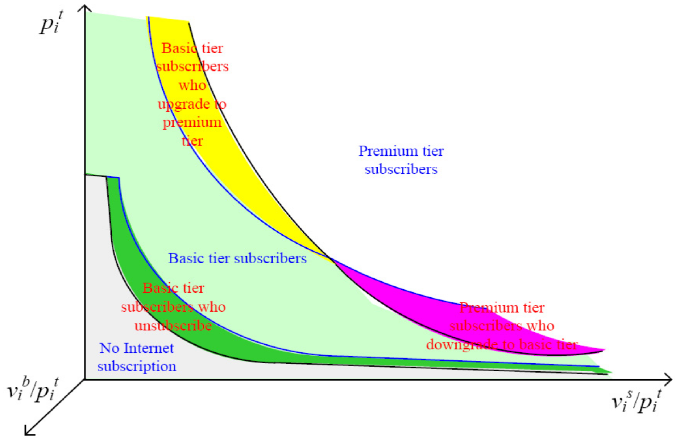


Fig. 2. Comparing service tier choices when a profit-maximizing cap is implemented.

6.1. Impact of Data Caps upon User Tier Choices

In Figure 2, we illustrate the change of service tier choice when a profit-maximizing cap is implemented. We first consider marginal basic subscribers under a pricing plan without caps, that is, users who are indifferent between the basic tier and no Internet subscription, which form the boundary between the green and light grey areas in the figure. The impact of profit-maximizing data caps upon these subscribers is fairly straightforward: by Theorem 5.4, basic tier subscribers will not see a significant change in price but will be hurt by the decreased tier rate X_1. Thus, marginal basic tier users will drop their Internet subscriptions because of the decreased tier rate. In addition, a set of basic tier subscribers with valuations slightly above that of marginal basic subscribers (the dark green region) will also drop. It is also straightforward that households that do not subscribe to Internet access under a pricing plan without caps (the light grey region) will not subscribe to a plan with data caps.

We next consider marginal premium subscribers under a pricing plan without caps, that is, users who are indifferent between the basic tier and the premium tier, which form the boundary between the dark yellow and white regions and the boundary between the dark pink and light green regions. The impact of profit-maximizing data caps upon these subscribers is more complex. By Theorem 5.4, premium tier subscribers will see a reduced premium tier price P_2 and an increased tier rate X_2. However, some of these users will be capped. We analyze them in subsets, as illustrated in Figure 1: those who would not be capped, those who would be capped but not pay an overuse charge, and those who would be capped and pay an overuse charge.

Marginal premium subscribers under a pricing plan without caps, who would not be capped when a data cap is implemented, will benefit from the reduced premium tier price P_2 and increased tier rate X_2. Hence, they will remain premium tier subscribers. In addition, a set of basic tier subscribers with valuations just below that of marginal premium subscribers (a subset of the dark yellow region) will upgrade from the basic tier to the premium tier.

Marginal premium subscribers under a pricing plan without caps who find themselves capped under a profit-maximizing cap pricing plan will have to compare the benefit of the reduced premium tier price P_2 and an increased tier rate X_2 with the drop in utility resulting from the cap. The combination of the increased rate X_2, reduced tier price P_2, and the cap decreases the amount of video streaming these users can

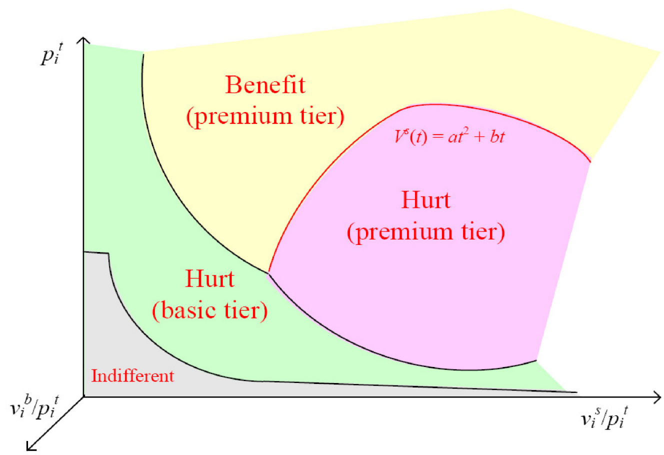


Fig. 3. Comparing user surplus when a profit-maximizing cap is implemented.

do before running into the cap. From Equations (8)–(10), we can show that marginal premium subscribers who place higher value on their time (i.e., larger p^t) but smaller relative value on video streaming (i.e., smaller v^s/p^t) will continue to subscribe to the premium tier in the presence of data caps, since their extra surplus obtained from the increased tier rate X_2 and reduced tier price P_2 outweighs the lost surplus from the reduced amount of time spent on video streaming. In addition, a set of basic tier subscribers with valuations just below that of marginal premium subscribers (a subset of the dark yellow region) will upgrade from the basic tier to the premium tier.

In contrast, marginal premium subscribers under a pricing plan without caps with smaller value on their time (i.e., smaller p^t) will downgrade to the basic tier in the presence of data caps, since their lost surplus from the reduced amount of time devoted to video streaming outweighs the benefit from the increased tier rate and decreased tier price. Similarly, a set of premium tier subscribers with valuations slightly above that of marginal premium subscribers (the dark pink region) will also downgrade from the premium tier to the basic tier.

6.2. Impact of Data Caps upon User Surplus

We now turn to the change of user surplus when data caps are adopted. In Figure 3, we illustrate the change in user surplus when a profit-maximizing data cap is implemented.

According to the preceding analysis of the impact of data caps upon user tier choices, users who place low values on video streaming $\left(\text{i.e., } v^s/p^t\right)$ or time $\left(\text{i.e., } p^t\right)$ (shaded in grey) do not subscribe to Internet access under either pricing plan. Thus, they are indifferent to data caps, because their surplus is zero under both plans. Subscribers who remain in the basic tier when data caps are adopted (shaded in light green) are hurt by the data caps because of the decreased tier rate X_1 and the same tier price P_1, according to Theorem 5.4. Premium subscribers with moderate valuations on video streaming and moderate-to-high incomes (shaded in light yellow) benefit from a data cap; these users place a high value on their time but do not spend much time on video streaming, and consequently also benefit from the reduction in tier price and increase in tier rate. In contrast, premium subscribers with moderate-to-high valuations on video streaming and low-to-moderate incomes (shaded in light pink) are hurt by a data cap; the effect of the cap and overage charges outweigh the reduction in tier price and the increase in tier rate.

We would like to understand the shape of the boundary between the light yellow and light pink regions, and in particular whether this boundary is monotonically increasing or not. Select a marginal premium subscriber under a pricing plan without caps would not be capped when a data cap is implemented, that is, a subscriber on the boundary between the white and dark yellow regions in Figure 2. Denote by $h_1(p^t, v^b) = v^{s1}(\infty, P_{21}, X_1, X_2, 0, p^t, v^b)$ the value placed on streaming by the selected marginal premium subscriber. The set of users who place the same value on time (i.e., p^t) and the same relative value on Web browsing (i.e., v^b/p^t) but higher values on video streaming (v^s) as this selected user constitute a horizontal half-line in the $(v^s/p^t, p^t)$ plane of Figure 2 (i.e., $v^s \geq h_1(p^t, v^b)$).

When profit-maximizing data caps are adopted, denote the changes in X_1, X_2, and P_{21} by ΔX_1, ΔX_2, and ΔP, respectively. By Theorem 5.4, $\Delta X_1 < 0$, $\Delta X_2 > 0$, and $\Delta P < 0$. Similarly, denote the change in user surplus by $\Delta S(p^t, v^b, v^s)$. Denote value of v^s at which the minimum of $\Delta S(p^t, v^b, v^s)$ occurs by $v^{s4} = \arg\min_{v^s \in \{v^s \geq h_1(p^t, v^b)\}} \Delta S(p^t, v^b, v^s)$. Consequently, as previously discussed, the users in the dark yellow region in Figure 2 upgrade from the basic tier to the premium tier when profit-maximizing data caps are implemented. Denote by $h_2(p^t, v^b) = v^{s1}(C_2, P_{21} + \Delta P, X_1 + \Delta X_1, X_2 + \Delta X_2, p^o, p^t, v^b) < h_1(p^t, v^b)$ the value placed on streaming by a marginal premium subscriber under a pricing plan with profit-maximizing caps, that is, a subscriber on the boundary between the light green and dark yellow regions in Figure 2.

In the special case in which there is an absolute cap placed on usage (i.e., $p^o = \infty$), we can show that the boundary is monotonically increasing to infinity.

THEOREM 6.1. *If $p^o = \infty$, for a fixed (p^t, v^b), then there exists a unique root in v^s of $\Delta S(p^t, v^b, v^s)$, denoted v^{th}. Furthermore, $\Delta S(p^t, v^b, v^s) > 0, \forall v^s : h_2(p^t, v^b) < v^s < v^{th}$ and $\Delta S(p^t, v^b, v^s) \leq 0, \forall v^s : v^s \geq v^{th}$.*

Theorem 6.1 is illustrated in Figure 4. The black curve shows user surplus under a pricing plan without data caps, and the blue curve shows the user surplus under profit-maximizing data caps. The user surplus under an absolute cap is a linear increasing function of v^s for $v^s \geq v^{s2}(C_2, X_2 + \Delta X_2, p^t)$. The theorem guarantees that there is a unique threshold: premium tier subscribers with valuations on video streaming below v^{th} benefit from data caps because the reduction in price and increase in tier rate outweigh the impact of the cap (if any), whereas those with valuations above v^{th} are hurt by data caps because the impact of being capped outweighs the reduction in price and increase in tier rate.

In the general case in which there exist overage charges for use above the cap, the shape of the boundary between the light yellow and light pink regions depends on the shape of $V^s(t)$. We examine two special cases here: a quadratic function $V^s(t) = at^2 + bt$, $(0 < t < -b/2a = t^{max}, a < 0, b > 0)$ which produces a linear demand curve [Besanko and Braeutigam 2005], and a concave polynomial function $V^s(t) = ct^k$, $(0 < t < t^{max}, 0 < k < 1)$ which produces a constant elasticity demand curve [Besanko and Braeutigam 2005].

In the case of a linear demand curve, we can show that the boundary is not monotonically increasing, and in fact that it is a unimodal function.

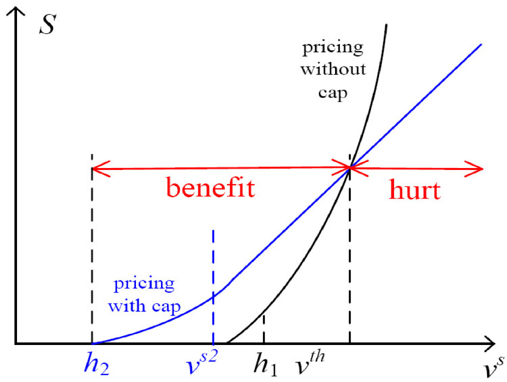


Fig. 4. User surplus under the pricing plan without caps and the pricing plan with absolute caps (i.e., $p^o = \infty$).

THEOREM 6.2. *If* $V^s(t) = at^2 + bt$, $\left(0 < t < -b/2a = t^{max}, a < 0, b > 0\right)$.

(1) *For a fixed* (p^t, v^b), *if* $\Delta S(p^t, v^b, v^{s4}) \geq 0$, *then* $\Delta S(p^t, v^b, v^s) \geq 0, \forall v^s : v^s \geq h_2(p^t, v^b)$.
(2) *For a fixed* (p^t, v^b), *if* $\Delta S(p^t, v^b, v^{s4}) < 0$, *then there exists exactly two roots in* v^s *of* $\Delta S(p^t, v^b, v^s)$, *denoted* $v^{th,1}$ *and* $v^{th,2}$, $\left(v^{th,1} \leq v^{th,2}\right)$. *Furthermore,* $\Delta S(p^t, v^b, v^s) \geq 0$, $\forall v^s : h_2(p^t, v^b) \leq v^s \leq v^{th,1}$, $\Delta S(p^t, v^b, v^s) \leq 0, \forall v^s : v^{th,1} \leq v^s \leq v^{th,2}$, *and* $\Delta S(p^t, v^b, v^s) \geq 0, \forall v^s : v^s \geq v^{th,2}$.

Theorem 6.2 is illustrated in Figure 5. The first case occurs when the cross section lies above the maximum point on the boundary, that is, entirely within the light yellow region of Figure 3. In Figure 5(a), the user surplus under profit-maximizing data caps is now increasing convex. In this case, the surplus increases for all premium tier subscribers (as well as users who upgrade from the basic to premium tier). The second case occurs when the cross section lies below the maximum point on the boundary, that is, it crosses from the light yellow region to the light pink region and back into the light yellow region. In Figure 5(b), at the point v^{s4}, $\Delta S(p^t, v^b, v^{s4}) < 0$, that is, there exists premium tier subscribers that are hurt by data caps. The theorem guarantees that there are two thresholds: premium tier subscribers with valuations on video streaming below $v^{th,1}$ benefit from data caps because the reduction in tier price and increase in tier rate outweigh the impact of the cap (if any), those with valuations between $v^{th,1}$ and $v^{th,2}$ are hurt by data caps because the impact of being capped outweighs the reduction in tier price and increase in tier rate, and those with valuations above $v^{th,2}$ benefit because the increase in tier rate outweighs the overage charge. Furthermore, $v^{th,1}$ is increasing in p^t, and $v^{th,2}$ is decreasing in p^t. Thus the boundary between the light yellow and light pink is a unimodal function of v^s.

In contrast, in the case of a constant elasticity demand curve, we can show that the boundary is monotonically increasing.

THEOREM 6.3. *If* $V^s(t) = ct^k$, $\left(0 < t < t^{max}, 0 < k < 1\right)$.

(1) *For a fixed* (p^t, v^b), *if* $\left(\frac{p^t + p^o\left(X_2 + \Delta X_2\right)}{p^t}\right)^2 \leq \frac{Q^s\left(X_2 + \Delta X_2\right)}{Q^s\left(X_2\right)}$, *then* $\Delta S(p^t, v^b, v^s) \geq 0, \forall v^s : v^s \geq h_2(p^t, v^b)$.

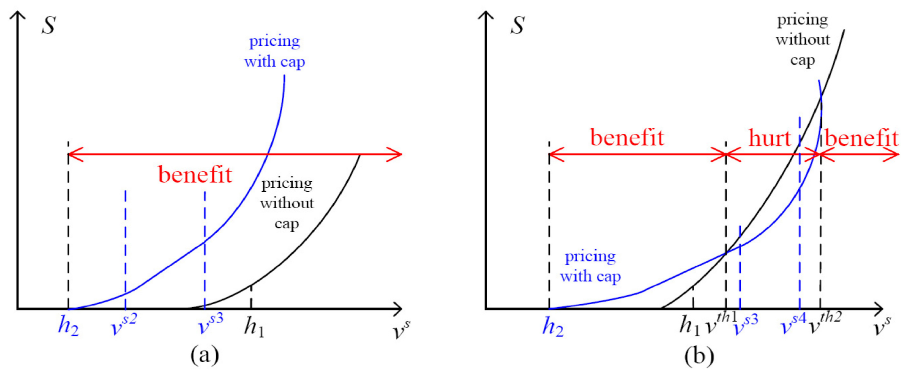


Fig. 5. User surplus under the pricing plan without data caps and the pricing plan with data caps.

(2) *For a fixed* (p^t, v^b)*, if* $\left(\frac{p^t+p^o(X_2+\Delta X_2)}{p^t}\right)^2 > \frac{Q^s(X_2+\Delta X_2)}{Q^s(X_2)}$*, then there exists a unique root in* v^s *of* $\Delta S(p^t, v^b, v^s)$*, denoted* v^{th}*. Furthermore,* $\Delta S(p^t, v^b, v^s) \geq 0, \forall v^s : h_2(p^t, v^b) \leq v^s \leq v^{th}$*, and* $\Delta S(p^t, v^b, v^s) \leq 0, \forall v^s : v^s \geq v^{th}$*.*

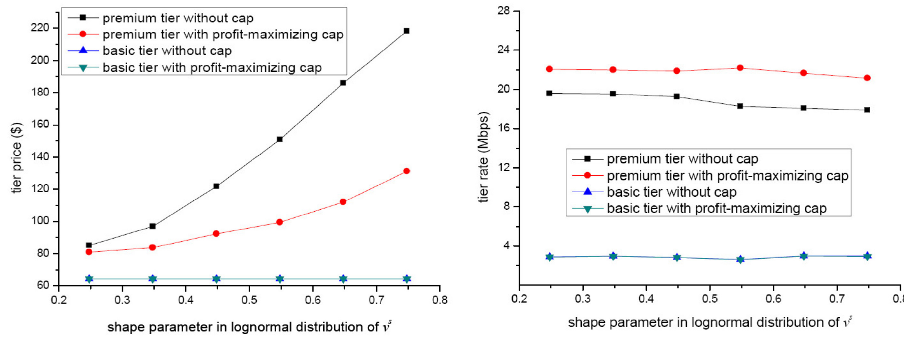
The first case occurs when the cross section lies above the boundary, that is, entirely within the light yellow region of Figure 3. The second case occurs when the cross section lies crosses the boundary from the light yellow region to the light pink region.

When there is linear demand, there an upper limit on the number of hours of video streaming by any subscriber. In contrast, when there is constant elasticity demand, there is no such limit, and hence users with high valuations spend more time on video streaming when data caps are absent, which outweighs the benefit from the increased tier rate when data caps are present.

7. NUMERICAL RESULTS

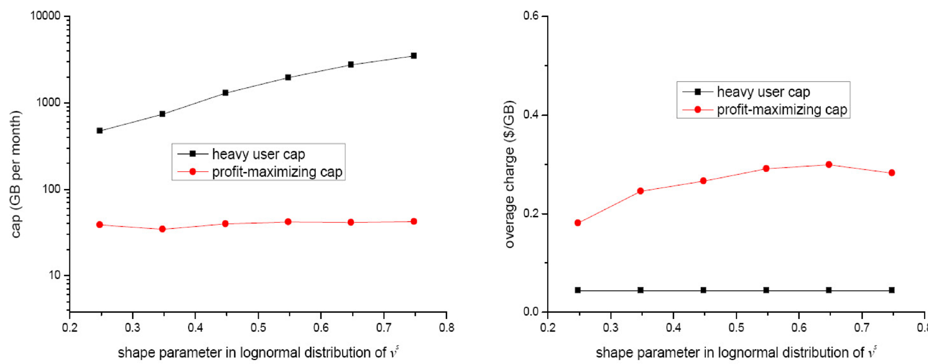
In this section, we evaluate the impact of data caps on the pricing plan, users, the ISP, and social welfare. We simulate 20,000 light users and 20,000 moderate-to-heavy users. For the light users, $(v^b/p^t, p^t)$ follows a multivariate lognormal distribution with parameters set to match demand and income statistics in Cesario [1976], GTRC [1998], and the U.S. Census Bureau [2009], and with a constant elasticity demand curve given by $V^b(t) = ct^k$ with parameters set to match the Web browsing statistics in GTRC [1998]. For the heavy users, $(v^s/p^t, p^t)$ follows a multivariate lognormal distribution with parameters set to match demand and income statistics in Cesario [1976], Burstmedia.com [2011], and U.S. Census Bureau [2009], and with a constant elasticity demand curve given by $V^s(t) = ct^k$ with parameters set to match the video streaming statistics in Burstmedia.com [2011]. Video streaming performance $Q^s(x)$ as a function of throughput is taken from Weber and Veeraraghavan [2007]. The load threshold $\rho^{th} = 0.7$ [Dai and Jordan 2013c]. The marginal cost per unit capacity p^μ =\$10/Mbps/month [CCS Leeds 2012].

In Figure 6(a), we plot the ISP's basic tier price P_1 and premium tier price P_2 with and without profit-maximizing data caps. The premium tier price P_2 depends strongly on the distribution of the scale factor v^s for users' utility for video streaming, and hence it is plotted as a function of the shape parameter (denoted σ) of the lognormal distribution. Both the mean and variance of v^s increase with σ, while the median is fixed, reflecting a higher proportion of heavy users. Without a data cap, the price increases



(a) (b)

Fig. 6. Tier price (a) and tier rate (b) versus proportion of heavy users.



(a) (b)

Fig. 7. Data caps (a) and overage charges (b) versus proportion of heavy users.

rapidly with the proportion of heavy users due to both the higher willingness-to-pay of heavy users and the much higher usage of heavy users. When profit-maximizing data caps are used, the premium tier price P_2 decreases substantially, as predicted by the preceding analytical results. The prices both with and without caps would fall if there were competing ISPs. The basic tier price P_1 remains approximately unchanged after profit-maximizing data caps are added, as predicted by the analytical results.

In Figure 6(b), we plot the ISP's basic tier rate X_1 and premium tier rate X_2 with and without profit-maximizing data caps. The premium tier rate X_2 decreases slightly with the proportion of heavy users in an attempt to maintain acceptable performance and cost. When profit-maximizing data caps are used, the premium tier rate X_2 increases moderately as predicted by the analytical results. However, the basic tier rate X_1 only decreases slightly after profit-maximizing data caps are added, because changing X_1 only slightly changes the quality of Web browsing and video streaming.

The decrease in the premium tier price when an ISP uses profit-maximizing data caps does not necessarily decrease the total price paid by the subscriber, since there are also overage charges. In Figure 7, we illustrate the caps and overage charges under the profit-maximizing cap and the less aggressive heavy users cap previously discussed. Under the heavy users cap, the caps are quite high and the overage charges are quite low, since the cap and overage charges are only intended to recover the imputed cost from heavy users. In contrast, under the profit-maximizing cap, the caps are quite low and the overage charges are quite high. Current wireline ISP caps are roughly in this

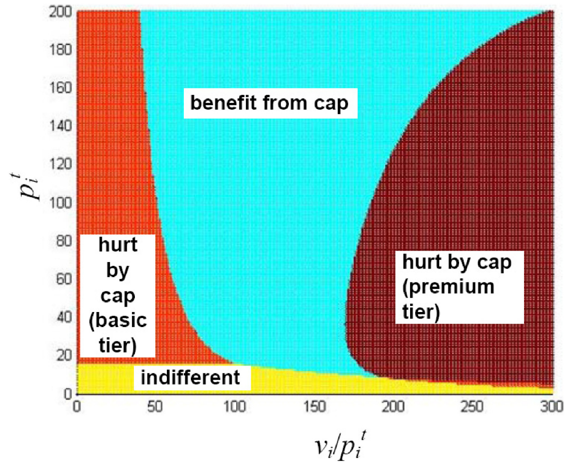
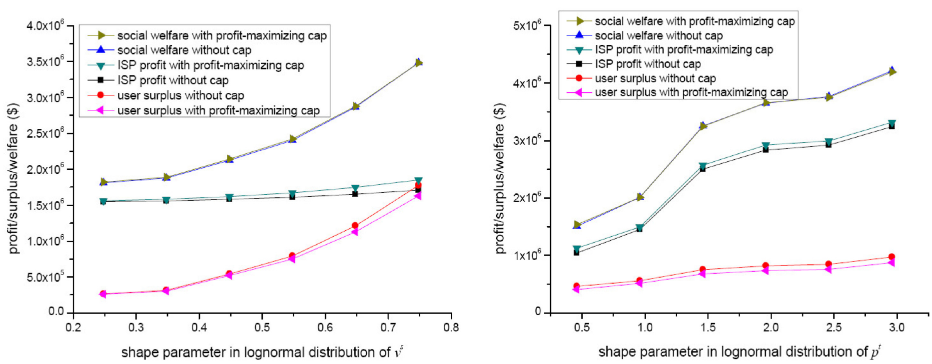


Fig. 8. User surplus under profit-maximizing caps.



(a) (b)

Fig. 9. ISP profit, user surplus, and social welfare versus (a) the proportion of heavy users and (b) the proportion of wealthy users.

range (e.g., AT&T offers pricing plans with a cap of 150 or 250 GB per month and an overage charge of $10 for an additional 50GB [Higginbotham 2012]).

To understand the impact upon different types of users, in Figure 8, we investigate which users are better off with profit-maximizing data caps, as a function of $\left(v_i^s/p_i^t, p_i^t\right)$, when $\sigma = 0.548$. Users in the yellow region are indifferent, since they do not subscribe to the Internet in either case. Users in the blue region have a larger surplus when profit-maximizing data caps are used, since the benefit of the decreased premium tier prices and increased premium tier rates outweighs the impact of the caps. Premium tier subscribers in the dark red region have a smaller surplus when profit-maximizing data caps are used, since the impact of the caps outweighs the benefit of decreased premium tier price and increased premium tier rate. Basic tier subscribers in the light red region also have a smaller surplus under the profit-maximizing data caps, since the basic tier rate is reduced slightly, while the basic tier price remains the same.

In Figure 9(a), we give the ISP profit, user surplus $\left(\text{defined as } \Sigma_i S_i\right)$, and social welfare (defined as user surplus plus ISP profit) resulting from each plan. Without data caps, ISP profit increases as the proportion of heavy users increases due to increases in subscriptions to the premium tier. When profit-maximizing data caps are used, ISP profit further increases. The increase reflects the new overage charges minus some

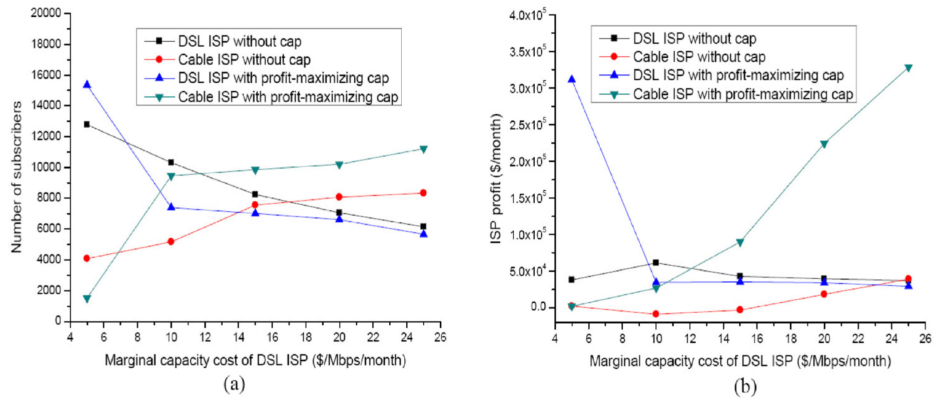


Fig. 10. (a) Number of subscribers and (b) ISP profits versus marginal cost per unit capacity of the DSL ISP.

reductions due to lower tier prices and some changes in premium tier subscriptions. The revenue and profit are both more sensitive to the proportion of heavy users when data caps are used because of the strong correlation between the proportion of heavy users and revenue raised through overage charges.

User surplus increases with the proportion of heavy users, since heavy users have high surpluses. When profit-maximizing data caps are present, user surplus decreases when there is a high proportion of heavy users. Social welfare might decease or increase when profit-maximizing data caps are used, depending on parameters. In this plot, it increases slightly.

Social welfare also depends on the distribution of wealth in the society. In Figure 9(b), we plot it versus the shape parameter of the lognormal distribution for a user's value on time, which is also proportional to income. Social welfare increases with wealth inequality, primarily since the mean income is increasing. Social welfare is observed here to increase slightly under profit-maximizing data caps when inequality is low, but to decrease slightly when inequality is high. We warn, however, that social welfare also depends on the shape of the utility function, and that different utility functions may result in different conclusions about changes in social welfare.

Finally, we briefly consider the application of our user utility model to analyze an ISP duopoly. We simulate one DSL ISP and one cable ISP, who compete to serve 20,000 moderate-to-heavy users by setting premium tier rates, premium tier prices, data caps, and overage charges for profit maximization. The cable ISP is assumed to have a higher fixed cost ($20 per user per month) than the DSL ISP ($15 per user per month). We set the marginal network capacity cost of the cable ISP at $5/Mbps/month, and vary the marginal network capacity cost of the DSL ISP from $5/Mbps/month to $25/Mbps/month. In Figure 10, we illustrate the resulting number of subscribers to and profits of each ISP both with and without data caps.

Without data caps, when the ISPs have the same marginal capacity cost, the DSL ISP dominates the Internet access market because it has a lower fixed cost per user. However, as the DSL ISP's marginal capacity cost increases, the DSL ISP loses customers and the cable ISP gains customers, as expected. The DSL ISP serves relatively light users at a lower premium tier rate and price, and the cable ISP serves relatively heavy users at a higher premium tier rate and price.

When profit-maximizing data caps are added into the pricing plans, the cable ISP serving heavier users increases the premium tier rate, reduces the premium tier price, reduces data caps below a heavy users cap, and increases overage charges above a heavy users cap, as predicted in the monopoly case in Theorems 5.3 and 5.4. In contrast, the DSL ISP does not have much incentive to set data caps, because it is serving light users who do not consume much data. The use of data caps thus results in an advantage for the cable ISP over the DSL ISP. As a result, a significant number of users switch from the DSL ISP to the cable ISP. While the DSL ISP's profit decreases slightly under data caps, the cable ISP's profit increases dramatically.

8. CONCLUSION AND FUTURE WORK

This article proposes novel user utility models that incorporate the time users devote to Internet applications and the opportunity cost of a user's leisure time, thus differentiating light and heavy users on an economic basis. It considers a monopoly ISP that maximizes profit by setting tier prices, tier rates, network capacity, data caps, and overage charges, thereby presenting what we believe is the first model in the academic literature of how an ISP may set data caps and overage charges. We show that users with small relative value on streaming and large income subscribe to the premium tier but are not capped due to their low interest in streaming; users with moderate relative value on streaming and moderate or high incomes subscribe to the premium tier and are capped; and that users with high relative value on streaming and/or high incomes subscribe to the premium tier and are willing to pay overage charges.

Analytical and numerical results show that the ISP will increase the premium tier rate, decrease the basic tier rate and premium tier price, and keep the basic tier price approximately unchanged when data caps are used to maximize profit. The ISP will set smaller caps and higher overage charges than when caps are used only to ensure that heavy users pay for their usage. As a result, light premium users benefit from data caps because of the increased tier rate and reduced tier price, while heavy premium users are hurt by the caps and overage charges. User surplus and social welfare may increase or decrease depending on the shape of the market density function and user utility function.

These results may be used by policymakers to address the public policy debate over usage-based pricing and data caps. Although the monopoly case is interesting in its own right, an excellent topic for future research would be considering multiple ISPs and content providers competing in a market.

APPENDIXES

A.1. Proof of Theorem 3.1

Using assumptions A, C, and D, user i's willingness to pay for Web browsing is the same in both tiers $\left(\text{i.e. } W_i^{b,1} = W_i^{b,2}\right)$, since the performance of Web browsing is identical in the two tiers and the presence of a data cap does not affect the user's Web browsing. Using Assumption B and $W_i^{b,1} = W_i^{b,2}$, it follows that the user tier choice in Equation (6) simplifies to Equation (11).

A.2. Proof of Theorem 5.1

The partial derivatives follow directly from earlier equations.

A.3. Proof of Theorem 5.2

According to Equation (11) in Theorem 3.1, N_2 only depends on $P_2 - P_1$ $\left(\text{or } P_{21}\right), X_1, X_2$, C_2, and p^o; $N_1 + N_2$ only depends on P_1, X_1. Thus, $\partial N_2 / \partial P_1 = 0$, $\partial\left(N_1 + N_2\right) / \partial X_2 = 0$, $\partial\left(N_1 + N_2\right) / \partial P_{21} = 0$, and $\partial\left(N_1 + N_2\right) / \partial C_2 = 0$, $\partial\left(N_1 + N_2\right) / \partial p^o = 0$. Thus, Equation (16) can be obtained by replacing the partial derivatives of the ISP profit with cap defined in Equation (15).

A.4. Proof of Theorem 5.3

To simplify notation, we henceforth denote the vector $(\upsilon^b/p^t, \upsilon^s/p^t, p^t)$ by υ.

We first consider the overage charge p^o; Equation (16) gives

$$\partial Profit/\partial p^o = P_{21}\partial N_2/\partial p^o + O + p^o\partial O/\partial p^o - (p^\mu/\rho^{th})\partial\lambda^s/\partial p^o.$$

Denote the traffic of video streaming from users in the basic tier and premium tier by $\lambda^{s,1}$ and $\lambda^{s,2}$, respectively, so that $\lambda^s = \lambda^{s,1} + \lambda^{s,2}$. Denote the time a user in the basic tier and premium tier devotes to videos streaming by $t^{s,1}$ and $t^{s,2}$, respectively. If the overage charge p^o is changed by Δp^o, some marginal premium subscribers who are paying overage charges will switch between the basic tier and the premium tier. Denote this set of users by ΔG_o; thus,

$$\begin{aligned}
&\partial N_2/\partial p^o = \textstyle\int_{\Delta G_o} f(\upsilon)d\upsilon/\Delta p^o\\
&\partial O/\partial p^o = \textstyle\int_{\Delta G_o}(X_2t^{s,2} + x^bt^b - C_2)f(\upsilon)d\upsilon/\Delta p^o + \int_{G_o} X_2 f(\upsilon)\partial t^{s,2}/\partial p^o d\upsilon\\
&\partial\lambda^{s,1}/\partial p^o = -\textstyle\int_{\Delta G_o} X_1t^{s,1}f(\upsilon)d\upsilon/\Delta p^o\\
&\partial\lambda^{s,2}/\partial p^o = \textstyle\int_{\Delta G_o} X_2t^{s,2}f(\upsilon)d\upsilon/\Delta p^o + \int_{G_o} X_2 f(\upsilon)\partial t^{s,2}/\partial p^o d\upsilon.
\end{aligned}$$

Considering $p^o = p^\mu/\rho^{th}$, we have

$$\frac{\partial Profit}{\partial p^o} = \frac{\int_{\Delta G_o}(P_{21} - p^o(C_2 - x^bt^b) + p^oX_1t^{s,1})f(\upsilon)d\upsilon}{\Delta p^o} + \int_{G_o}(X_2t^{s,2} + x^bt^b - C_2)f(\upsilon)d\upsilon.$$

The data cap is set to be $C_2 = P_{21}/p^o + x^bt^{b,1}_{max} + X_1t^{s,1}_{max}$. Thus, we have

$$P_{21} - p^o(C_2 - x^bt^b) + p^oX_1t^{s,1} = p^o\left(x^b(t^b - t^{b,1}_{max}) + X_1(t^{s,1} - t^{s,1}_{max})\right) \le 0.$$

As the overage charge p^o decreases, more users subscribe to the premium tier. Thus, the first term in $\partial Profit/\partial p^o$ is a nonnegative value. Marginal premium subscribers in G_o consume more data than the cap C_2 if they subscribe to the premium tier, and hence $X_2t^{s,2} + x^bt^b - C^2 \ge 0$. Thus, the second term is also a nonnegative value. So, $\partial Profit/\partial p^o \ge 0$.

We now turn to cap C_2; Equation (16) also gives

$$\partial Profit/\partial C_2 = P_{21}\partial N_2/\partial C_2 + p^o\partial O/\partial C_2 - (p^\mu/\rho^{th})\partial\lambda^s/\partial C_2.$$

Similarly, if cap C_2 is changed by ΔC_2, some marginal premium subscribers who are capped but not paying overage charges and some marginal premium subscribers who are paying overage charges will switch between the basic tier and the premium tier. Denote these sets of users by ΔG_c and ΔG_o. Thus,

$$\begin{aligned}
&\partial N_2/\partial C_2 = \textstyle\int_{\Delta G_c\cup\Delta G_o} f(\upsilon)d\upsilon/\Delta C_2\\
&\partial O/\partial C_2 = \textstyle\int_{\Delta G_o}(X_2t^{s,2} + x^bt^b - C_2)f(\upsilon)d\upsilon/\Delta C_2 - \int_{G_o} f(\upsilon)d\upsilon\\
&\partial\lambda^{s,1}/\partial p^o = -\textstyle\int_{\Delta G_c\cup\Delta G_o} X_1t^{s,1}f(\upsilon)d\upsilon/\Delta C_2\\
&\partial\lambda^{s,2}/\partial p^o = \textstyle\int_{\Delta G_c\cup\Delta G_o} X_2t^{s,2}f(\upsilon)d\upsilon/\Delta C_2.
\end{aligned}$$

Considering $p^o = p^\mu/\rho^{th}$ and $C_2 = P_{21}/p^o + x^bt^{b,1}_{max} + X_1t^{s,1}_{max}$, we have

$$\frac{\partial Profit}{\partial C_2} = \frac{\int_{\Delta G_c\cup\Delta G_o} p^o(x^bt^b - x^bt^{b,1}_{max} + X_1t^{s,1} - X_1t^{s,1}_{max})f(\upsilon)d\upsilon}{\Delta C_2} - \int_{G_o} p^of(\upsilon)d\upsilon.$$

As cap C_2 increases, more capped users subscribe to the premium tier. Thus, the first term is a nonpositive value. The second term is also a nonpositive value. So, $\partial Profit/\partial C_2 \le 0$.

A.5. Proof of Theorem 5.4

We first consider the basic tier price P_1. According to Theorem 5.1, the optimal tiered pricing plan without data caps satisfies

$$\partial Profit_0/\partial P_1 = N_1 + N_2 + (P_1 - k)\partial(N_1 + N_2)/\partial P_1 - (p^\mu/\rho^{th})\partial\lambda/\partial P_1 = 0.$$

From Equation (16) in Theorem 5.2, we have

$$\partial Profit/\partial P_1 = N_1 + N_2 + (P_1 - k)\partial(N_1 + N_2)/\partial P_1 - (p^\mu/\rho^{th})\partial\lambda/\partial P_1 = 0.$$

According to Theorem 3.1, $N_1 + N_2$ is a function of X_1 and P_1. $\partial\lambda/\partial P_1$ is also a function of X_1 and P_1, because changing P_1 only makes marginal basic subscribers switch between no Internet subscription and the basic tier, whose choices only depend on X_1 and P_1, according to Equation (11).Thus, we have

$$\partial Profit/\partial P_1 = \partial Profit_0/\partial P_1 = 0.$$

We then turn to the tier differential price P_{21}. In the absence of data caps, if the price P_{21} is changed by ΔP_{21}, some marginal premium subscribers in sets G_b and $G_u \cup G_d$ will switch between the basic tier and the premium tier; see the black curve that partitions G_b and $G_u \cup G_d$ in Figure 2. Denote these sets of users by ΔG_u and ΔG_d. Thus, Theorem 5.1 gives

$$\begin{aligned}\frac{\partial Profit_0}{\partial P_{21}} &= N_2 + P_{21}\frac{\partial N_2}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial P_{21}}\\ &= \int_{G_u\cup G_c\cup G_o\cup G_d} f(\upsilon)d\upsilon + P_{21}\frac{\int_{\Delta G_u\cup\Delta G_d} f(\upsilon)d\upsilon}{\Delta P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u\cup\Delta G_d}\left(X_2t^{s,2} - X_1t^{s,1}\right)f(\upsilon)d\upsilon}{\Delta P_{21}} = 0.\end{aligned}$$

In the presence of data caps, if price P_{21} is changed by ΔP_{21}, some marginal premium subscribers in sets $G_b \cup G_d$ and $G_u \cup G_c \cup G_o$ may switch between the basic tier and the premium tier; see the partial black curve that partitions G_b and G_u, and the blue curve that partitions G_d and $G_c \cup G_o$ in Figure 1. Denote these sets of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 5.2 gives

$$\begin{aligned}\frac{\partial Profit}{\partial P_{21}} = N_2 + P_{21}\frac{\partial N_2}{\partial P_{21}} + p^o\frac{\partial O}{\partial P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial P_{21}} &= \int_{G_u\cup G_c\cup G_o} f(\upsilon)d\upsilon + P_{21}\frac{\int_{\Delta G_u\cup\Delta G_c\cup\Delta G_o} f(\upsilon)d\upsilon}{\Delta P_{21}}\\ &+ p^o\frac{\int_{\Delta G_o}\left(X_2t^{s,2} + x^bt^b - C_2\right)f(\upsilon)d\upsilon}{\Delta P_{21}} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u\cup\Delta G_c\cup\Delta G_o}\left(X_2t^{s,2} - X_1t^{s,1}\right)f(\upsilon)d\upsilon}{\Delta P_{21}}.\end{aligned}$$

Thus, we have

$$\begin{aligned}\frac{\partial\left(Profit - Profit_0\right)}{\partial P_{21}} &= -\int_{G_d} f(\upsilon)d\upsilon - \frac{\int_{\Delta G_d}\left(P_{21} - p^\mu\left(X_2t^{s,2} - X_1t^{s,1}\right)/\rho^{th}\right)f(\upsilon)d\upsilon}{\Delta P_{21}}\\ &+ \frac{\int_{\Delta G_c\cup\Delta G_o}\left(P_{21} - p^\mu\left(C_2 - x^bt^b - X_1t^{s,1}\right)/\rho^{th}\right)f(\upsilon)d\upsilon}{\Delta P_{21}} + \left(p^o - \frac{p^\mu}{\rho^{th}}\right)\frac{\int_{\Delta G_o}\left(X_2t^{s,2} + x^bt^b - C_2\right)f(\upsilon)d\upsilon}{\Delta P_{21}}.\end{aligned} \tag{17}$$

Obviously, the first term in Equation (17) is nonpositive. As P_{21} decreases, more users subscribe to the premium tier. Marginal premium subscribers in G_o consume more data than the cap C_2 if they subscribe to the premium tier, and thus $X_2t^{s,2} + x^bt^b - C_2 \geq 0$. Thus, the last term in Equation (17) is nonpositive, since $p^o \geq p^\mu/\rho^{th}$. The sum of the second and third term in Equation (17) can be expressed as

$$\begin{aligned}-\int_{\frac{v^b}{p^t}}\int_{\frac{v^s}{p^t}}\Bigg(\left(P_{21} - \frac{p^\mu\left(C_2 - x^bt^b - X_1t^{s,1}\right)}{\rho^{th}}\right)&\frac{f\left(v^b/p^t, v^s/p^t, p^{t,c}\right)}{\partial S^c/\partial p^t|_{p^t=p^{t,c}}}\\ -\left(P_{21} - \frac{p^\mu\left(X_2t^{s,2} - X_1t^{s,1}\right)}{\rho^{th}}\right)&\frac{f\left(v^b/p^t, v^s/p^t, p^{t,u}\right)}{\partial S^u/\partial p^t|_{p^t=p^{t,u}}}\Bigg)d\frac{v^s}{p^t}d\frac{v^b}{p^t},\end{aligned}$$

where $p^{t,u}$ and $p^{t,c}$ are the values placed on time by the marginal premium subscribers with relative values v^b/p^t and v^s/p^t, in the absence and presence of data caps, respectively. They can be obtained from the functions $v^{1,u}()$, $v^{1,c}()$, and $v^{1,o}()$. S^u and S^c are the corresponding surplus of the uncapped and capped marginal premium subscribers, which can be obtained from Equation (8).

We can easily prove that

$$p^{t,c} \geq p^{t,u}, \quad \partial S^u/\partial p^t\big|_{p^t=p^{t,u}} \geq \partial S^c/\partial p^t\big|_{p^t=p^{t,c}} > 0.$$

Using assumption E, we have

$$f(v^b/p^t, v^s/p^t, p^{t,c}) \geq f(v^b/p^t, v^s/p^t, p^{t,u}) \geq 0, \quad \frac{f(v^b/p^t, v^s/p^t, p^{t,c})}{\partial S^c/\partial p^t|_{p^t=p^{t,c}}} \geq \frac{f(v^b/p^t, v^s/p^t, p^{t,u})}{\partial S^u/\partial p^t|_{p^t=p^{t,u}}} \geq 0.$$

Since $C_2 \leq P_{21}/p^o$ and $p^o \geq p^\mu/\rho^{th}$, we have

$$P_{21} - p^\mu\left(C_2 - x^b t^b - X_1 t^{s,1}\right)/\rho^{th} > 0.$$

Marginal premium subscribers in G_o consume more data than cap C_2 if they subscribe to the premium tier, and hence $X_2 t^{s,2} + x^b t^b - C_2 \geq 0$. Thus, the sum of the second and third terms in Equation (17) can be expressed as

$$\begin{aligned}
&-\int_{\frac{v^b}{p^t}}\int_{\frac{v^s}{p^t}}\left(\left(P_{21} - \frac{p^\mu\left(C_2 - x^b t^b - X_1 t^{s,1}\right)}{\rho^{th}}\right)\left(\frac{f\left(v^b/p^t, v^s/p^t, p^{t,c}\right)}{\partial S^c/\partial p^t|_{p^t=p^{t,c}}} - \frac{f\left(v^b/p^t, v^s/p^t, p^{t,u}\right)}{\partial S^u/\partial p^t|_{p^t=p^{t,u}}}\right)\right.\\
&\qquad\left. + \frac{p^\mu\left(X_2 t^{s,2} + x^b t^b - C_2\right)}{\rho^{th}}\frac{f\left(v^b/p^t, v^s/p^t, p^{t,u}\right)}{\partial S^u/\partial p^t|_{p^t=p^{t,u}}}\right) d\frac{v^s}{p^t} d\frac{v^b}{p^t} \leq 0.
\end{aligned}$$

So, $\partial\left(Profit - Profit_0\right)/\partial P_{21} \leq 0$. Considering $\partial Profit_0/\partial P_{21} = 0$, it follows that $\partial Profit/\partial P_{21} \leq 0$.

We then turn to the basic tier rate X_1. In the absence of data caps, if price X_1 is changed by ΔX_1, some marginal premium subscribers in sets G_b and $G_u \cup G_d$ will switch between the basic tier and the premium tier. Denote these sets of users by ΔG_u and ΔG_d. Some marginal basic subscribers in sets G_n and G_b will switch between no Internet subscription and the basic tier. Denote this set of users by ΔG_n. Thus, Theorem 5.1 gives

$$\begin{aligned}
&\frac{\partial Profit_0}{\partial X_1} = \left(P_1 - k\right)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1} + P_{21}\frac{\partial N_2}{\partial X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial X_1} = \left(P_1 - k\right)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1}\\
&+ P_{21}\frac{\int_{\Delta G_u \cup \Delta G_d} f(v)dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda\left(G_b\right)}{\partial X_1} + \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_n \cup \Delta G_u \cup \Delta G_d} X_1 t^{s,1} f(v)dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u \cup \Delta G_d} X_2 t^{s,2} f(v)dv}{\Delta X_1} = 0,
\end{aligned}$$

where $\lambda(G_b)$ is the traffic from the users in set G_b. In the presence of data caps, if the rate X_1 is changed by ΔX_1, some marginal premium subscribers in sets $G_b \cup G_d$ and $G_u \cup G_c \cup G_o$ may switch between the basic tier and the premium tier. Denote these sets of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 5.2 gives

$$\begin{aligned}
&\frac{\partial Profit}{\partial X_1} = (P_1 - k)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1} + P_{21}\frac{\partial N_2}{\partial X_1} + p^o\frac{\partial O}{\partial X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial X_1}\\
&= \left(P_1 - k\right)\frac{\partial\left(N_1 + N_2\right)}{\partial X_1} + P_{21}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} f(v)dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\partial\left(\lambda\left(G_b\right) + \lambda\left(G_d\right)\right)}{\partial X_1}\\
&+ \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_n \cup \Delta G_u \cup \Delta G_c \cup \Delta G_o} X_1 t^{s,1} f(v)dv}{\Delta X_1} - \frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u \cup \Delta G_c \cup \Delta G_o} X_2 t^{s,2} f(v)dv}{\Delta X_1} + p^o\frac{\int_{\Delta G_o}\left(X_2 t^{s,2} + x^b t^b - C_2\right) f(v)dv}{\Delta X_1},
\end{aligned}$$

where $\lambda(G_d)$ is the traffic from the users in set G_d. Recall that N_1+N_2 is a function of X_1 and P_1. $\partial\lambda^b/\partial P_1$ is also a function of X_1 and P_1, according to Theorem 3.1. Thus, we have

$$\begin{aligned}\frac{\partial\left(Profit-Profit_0\right)}{\partial X_1} &= -\frac{p^\mu}{\rho^{th}}\frac{\partial\lambda\left(G_d\right)}{\partial X_1}+\frac{\int_{\Delta G_c\cup\Delta G_o}\left(P_{21}-p^\mu\left(C_2-x^b t^b-X_1t^{s,1}\right)/\rho^{th}\right)f(\upsilon)d\upsilon}{\Delta X_1}\\ &-\frac{\int_{\Delta G_d}\left(P_{21}-p^\mu\left(X_2t^{s,2}-X_1t^{s,1}\right)/\rho^{th}\right)}{\Delta X_1}+\left(p^o-\frac{p^\mu}{\rho^{th}}\right)\frac{\int_{\Delta G_o}\left(X_2t^{s,2}+x^bt^b-C_2\right)f(\upsilon)d\upsilon}{\delta X_1}.\end{aligned}\tag{18}$$

Obviously, the first term in Equation (18) is nonpositive. As X_1 decreases, more users subscribe to the premium tier. Marginal premium subscribers in G_o consume more data than data cap C_2 if they subscribe to the premium tier, and thus $X_2t^{s,2}+x^bt^b-C_2\geq 0$. Thus, the last term in Equation (18) is nonpositive, since $p^o\geq p^\mu/\rho^{th}$. Similar to Equation (17), we can prove the sum of the second and third terms in Equation (18) is also nonpositive. So, $\partial\left(Profit-Profit_0\right)/\partial X_1\leq 0$. Considering $\partial Profit_0/\partial X_1=0$, it follows that $\partial Profit/\partial X_1\leq 0$.

We finally turn to the premium tier rate X_2. In the absence of data caps, if price X_2 is changed by ΔX_2, some marginal premium subscribers in sets G_b and $G_u\cup G_d$ will switch between the basic tier and the premium tier. Denote these sets of users by ΔG_u and ΔG_d. Thus, Theorem 5.1 gives

$$\begin{aligned}\frac{\partial Profit_0}{\partial X_2} &= P_{21}\frac{\partial N_2}{\partial X_2}-\frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial X_2}=P_{21}\frac{\int_{\Delta G_u\cup\Delta G_d}f(\upsilon)d\upsilon}{\Delta X_2}\\ &-\frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u\cup\Delta G_d}\left(X_2t^{s,2}-X_1t^{s,1}\right)f(\upsilon)d\upsilon}{\Delta X_2}-\frac{p^\mu}{\rho^{th}}\int_{G_u\cup G_c\cup G_o\cup G_d}\left(t^{s,2}+X_2\frac{\partial t^{s,2}}{\partial X_2}\right)f(\upsilon)d\upsilon.\end{aligned}$$

In the presence of data caps, if the rate X_2 is changed by ΔX_2, some marginal premium subscribers in sets $G_b\cup G_d$ and $G_u\cup G_c\cup G_o$ may switch between the basic tier and the premium tier. Denote these sets of users by ΔG_u, ΔG_c, and ΔG_o. Thus, Theorem 5.2 gives

$$\begin{aligned}\frac{\partial Profit}{\partial X_2} &= P_{21}\frac{\partial N_2}{\partial X_2}+p^o\frac{\partial O}{\partial X_2}-\frac{p^\mu}{\rho^{th}}\frac{\partial\lambda}{\partial X_2}=P_{21}\frac{\int_{\Delta G_u\cup\Delta G_c\cup\Delta G_o}f(\upsilon)d\upsilon}{\Delta X_2}+p^o\frac{\int_{\Delta G_o}\left(X_2t^{s,2}+x^bt^b-C_2\right)f(\upsilon)d\upsilon}{\Delta X_2}\\ &-\frac{p^\mu}{\rho^{th}}\frac{\int_{\Delta G_u\cup\Delta G_c\cup\Delta G_o}\left(X_2t^{s,2}-X_1t^{s,1}\right)f(\upsilon)d\upsilon}{\Delta X_2}-\frac{p^\mu}{\rho^{th}}\int_{G_u}\left(t^{s,2}+X_2\frac{\partial t^{s,2}}{\partial X_2}\right)f(\upsilon)d\upsilon\\ &+\left(p^o-\frac{p^\mu}{\rho^{th}}\right)\int_{G_o}\left(t^{s,2}+X_2\frac{\partial t^{s,2}}{\partial X_2}\right)f(\upsilon)d\upsilon.\end{aligned}$$

Thus, we have

$$\begin{aligned}\frac{\partial\left(Profit-Profit_0\right)}{\partial X_2} &= \frac{p^\mu}{\rho^{th}}\int_{G_c\cup G_o\cup G_d}\left(t^{s,2}+X_2\frac{\partial t^{s,2}}{\partial X_2}\right)f(\upsilon)d\upsilon\\ &+\frac{\int_{\Delta G_c\cup\Delta G_o}\left(P_{21}-p^\mu\left(C_2-x^bt^b-X_1t^{s,1}\right)/\rho^{th}\right)f(\upsilon)d\upsilon}{\Delta X_2}\\ &-\frac{\int_{\Delta G_d}\left(P_{21}-p^\mu\left(X_2t^{s,2}-X_1t^{s,1}\right)/\rho^{th}\right)f(\upsilon)d\upsilon}{\Delta X_2}\\ &+\left(p^o-\frac{p^\mu}{\rho^{th}}\right)\left(\int_{G_o}\left(t^{s,2}+X_2\frac{\partial t^{s,2}}{\partial X_2}\right)f(\upsilon)d\upsilon+\frac{\int_{\Delta G_o}\left(X_2t^{s,2}+x^bt^b-C_2\right)f(\upsilon)d\upsilon}{\Delta X_2}\right).\end{aligned}\tag{19}$$

In the absence of data caps, we can easily prove that $\partial t^{s,2}/\partial X^2\geq 0$ from Equation (3), since users will devote more time to video streaming when the tier rate X_2 improves. Thus, the first term in Equation (19) is nonnegative. As X_2 increases, more users subscribe to the premium tier. Marginal premium subscribers in G_o consume more data than cap C_2 if they subscribe to the premium tier, and thus $X_2 t^{s,2} + x^b t^b - C_2 \geq 0$. Thus, the last term in Equation (19) is nonnegative, since $p^o \geq p^\mu / \rho^{th}$. Similar to Equation (17), we can prove the sum of the second and third terms in Equation (19) is also nonnegative. So, $\partial(Profit - Profit_0)/\partial X_2 \geq 0$. Considering $\partial Profit_0/\partial X_2 = 0$, it follows that $\partial Profit/\partial X_2 \geq 0$.

A.6. Proof of Theorem 6.1

The theorem follows directly from the following: user surplus with data caps is a linear increasing function of $v^s, \forall v^s \geq v^{s2}$ when $p^o = \infty$; user surplus without data caps is an increasing convex function of $v^s, \forall v^s \geq 0$.

A.7. Proof of Theorem 6.2

See the appendix in Dai and Jordan [2013b].

A.8. Proof of Theorem 6.3

The proof is similar to that of Theorem 6.2.

REFERENCES

ACLP NYU. 2012. A primer on data consumption: Trends & emerging business models. Technical Report, Advanced Communications Law & Policy Institute at New York Law School.

Richard Bennett. 2012. Comcast raises invisible data cap. http://www.innovationfiles.org.

David Besanko and Ronald Braeutigam. 2005. *Microeconomics*. John Wiley & Sons.

Hemant K. Bhargava and Vidyanand Choudhary. 2008. Research note—When is versioning optimal for information goods? *Manag. Sci. 54*, 5, 1029–1035.

Burst Media. 2011. Online video content & advertising video preferences habits and actions in Q4 2011. http://burstmedia.com.

N. Cardwell, S. Savage, and T. Anderson. 2000. Modeling TCP Latency. In *Proceedings of the IEEE INFOCOM*.

CCS Leeds. 2012. UK leased lines pricing. http://ccsleeds.co.uk.

Frank J. Cesario. 1976. Value of time in recreation benefit studies. *Land Econ. 52*, 1, 32–41.

Anindya Chaudhuri, Kenneth S. Flamm, and John Horrigan. 2005. An analysis of the determinants of internet access. *Telecommun. Policy 29*, 9–10, 731–755.

M. Chetty, R. Banks, A. Brush, J. Donner, and R. Grinter. 2012. You're capped: Understanding the effects of bandwidth caps on broadband use in the home. In *Proceedings of the ACM CHI Conference*.

Wei Dai and Scott Jordan. 2013a. Design and impact of data caps. In *Proceedings of the IEEE GLOBECOM*.

Wei Dai and Scott Jordan. 2013b. How do ISP data caps affect subscribers? In *Proceedings of the Telecommunications Policy Research Conference (TPRC)*.

Wei Dai and Scott Jordan. 2013c. Modeling ISP tier design. In *Proceedings of the 25th International Teletraffic Congress*.

Nicholas Economides. 2008. Net neutrality non-discrimination and digital distribution of content through the internet. *J. Law Policy Inform. Soc. 4*, 2.

Tom R. Eikebrokk and Oystein Sorebo. 1999. Technology acceptance in situations with alternative technologies: An empirical evaluation of the technology acceptance model in a multiple-choice situation. In *Proceedings of the 7th European Conference on Information Systems (ECIS'99)*.

George S. Ford. 2012. A most egregious act? The impact on consumers of usage-based pricing. Phoenix Center Policy Perspective No. 12-02, Technical Report, Phoenix Center for Advanced Legal & Economic Public Policy Studies.

GTRC. 1998. GVN's 10th WWW user survey. Georgia Tech Research Corporation. http://www.gtrc.gatech.edu.

H. Gerber and G. Pafumi. 1998. Utility functions: From risk theory to finance. *North Am. Actuarial J. 2*, 3.

Stacey Higginbotham. 2012. Which ISPs are capping your broadband, and why? http://gigaaom.com/2012/10/01/data-caps-chart.

Hibah Hussain, Danielle Kehl, Benjamin Lennett, and Patrick Lucey. 2012b. *Capping the nation's broadband future?* Technical Report. New America Foundation, Open Technology Institute.

Hibah Hussain, Danielle Kehl, and Patrick Lucey. 2012a. The destructive power of data caps. Technical Report. http://www.freepress.net/blog/2012/12/19/destructive-power-data-caps.

Libin Jiang, Shyam Parekh, and Jean Walrand. 2008. Time-dependent network pricing and bandwidth trading. In *Proceedings of the Network Operations and Management Symposium Workshops*.

Carlee Joe-Wong, Soumya Sen, and Sangtae Ha. 2013. Offering supplementary wireless technologies: Adoption behavior and offloading benefits. In *Proceedings of the IEEE INFOCOM*.

Dilip Joseph, Nikhil Shetty, John Chuang, and Ion Stoica. 2007. Modeling the adoption of new network architectures. In *Proceedings of the ACM CoNEXT Conference*.

Joel Kelsey. 2012. Comcast should eliminate punitive data caps altogether. Free Press Press Release.

Daniel A. Lyons. 2012. The impact of data caps and other forms of usage-based pricing for broadband access. Technical Report, Mercatus Center at George Mason University.

Jacob Joseph Orion Minne. 2012. Data caps: How ISPs are stunting the growth of online video distributors and what regulators can do about it. SSRN Technical Report, Social Science Research Network.

Papak Nabipay, Andrew Odlyzko, and Zhi-Li Zhang. 2011. Flat versus metered rates, bundling, and bandwidth hogs. In *Proceedings of NetEcon*.

Andrew Odlyzko, Bill St. Arnaud, Erik Stallman, and Michael Weinberg. 2012. Know your limits considering the role of data caps and usage based billing in internet access service. White Paper, Public Knowledge.

J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. 1998. Modeling TCP throughput: A simple model and its empirical validation. In *Proceedings of the ACM SIGCOMM*.

PK and NAF. 2011. Letter to FCC asking for an investigation of AT&T broadband data caps. Public Knowledge and New America Foundation.
http://www.publicknowledge.org/documents/letter-to-fcc-on-att-data-caps.

Sandvine. 2012. Global Internet phenomena report. Sandvine Technical Report.

Soumya Sen, Youngmi Jin, Roch Guerin, and Kartik Hosanagar. 2010. Modeling the dynamics of network technology adoption and the role of converters. *IEEE/ACM Trans. Netw. 18*, 1793–1805.

Soumya Sen, Carlee Joe-Wong, and Sangtae Ha. 2012a. The economics of shared data plans. In *Proceedings of the Workshop on Information Technologies and Systems (WITS)*.

Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. 2012b. Incentivizing time-shifting of data: A survey of time-dependent pricing for internet access. *IEEE Commun. Magazine 15*, 11, 91–99.

Soumya Sen, Carlee Joe-Wong, Sangtae Ha, Jasika Bawa, and Mung Chiang. 2013. When the price is right: Enabling time-dependent pricing of broadband data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

U.S. Census Bureau 2009. Money income of households 2009. United States Census Bureau Statistical Abstract.

U.S. Congress. 2012. S.3703 Data Cap Integrity Act, 112th Congress. Senate Bill.
http://www.congress.gov/bill/112th-congress/Senate-bill/3703.

David Waterman, Ryland Sherman, and Sung Wook Ji. 2012. The economics of online television: Revenue models aggregation and TV everywhere. In *Proceedings of the Telecommunications Policy Research Conference (TPRC)*.

Steven Weber and Vilas Veeraraghavan. 2007. Distributed algorithms for rate-adaptive media streams. In *Proceedings of the 8th INFORMS Telecommunications Conference*.

Roger Yu. 2012. Cable companies cap data use for revenue. http://www.usatoday.com. October 1, 2012.

Received November 2013; revised May, November 2014; accepted May 2015